PROCEEDINGS

ISBN 978-969-8858-24-7

Vol. 33

17th International Conference on Statistical Sciences:

THEME:

Sustainable Development Goals (SDGs) and Data Gap Challenges

January 21-23, 2019

at

National College of Business Administration & Economics

Lahore (Pakistan)

Jointly Organized by:

ISOSS

NCBA&E

PAK INSTITUTE OF STATISTICAL TRAINING AND RESEARCH

PISTAR

Sponsor

ISOSS Publications

44-A, Civic Centre, Sabzazar, Multan Road, Lahore, Pakistan

URL: www.isoss.net

Published by: ISOSS, Lahore, Pakistan.

"All papers published in the

PROCEEDINGS

were accepted after formal peer review

by the experts in the relevant field.

Dr. Munir Ahmad
Editor

CONTENTS

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 1-14

QUALITY OF LIFE OF HOME-BASED WORKERS IN THE PUNJAB: GENDERWISE COMPARISON

Saleha Naghmi Habibullah§, Farah Anjum, Khadija Noor Butt
Zohra Dar, Iqra Qurat-ul-Ain and **Roma Rehan**
Department of Statistics, Kinnaird College for Women
Lahore, Pakistan
§Email: saleha.habibullah@kinnaird.edu.pk

ABSTRACT

Home-based work plays a significant role in the informal economy. As such, it is important for the government to formulate policies and programs aimed at enhancing the quality of life of this segment of the society. In this regard, it is important to explore whether there exist any significant differences between the situations of various categories of home-based workers. This paper presents a statistical study aimed at comparing the circumstances of female home-based workers with those of the men in the Punjab province of Pakistan.

The study is based on data pertaining to the Home-Based Workers Survey 2017 carried out by the Bureau of Statistics, Punjab. Tests of hypotheses and confidence intervals contained in the paper throw light on the existing situation of approximately twelve million women and about one million men in the Punjab who are engaged in home-based work for purposes of income-generation. The non-trivial differences between the conditions of male and female home-based workers indicate there is a need to strive for more favorable conditions for the female HBWs which will be a step forward in the direction of gender equality, one of the Sustainable Development Goals adopted by the world.

1. INTRODUCTION

Home-based work comprises that sector of the informal economy in which items are produced inside homes for income-generation. This phenomenon exists in many countries of the world, particularly in the developing countries. As far as the Punjab province of Pakistan is concerned, it is estimated that, at this point in time, there exist approximately thirteen million home-based workers, a huge majority of whom are females. Evidently, it is an important segment of Punjab's labour-force and the government needs to cater for the needs of these people as much as those of other types of workers.

In order to launch projects and programs targeted at improving the quality of life of home-based workers (HBWs), it is useful to focus on the differences that occur due to gender, literacy level, area of residence (urban/rural), etc. This research-study focuses on gender and utilizes data related to the Home-Based Workers Survey 2017 carried out by the Bureau of Statistics, Punjab (BoS, 2017) to present a comparison between the situations of female and male home-based workers in the province. The Survey questionnaire invites responses not only from male and female but also, appropriately so, from transgender home-based workers but in the 2017 sample, only a very small number of HBWs were found to have transgender identities. As such, whereas for the male and female respondents, we have applied techniques of inferential statistics such as chi square test of independence and confidence intervals for proportions in the case of qualitative variables, and the independent samples t-test for differences between means and confidence intervals for means in the case of quantitative variables, for the transgender HBWs, we have simply presented point estimates of (i) means of quantitative variables of interest and (ii) proportions of HBWs falling in various categories in the case of qualitative variables.

A huge proportion of home-based workers in the Punjab being women, our primary interest in this study is to assess the situation of the female HBWs in comparison with the situation of their male counterparts. The various statistical analyses that have been performed throw light on differences that exist between the conditions of male and female HBWs. This indicates that, in a number of areas, there is a need for creating more favorable conditions for the females. Pakistan is a signatory of the Sustainable Development Goals adopted by the world and, as such, efforts for improving the situation of female home-based workers will be a step forward in the direction of SDG 5 which calls for achieving gender equality and empowerment of all women and girls.

2. LITERATURE REVIEW

During the past few decades, a number of papers have emerged regarding home-based workers in Pakistan. In this section, we present a brief review of some of these research-works.

Ali (1990) conducted a survey was conducted in 8 districts, Multan, Kasur, Bahawalpur, Layyah, Sargodha, Jhang, Attock and Sialkot, the objective being to carry out research on the problems of women workers in the informal sector. Five villages were surveyed from each district and a sample of up to 30 to 40 respondents was selected from each village. On the basis of findings, the researcher concluded that most of these women workers started work in early ages. They were illiterate, unskilled and in poor health. Their income was seen to be low and they were working to meet the basic requirements of their large families. The researcher suggested that their conditions could be improved by providing them reasonable wages according to their work, education, training, and asserted that they should be given access to better raw materials and markets and other facilities so that they can work in a better environment and play their part in the development of the country.

Khan and Kazmi (2003) asserted that various factors have generated unemployment in Pakistan which has created pressures on household members to increase household income through home-based subcontracted work in the informal sector. The researchers documented home-based work in four sectors in Pakistan in a value-chain context and attempted to track the distribution of revenue across the value-chain in order to demonstrate the extent to which home-based workers are deprived of the value that they create. The researchers concluded by making a case for welfare funds for home-based workers across sectors.

Akhtar (2011) designed a questionnaire to determine the demographic details, technical, education, and vocational training of HBWs, and the type of work undertaken by them. A sample of 525 home-based workers who had specialized in stitching, embroidery, handicraft, and paperwork was selected from Rawalpindi/Islamabad. On the basis of findings, the researcher concluded that education and training among women are growing (which is good) but they are still facing problems. The main problem is that their wages are low and they receive late receipt of payments. Also, they do not have any social and security protection.

Hassan and Azman (2014) carried out a study on female home-based workers of the 20-60 age groups in Pakistan. The data were collected from different areas of Lahore including Shahdra Town, Badami Bagh, Shalimar Town and others. The result showed that most of the home-based women workers work 12-14 hours per day and get wages between 30-40 rupees per day only. The researchers concluded that the poor in Pakistan, especially the female home-based workers have not only economic problems but they also face social issues. Neither is there any specific legislation for female home-based workers nor do they have social/legal recognition of their work.

3. HOME-BASED WORKERS SURVEY PUNJAB 2017

The Punjab Bureau of Statistics regularly conducts surveys on various socio-economic phenomena such as living standard, health, education, etc. The Home-Based Workers Survey 2017 provides information on approximately 200 variables collected from 13,682 home-based workers. Whereas many of the initial variables are demographic in nature, a majority of the subsequent variables pertain to the working conditions and quality of life of the HBWs. The Bureau has carried out descriptive analyses of the collected data including cross-tabulations, bar charts, pie charts and has produced a Report in order to equip the government with important information regarding various aspects of the HBWs work-lives which will enable the government to devise policies and programs for the well-being of the home-based workers.

4. ANALYSIS OF THE 2017 SURVEY DATA

In this section, we present results of inferential analyses pertaining to variables that can be regarded as indicators of quality of life of the home-based workers, our target being to bring to light the differences that exist between the situations of the male and female HBWs.

For the quantitative variables, we (i) compare the values of the variables of interest for male and female HBWs through appropriate tests of significance as well as confidence intervals computed for the mean values of these variables for the male and female HBWs separately. On the other hand, for the qualitative variables, we compare the situations of male and female HBWs through cross-tabulations accompanied by the chi square test of independence as well as confidence intervals for proportions of male and female HBWs possessing particular attributes.

As far as transgender HBWs concerned, out of the information on various variables pertaining to 13,682 HBWs the Home-Based Workers Survey 2017 (Punjab) contains data pertaining to 13,682 home-based workers out of which only 8 persons possess transgender identities. As this number is too small for any reliable inferential analyses, we simply present descriptive statistics i.e. means of quantitative variables of interest and proportions of transgender HBWs falling in various categories in the case of qualitative variables.

4.1 Variable No. 96: Basic Income

Obviously, one of the most important questions in the Survey questionnaire is regarding the basic income of the home-based workers. As information regarding income has been provided by each respondent in one out of four different modes (i.e. daily, weekly, fortnightly or monthly income), it is not possible to compute the mean income directly. Table 4.1 contains the results of calculations that have been performed in order to (i) find the mean income mode-wise and (ii) convert all means into one and the same unit i.e. mean monthly income. The last column of Table 4.1 contains the mean monthly income-figures gender wise. From this column, it is obvious that the mean monthly income of male home-based workers is approximately double of the mean monthly income of female HBWs. The mean monthly income of transgender HBWs seems to be more or less the same as that of the female home-based workers.

The pooled variances of the monthly income-figures of the male and female HBWs are 102689261.9 and 9711397.291 respectively. Hence:

(i) The 99% confidence interval for the true mean monthly income of the male home-based workers in the Punjab is (Rs. 12834.90, Rs. 14380.72);

(ii) The 99% confidence interval for the true mean monthly income of the female home-based workers in the Punjab is (Rs. 6298.89, Rs. 6444.51).

4.2 Variable No. 92: How Many Hours did you do Current (Main) Work during the Last 24 Hours?

It seems logical to assume that, in a vast majority of cases, the amount of income earned by home-based workers should be directly related to the ***number of hours*** that are put in by the HBWs for producing marketable items. In order to compare the number of hours spent by male HBWs with the number of hours spent by females, we opt for the Mann Whitney U Test due to the fact that the distribution of income for the male HBWs as well as for the female HBWs is not normal. Application of the Test to Variable No. 92

Table 4.1
Mean Monthly Income of Male, Female & She-Male HBWs

Gender	Mode of Income	Number of HBWs	Mean Income Mode-Wise (Rs.)	Mean Monthly Income (Rs.)	Mean Monthly Income Gender Wise (Rs.)
Male	1=Daily	467	659.15	19774.50	13,607.81
	2=Weekly	482	2112.96	9055.54	
	3=Fortnightly	51	5682.35	11364.70	
	4=Monthly	109	8366.97	8366.97	
Female	1=Daily	4567	326.63	9798.90	6371.70
	2=Weekly	5633	1053.87	4516.59	
	3=Fortnightly	607	1577.97	3155.94	
	4=Monthly	1011	3156.81	3156.81	
She-Male	1=Daily	5	260.00	7800.00	6375.00
	2=Weekly	3	933.33	3999.99	
	3=Fortnightly	0	><	><	
	4=Monthly	0	><	><	

(Hours of work in a day) yields U = 4462603.000 with a p-value of 0.000 indicating that there exists a significant difference between the number of hours of work per day by male and female HBWs. From Table 4.2, it is evident that the males are spending a greater number of hours on the home-based work than the females. It seems obvious that the reason why the female HBWs are able to put in fewer hours for producing marketable items is that they have a heavy load of domestic responsibilities.

Table 4.2
Table of Ranks for Number of Hours of Work per Day by Male and Female HBWs

	Gender	n	Mean Rank	Sum of Ranks
How many hours did you do current (main) work during the last 24 hours?	Male	1116	8431.75	9409835.00
	Female	11872	6312.39	74940731.00
	Total	12988		

As far as transgender HBWs concerned, Table 4.3 presents the profile of the eight she-males in the 2017 sample. The mean comes out to be 6.13 hours.

4.3 Variable No. 65: Who Decides to Spend Income which you Earn from your Current (Main) Work?

It seems obvious that the decision regarding how to spend the income earned through home-based work should be in the hands of the HBWs themselves. However, cross-tabulation of gender with this variable yields the figures contained in Table 4.4. From the table, it can be seen that 27.2% of the female and 25.7% of the male HBWs do not decide themselves where and how to spend the income earned by them. Application of the chi square test of independence yields $\chi^2 = 1.17$ with a p-value of 0.279 which is greater than 0.05 indicating that decision of spending income is independent of gender. (There does not exist a significant difference between the proportions of male and female HBWs who are unable to take this decision themselves.)

Table 4.3
Number of Hours of Work per Day By Transgender HBWs

No. of Hours of Work Per Day	No. of She-Males
4	3
5	1
7	2
8	1
10	1
Total	**8**

Table 4.4
Cross-Tabulation of Gender with Variable No. 65

Gender	Who Decides to Spend Income which you Earn from your Current (Main) Work?		Total
	Self	Other Family Members	
Male	829 74.3%	287 25.7%	1116 100%
Female	8651 72.8%	3236 27.2%	11887 100%
Total	9480 72.9%	3523 27.1	13003 100%

The 99% confidence interval for the true proportion of home-based workers (males and females taken together) who are unable to take a decision themselves regarding where and how to spend their income is given by

$$0.272 \pm 2.54\sqrt{\frac{0.272(0.728)}{13003}} = (0.262, 0.282) = (26.2\%, 28.2\%)$$

Thus, we can say that we are 99% confident that, in the Punjab, between 26.2% and 28.2% of the home-based workers (males and females combined) are unable to take a decision themselves regarding where and how to spend their income. Obviously, this is a non-trivial proportion.

On the other hand, as far as the transgender HBWs are concerned, analysis of the 2017 Survey data revealed that all eight transgenders included in the 2017 sample take the decision regarding where and how to spend their income themselves.

4.4 Variable No. 71: Do you want to do another Work Instead of your Current (Main) Work?

One of the questions in the 2017 Survey questionnaire aims to determine the proportion of HBWs who are desirous of changing the nature of their work. Cross-tabulation of gender with this variable yields the information given in Table 4.5.

We find that, in the 2017 sample, approximately 50.6% of the female and 48.2% of the male HBWs do wish to replace their current work by some other type of home-based work. Application of the chi square test of independence yields $\chi^2 = 2.347$ with a p-value of 0.126 which is greater than 0.05 indicating that the desire to replace current work by some other type of home-based work is independent of gender. (There does not exist a significant difference between the proportions of male and female HBWs who are desirous of replacing their current work by another type of work.)

Table 4.5
Cross-Tabulation of Gender with Variable No. 71

Gender	Do you want to do another Work Instead of your Current (Main) Work?		Total
	Yes	No	
Male	538 48.2%	578 51.8%	1116 100%
Female	6012 50.6%	5868 49.4%	11880 100%
Total	6550 50.4%	6446 49.6%	12987 100%

The 99% confidence interval for the true proportion of home-based workers (males and females put together) who wish to replace their current work by some other type of home-based work is given by

$$0.504 \pm 2.54\sqrt{\frac{0.504(1-0.504)}{12987}} = (0.493, 0.515) = (49.3\%, 51.5\%)$$

We are 99% confident that between 49.3% and 51.5% of the HBWs (males and females combined) do wish to replace their current work by some other type of home-based work. As it is a large proportion, it will definitely be helpful to include a question on the main reason for this desire in a survey to be conducted in the not too distant future.

Regarding transgender HBWs, Table 4.6 presents the profile of the eight she-males in the 2017 sample. The proportion of she-males who do wish to replace their current work by another type of work is 37.5%.

Table 4.6
Number of Transgender HBWs Who Want to Replace their Current (Main) Work by another Type of Work

Do you want to do another work instead of your current (main) work?	No. of She-Males
Yes	3
No	5
Total	8

4.5 Variable 152: Have You Suffered Any Injury / Accident during Current / Additional Work?

A risk-free environment minimizing the possibility of physical injury to workers is one of the important concerns in any industry. One of the questions in the 2017 Survey was aimed at determining the proportion of HBWs who had suffered any injury / accident during their work. Cross-tabulation of gender with this variable has led to the figures contained in Table 4.7. From the table, it is obvious that, in the 2017 sample, approximately 18.5% of the HBWs (males and females put together) had suffered an injury/accident during their work and the proportion of male HBWs in this situation is approximately 6.5% higher than the proportion of females who have experienced an injury/accident during their work-hours. Application of the chi square test of independence yields χ^2=28.482 with a p-value of 0.000 which is highly significant indicating an association between gender and vulnerability to injury/accident during the home-based work. (A higher proportion of male than female HBWs seem to be in danger of injury/accident, and one of the possible reasons may be the difference that might exist between the nature of work carried out by male HBWs and the nature of work carried out by their female counterparts. it will definitely be helpful to explore the reasons for this phenomenon in an upcoming survey.)

Table 4.7
Cross-Tabulation of Gender with Variable No. 152

Gender	Have you suffered any injury / accident during current / additional work?		Total
	Yes	No	
Male	272 24.4%	843 75.6%	1115 100%
Female	2127 17.9%	9750 82.1%	11877 100%
Total	2399 18.5%	10593 81.5%	12992 100%

The 99% confidence interval for the true proportion of female HBWs susceptible to injury/accident during the home-based work is

$$0.179 \pm 2.54\sqrt{\frac{0.179(0.821)}{11877}} = (0.170, 0.187) = (17\%, 18.7\%)$$

On the other hand, the 99% confidence interval for the true proportion of male HBWs susceptible to injury/accident during work-hours is

$$0.244 \pm 2.54\sqrt{\frac{0.244(0.756)}{1115}} = (0.211, 0.276) = (21.1\%, 27.6\%)$$

Focusing on transgender HBWs, analysis of the 2017 Survey data revealed that one of the eight transgenders included in the 2017 sample had suffered an injury/accident during work-hours.

4.6 Variable No. 163: Are / Did you get any Benefit from Micro Finance Group / Organization / Institution?

Microfinance or microcredit is a type of financial service that is provided to low-income individuals or groups who otherwise would have no other access to financial services. Microloans can be small or large, the goal of microfinancing being to give needy people an opportunity to become self-sufficient.

One of the questions in the 2017 Survey questionnaire was: "Are / Did You Get Any Benefit From Micro Finance Group / Organization / Institution?" Cross-tabulation of gender with this variable yields the figures given in Table 4.8:

Table 4.8
Cross-Tabulation of Gender with Variable No. 163

Gender	Are / Did you get any benefit from Micro Finance Group / Organization / Institution?		Total
	Yes	No	
Male	212 19.0%	905 81.0%	1117 100%
Female	1763 14.9%	10107 85.1%	11870 100%
Total	1975 15.2%	11012 84.8%	12987 100%

From Table 4.8, we see that, in the 2017 sample, overall, only 15.2% of the HBWs who are getting some type of benefit from a micro finance group /organization / institution. It seems that there is a substantial need to promote awareness among home-based workers regarding the financial benefits of microcredit and other products of microfinance groups/organizations.

As far as comparison between the male and female HBWs is concerned, application of the chi square test of independence yields $\chi^2 = 13.484$ with a p-value of 0.000 which is highly significant indicating that gender and get benefit from micro finance group is not independent. (A slightly higher proportion of male than female HBWs seem to be utilizing the financial products of microfinance groups/organizations.)

The 99% confidence interval for the true proportion of male home-based workers who gets any benefit from micro finance group / organization / institution is

$$0.190 \pm 2.54\sqrt{\frac{0.190(0.81)}{1117}} = (0.160, 0.219) = (16.0\%, 21.9\%)$$

On the other hand, the 99% confidence interval for the true proportion of female home-based workers who gets any benefit from micro finance group / organization / institution is

$$0.149 \pm 2.54\sqrt{\frac{0.149(0.851)}{11870}} = (0.141, 0.157) = (14.1\%, 15.7\%)$$

As such, we can say that we are 99% confident that the true proportion of male home-based workers who get some type of benefit from micro financial group/institution is between 16.0% and 21.9% whereas the true proportion of the female HBWs who get such a benefit is between 14.1% and 15.7%. The difference is not very large but there is a difference.

As far as transgender HBWs are concerned, analysis of the 2017 Survey data revealed that not a single one of the eight transgenders included in the 2017 sample had obtained any benefit from a micro financial group/institution.

4.7 Variable No. 166: Are you aware about any Institute/Organization/Trade Union which is Working for Rights of Home-Based Workers?

Indubitably, awareness regarding one's rights as a member of the labour-force is of great importance in any profession. The ground reality, however, is that in Pakistan as well as in many other developing countries, due to illiteracy/little education, a large number of low-paid workers are unaware of their rights. One of the questions in the 2017 Survey questionnaire was: "Are you aware about any institute/organization/trade union which is working for rights of home-based workers?". Analysis of the Survey data reveals that, in the 2017 sample, as few as 6.3% of the home-based workers (males and females taken together) are aware of any such organization/trade union! As far as comparison between the two genders is concerned, Table 4.9 presents the proportions of male and female HBWs who are aware about at least one institute/organization/trade union which is working for the rights of home-based workers. Application of the chi square test of independence yields $\chi^2 = 3.005$ with a p-value of 0.083 which is greater than 0.05 indicating that there does not exist an association between gender and awareness regarding organizations/trade unions working for the rights of home-based workers. (There does not exist a significant difference between the proportions of male and female HBWs who possess this awareness).

Table 4.9
Cross-Tabulation of Gender with Variable No. 166

Gender	Are you aware about any institute / organization / trade union which is working for the rights of home-based workers?		Total
	Yes	No	
Male	59 7.7%	708 92.3%	767 100%
Female	446 6.1%	6866 93.9%	7312 100%
Total	505 6.3%	7574 93.7%	8079 100%

The 99% confidence interval for the true proportion of home-based workers (males and females put together) who are aware about institute/organization trade unions that are working for the rights of home-based workers is given by

$$0.0625 \pm 2.54\sqrt{\frac{0.0625(1-0.0625)}{8079}} = (0.056\,, 0.069\,) = (5.6\%, 6.9\%)$$

Focusing on transgender HBWs, the 2017 Survey data revealed that not a single one of the eight transgenders included in the 2017 sample possessed awareness regarding any institute/organization/trade union that is working for the rights of home-based workers.

4.8 Variable No. 180: Do you have some Awareness about the Security Laws Against the (Sexual) Harassment?

The menace of sexual harassment at the work-place exists all over the world. In many countries including Pakistan, laws are in place to curb this practice. The law entitled "The Protection Against Harassment of Women at Workplace Act, 2010" is a legislative instrument that defines sexual harassment clearly and prohibits harassment at the workplace. The important question is: is a huge majority of working women aware of this law? One of the questions in the 2017 Survey was: 'Do you have some awareness about the security laws against the (sexual) harassment?' Cross-tabulation of gender with this variable yields the information presented in Table 4.10. From the table we see that, in the 2017 sample, 19% of the male HBWs and only 11% of the female HBWs have some awareness about the security laws against sexual harassment. Application of the chi square test of independence yields $\chi^2 = 63.26$ with a p-value of 0.000 which is much less than 0.05 indicating that there exists an association between awareness regarding the security laws against sexual harassment and gender. (A significantly lower proportion of female than male HBWs seem to possess some awareness about these laws.)

The 99% confidence interval for the true proportion of male home-based workers in the Punjab who are aware about security laws against sexual harassment is

Table 4.10
Cross-Tabulation of Gender with Variable No. 180

Gender	**Do you have some awareness about the security laws against the (sexual) Harassment?**		**Total**
	Yes	**No**	
Male	212 19.0%	904 81.0%	1116 100%
Female	1306 11.0%	10570 89.0%	11876 100%
Total	1518 11.7%	11474 88.3%	12992 100%

$$0.190 \pm 2.54\sqrt{\frac{0.190(1-0.190)}{1116}} = (0.160, 0.219) = (16\%, 21.9\%)$$

On the other hand, the 99% confidence interval for the true proportion of female home-based workers in the Punjab who possess this awareness is

$$0.11 \pm 2.54\sqrt{\frac{0.11(1-0.11)}{11876}} = (0.102, 0.117) = (10.2\%, 11.7\%)$$

There seems to be a need for launching a *campaign* to bring about wide-spread awareness among female home-based workers regarding security laws against sexual harassment as they are the ones who are much more vulnerable to this type of threat than their male counterparts.

As far as transgender HBWs concerned, inspection of the responses obtained from the eight transgenders in the 2017 sample reveals that only one of them has some awareness of the security laws against sexual harassment.

5. COMMENTS AND CONCLUSION

In this paper, we have presented a study based on data pertaining to the Home-Based Workers Survey 2017 carried out by the Bureau of Statistics, Punjab, the primary objective being to bring to light the differences that exist between the situations of male and female home-based workers. Inferential analyses including tests of hypotheses and confidence intervals have revealed that:

i) The mean monthly income of male home-based workers is approximately Rs. 13,500/- whereas the mean monthly income of female HBWs is approximately Rs. 6375/-. It is obvious that the income of male HBWs is pretty low but the income of their female counterparts is *less than half* of the income of their male counterparts;

ii) Females having a heavy, the male HBWs are spending a greater number of hours per day on home-based work leading to production of marketable items than the females;

iii) A non-trivial proportion of 26% to 28% of the home-based workers are unable to take a decision themselves regarding where and how to spend their income which is a cause for concern. However, interestingly, no significant difference was found between the situations of the males and females in this regard;

iv) Approximately 50% of the HBWs wish to replace their current work by some other type of home-based work; no significant difference was found between the situations of the males and females in this regard but it will certainly be helpful to explore the reason(s) for this desire in an upcoming survey;

v) The overall proportion of HBWs who are getting some type of benefit from a Micro finance group /organization / institution is approximately 15% (i.e. pretty low) with a slightly higher proportion of male than female HBWs utilizing the financial products of such groups/organizations;

vi) As far as awareness regarding institutes/organizations/trade unions that are working for the rights of home-based workers is concerned, a very small proportion of HBWs seem to be aware of any such organization/trade union; interestingly, there does not exist a significant difference between the proportions of male and female HBWs who possess this awareness;

vii) Focusing on awareness regarding security laws against sexual harassment, it appears that about one-fifth of the male HBWs and only about one-tenth of the female HBWs have some awareness about these laws.

Among the phenomena considered in this paper, the only one where the female HBWs seem to have an edge over their male counterparts is the risk of injury/accident during work-hours. (A higher proportion of male than female HBWs seem to be in danger of injury/accident, and it will definitely be useful to probe into the factors responsible for this phenomenon in a future study).

The above-mentioned results have brought into light non-trivial differences between the conditions of male and female home-based workers in the Punjab province of Pakistan. Evidently, there is a need to struggle for the uplift of the female HBWs who comprise an important segment of the informal economy. Multi-faceted efforts to *empower* the female HBWs by bringing their rates of remuneration closer to those of their male counterparts, reducing their load of domestic responsibilities allowing them more time to produce marketable items, increasing their awareness-level regarding financial products such as microloans and regarding their rights will bring the province closer to the attainment of Sustainable Development Goal No. 5 i.e. 'gender equality'.

REFERENCES

1. Akhtar, S. (2011). *Searching for the Invisible Workers: A Statistical Study of Home Based Workers in Pakistan*. Islamabad: ILO.
2. Ali, K. (1990). Problems of Working Women in the Rural Informal Sector of Multan District. *Pakistan Economic and Social Review*, 27(2), 89-104.
3. BoS (2017). *Home Based Worker Survey. Punjab*, Bureau of Statistics Punjab, Lahore.
4. Hassan, S.M. and Azman, A. (2014). Visible work, invisible workers: A study of women home based workers in Pakistan. *International Journal of Social Work and Human Services Practice*, 2(2), 48-55.
5. Khan, S.R. and Kazmi, S. (2003). Revenue Distribution across Value Chains: The Case of Home-based Subcontracted Workers in Pakistan. *Working Paper No: 2003-04*, *June 2003*, Department of Economics, University of Utah, USA.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 15-20

COMPARISON OF FINANCIAL RATIOS OF ENGRO FOODS VS NESTLE

Ammara Nawaz Cheema
Department of Mathematics, Air University, Islamabad, Pakistan
Email: ammara.au@gmailcom

ABSTRACT

Engro Foods Limited is among the greatest and quickest developing organizations in Pakistan. Nestle Pakistan is pleased with its responsibility regarding brilliance in item security and quality. There are different products of Engro Foods and Nestle. We study financial ratios of these two as financial ratios are the basic and most important part of any business. It describes the firm's financial position. The data set consists 2014-2016 financial ratios of Engro Foods and Nestle. Results indicates that nestle is an international brand and has expanded its business on the large geographical area and also offers the large range of products. But on the other side Engro Foods offers the limited range of products and most of them are dairy product. From the financial statement it is clear that the financial position of the Engro Foods is far better than Nestle because nowadays it is more preferred by the costumers. It gives more return because it gain more profits then Nestle. Engro deals with the limited product in limited geographical area but on the basis of financial ratios Engro has a better financial position and also has an opportunity to expand its business. Both the companies have some opportunities and threats and they need to work on it.

1. INTRODUCTION

Engro Foods Limited is among the greatest and quickest developing organizations in Pakistan with a dream to take into account neighborhood needs with items fitting in with worldwide measures. Engro Foods started its business in 2005-06.

Very energetic about giving a great many individuals over the length and expansiveness of Pakistan and past with a definitive brand understanding, their item portfolio contains a portion of the nation's greatest and top brands.

In any case, regardless of whether it is their thick, rich generally useful drain, tasty dessert high on sustenance content or invigorating scope of refreshment, their approach remains to a great extent the same as they endeavor to keep item advancement at the cutting edge of their controlling rationality and shopper fulfillment at the heart of their operational procedure.

What's more, that is unequivocally what empowers them to advance as Pakistan's chief nourishments organization conveying top quality items implanted with a nearby flavor.

Manufacture branded sustenance business to enhance personal satisfaction by offering top notch, moderate and high healthful items to their customers while amplifying partner's esteem**.** Aim to changing the organization inside five years into first national sustenance industry giant, then in to provincial drive lastly into a worldwide player.

Nestle Pakistan is pleased with its responsibility regarding brilliance in item security and quality and to offering some incentive and expects to be the main Nutrition, Health and Wellness company. As a socially capable corporate, they generally concentrate on conditions cordial operations, moral business rehearse and their obligation towards the groups.

Nestle in Pakistan is working since 1988 under a joint wander with Milk Pak ltd and assumed control administration in 1992.

The organization's technique is guided by Nestle's corporate business principles which are in accordance with globally acknowledged prescribed procedures and moral execution culture. Nestle's current items develop through advancement and remodel while keeping up an adjustment in geographic exercises and product offerings. Long haul potential is never yielded for here and now execution. The company's need is to bring the best and most significant items to individuals, wherever they are, whatever their needs are, and for all age gatherings.

To decidedly upgrade the personal satisfaction of individuals of Pakistan by all that they do through their kin, their brands, items and their creating shared value activities**.** Nestle vision is to be the all-around perceived driving Nutrition, Health and Wellness company. Nestle Pakistan subscribes completely to this vision and the qualities that accompanied it.

Specifically, they imagine:

- Driving a dynamic, energetic and proficient workforce, glad for their legacy and positive about what's to come.
- Meeting the health needs of buyers of any age from outset to seniority, from nourishment to delight, through an imaginative arrangement of marked sustenance and drink results of the most elevated quality.
- Conveying shareholder esteem through beneficial long haul development, while proceeding to play a huge and dependable part in the social, monetary, and natural segments of Pakistan.

There are different products of Engro Foods and Nestle. Engro Foods includes dairy and beverages in which Olper's Milk, Olper's cream, Tarang, Dairy Omung etc. are included and also frozen desserts like Ice Creams, Omore Sticks, Cup Desserts, Cone Ice Creams etc.

Nestle includes chilled dairy (Nestle Yogurt, Nestle Milk Pak), Juices (Nestle Fruita Vital, Nestle Fruita Vital Calorie SMART), Breakfast Cereals (Nestle Corn Flakes, Nestle Milo), Baby Foods (Nestle Cerelac) etc.

2. METHODOLOGY

In our research work on Engro Foods & nestle we have used only secondary data methodology. Because all the statistical data that we needed for our project was easily available on internet.

We collected Data from following sites:

- https://www.statista.com/topics/1439/nestle/
- http://www.nestle.com/investors/sharesadrsbonds/stockquotes/graphs
- http://www.engrofoods.com/downloads.html
- http://www.engro.com/downloads/
- https://pkfinance.info/kse/stock/efoods

Financial Position:

The status of the benefits (assets), liabilities and owner's equity of an association, as reflected in its money related explanations and also called financial condition or position.

We analyzed the financial position of Engro Foods and Nestle for the years 2014, 2015 and 2016. In which we have calculated current ratio, quick ratio, net working capital and debt ratio.

	Engro Foods				Nestle			
	2016	**2015**	**2014**	**Avg.**	**2016**	**2015**	**2014**	**Avg.**
Current Ratio	1.92	1.5	1.27	1.563	0.85	0.8	1.03	0.893
Quick Ratio	0.89	0.26	0.18	0.443	0.18	0.17	0.19	0.18
Net Working Capital	0.2	0.14	0.08	0.14	-0.041	-0.031	0.008	-0.021
Debt Ratio	0.3	0.43	0.55	0.426	0.49	0.48	0.46	0.476

The average current ratio of Engro Foods is greater than Nestle so it indicates that company is in well position to cover its current and short term liabilities. Similarly, the average quick ratio of Engro Foods is greater than Nestle which means that Engro Foods is sufficiently able to meet its short term obligations. The average networking capital of Engro Foods is also greater than Nestle that shows there will be less liquidity problems as compared to Nestle. Debt ratio of Engro Foods is less than 0.5 which means that company's asset are financed by equity whereas Debt ratio of Nestle is approximately 0.5 which means that company's assets are financed by debt.

After finding and analyzing above ratios we can say that financial position of Engro Foods is better than Nestle.

3. RESULTS

Engro Foods:

Measure of Central Tendency:

Sales Revenue(Rs)		**Units**	
Mean	108.67	Mean	0.76
Median	108	Median	0.76

On average, the sales revenue is 108.67**.** The median of sales revenue is 108**.** On average, the units are 0.76**.** The median of units is 0.76

Measure of Dispersion:

Sales revenue (Rs)		**Units**	
Range	6	Range	0.08
Variance	9.3	Variance	0.0016
Standard deviation	3.05	Standard deviations	0.04

The range of sales revenue is 6 and square deviation of the sales revenues from their average sales revenue is 9.3. The deviation of the sales revenue from their average sales revenue is 3.05. Range of units is 0.08. The square deviation of the units from their average units is 0.0016. The deviation of the units from their average units is 0.04.

The regression equation of Engro Foods according to sales revenue and units is:

Y = 51.67 + 75x

The correlation of Engro Foods according to sales revenue and units is 0.98. So there is positive strong association between sales revenue and units.

Nestle:

Measure of Central Tendency:

Sales Revenue(Rs)		**Units**	
Mean	95.9	Mean	0.62
Median	97.8	Median	0.65

On average, the sales revenue is 95.9. The median of sales revenue is 97.8. On average, the units are 0.62. The median of the units is 0.65

Measure of Dispersion:

Sales revenue (Rs)		**Units**	
Range	6.7	Range	0.15
Variance	13.93	Variance	0.063
Standard deviation	3.73	Standard deviations	0.07

The range of sales revenue is 6.7 and square deviation of the sales revenue from their average sales revenue is 13.93**.** The deviation of the sales revenue from their average sales revenue is 3.73. The range of units is 0.15 and square deviation of the units from their average units is 0.063. The deviation of the units from their average unit is 0.07.

The regression equation of Nestle according to sales revenue and units is:

Y = 66.96 + 46.67x

The correlation of sales revenue and units is 0.99. So there is positive strong association between sales revenue and units.

Comparison:

	Engro Foods		Nestle	
	Sales Revenue	**Units**	**Sales Revenue**	**Units**
Mean	108.67	0.76	95.9	0.62
Median	108	0.76	97.8	0.65
Range	6	0.08	6.7	0.15
Variance	9.3	0.0016	13.93	0.063
Standard deviation	3.05	0.04	3.73	0.07

The mean of sales revenue of Engro Foods is greater than the mean of sales revenue of Nestle. The median of sales revenue of Engro Foods is greater than the median of sales revenue of Nestle. The range of sales revenue of Engro Foods is less than the range of sales revenue of Nestle. The variance of sales revenue of Engro Foods is less than the variance of sales revenue of Nestle. The standard deviation of sales revenue of Engro Foods is less than the standard deviation of sales revenue of Nestle.

The mean of units of Engro Foods is greater than the mean of units of Nestle. The median of units of Engro Foods is greater than the median of units of Nestle. The range of units of Engro Foods is less than the range of units of Nestle. The variance of units of Engro Foods is less than the variance of units of Nestle. The standard deviation of units of Engro Foods is less than the standard deviation of Nestle.

	Engro Foods	Nestle
Regression	Y = 51.67 + 75x	Y = 66.96 + 46.67x
Correlation	0.98	0.99

The correlation of Engro Foods according to their sales revenue and units is less than the correlation of Nestle according to their sales revenue and its units by 0.01

4. COMMENTS AND CONCLUSION

After all finding, it is concluded that financial ratio are the basic and most important part of any business. It describes the firm s financial position. As the data indicates that nestle is an international brand and has expanded its business on the large geographical area and also offers the large range of products. But on the other side Engro Foods offers the limited range of products and most of them are dairy product.

From the financial statement it is clear that the financial position of the Engro Foods is far better than Nestle because nowadays it is more preferred by the costumers. It gives more return because it gain more profits then Nestle. Engro deals with the limited product in limited geographical area but on the basis of financial ratios Engro has a better financial position and also has an opportunity to expand its business. Both the companies have some opportunities and threats and they need to work on it

Suggestions:

Nestle doesn’t have any direct market and outlets so it can be disadvantages so they should facilitate their customers through pricing strategies and if they start direct market or open the outlets so the prices will fall automatically and costumers need not to pay any extra money to suppliers**.** Nestle Pakistan mostly depends on local raw material and sometimes the quality of raw material and some time the quality of raw material is not good as other countries so they should not rely on the local raw material if they want to provide the quality product.

Engro Foods should introduce other product lines and expand the business. Engro Foods should distribute their product to more geographically areas**.** As nestle is well known product and Engro Foods is not as known internationally as nestle is so they need to spend more money on the marketing activities. Engro food is better than nestle in financial analysis so if they expand their product line and cover the same area as nestle has covered so Engro can appear as a strong competitor of nestle and Haleeb.

REFERENCES

1. http://www.engrofoods.com/pdf/engro-foods-annual-report-2016.pdf
2. http://www.nestle.pk/asset-library/documents/financial_reports/nestl%C3%A9_financials_2015.pdf
3. http://searchdatacenter.techtarget.com/definition/statistical-mean-median-mode-and-range
4. http://www-ist.massey.ac.nz/dstirlin/cast/cast/scentre/centre5.html
5. http://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 21-34

MODELING AND FORECASTING OF THE GENERAL SHARE INDEX OF WHOLE SALE PRICE OF RAW MATERIAL

Ammara Nawaz Cheema
Department of Mathematics, Air University, Islamabad, Pakistan
Email: ammara.au@gmailcom

ABSTRACT

The aim of this study is to determine the model for the general share index of whole sale price of raw material in Pakistan. Here we have the time series data of whole sale price of raw material: cotton, tobacco, wool and skins from Jan, 1980 to Aug 2014. Firstly we have check the stationarity of the data, the HEGY test is used for testing the seasonal and the non seasonal unit roots. Box-Pierce Q-Statistic, Breusch-Godfrey Serial Correlation LM test are used for testing the white noise of the error term and serial correlation respectively. We find that there are more than one potential models which fit to this data, but the final and best model is selected using the criteria of the AIC, BIC and RMSE. At the end we have forecast observation for the suitable selected model. Since the forecasted values through the fitted model is close to the original values so we conclude that our fitted model is a good model.

KEYWORDS

Stationarity, HEGY test, unit roots, Box-Pierce Q-Statistic, Breusch-Godfrey Serial Correlation LM test

1. INTRODUCTION

Time series analysis and its applications have become increasingly important in various fields of research, such as business, economics, engineering, medicine, environmetrics, social sciences, politics, and others. Since Box and Jenkins (1970, 1976) published the seminal book *"Time Series Analysis: Forecasting and Control"*, a number of books and a vast number of research papers have been published in this area. The goal of this book is to distill and integrate these research results into cohesive and comprehensible methodologies, and to provide a streamlined approach to time series analysis and forecasting.

The use of computers and computer software is essential in any modern quantitative analysis, even more so in time series analysis where complex algorithms and extensive computations are often required. With the speed and capacity of modern computers, in many situations it is preferable to adopt a methodology that simplifies the means of conducting an analysis even if it is at the expense of computation time. Using such an approach, we are able to provide simplified and effective methodologies for complex subjects in time series analysis

Forecasting techniques range from simple to complex, and include the use of executive judgment, surveys, time-series analysis, correlation methods, market tests and simulation (Smith III et al. 1996). Wilson and Keating (2007) defined time series forecasting as the use of a technique to forecast future events based on known past events. Time Series Analysis explicitly recognizes the order in which experimental data are observed, as well as the statistical dependence of observed data (Abdelhamid and Everett 1999). The advantage of the time series method compared to other predictive methods, such as regression and neural networks, is that the other methods require future values of input variables that are not readily available, whereas, univariate time series analysis has been identified as the most effective way of carrying out work to develop predictive models for construction costs (Ashuri and Lu 2010). Therefore, univariate time series analysis was chosen in this paper as the research methodology due to its requirement of just one input variable for creating and calibrating models. Brockwell and Davis (2002) pointed out that time series analysis follows a standard procedure in sequence: examine the main features of data series; check dependency in data; choose a model to fit the series; diagnose the constructed model; and forecast and update data. Four major factors affect time series analysis: trend; seasonality; cycles; and irregular movements, whereas, seasonal component is concerned with the periodic fluctuation in the series within each year (Farnum and Stanton 1989). Simple Moving-Average (SMA), Holt Exponential Smoothing (Holt ES), Holt-Winters Exponential Smoothing (Holt-Winters ES), Auto-Regressive Integrated Moving-Average (ARIMA), and Seasonal ARIMA are some of the well-known time series approaches (Ashuri and Lu 2010). The SMA is properly employed when there is no trend or seasonality present in the data (Wilson and Keating 2007). The Holt ES method is recommended to handle time series data that display trends (Brockwell and Davis 2002). A third-factor called seasonal smoothing is presented into time series analysis and is an estimated value of seasonal growth rate reflecting the seasonal pattern of time series data (Wilson and Keating 2007). ARIMA is built upon Auto-Regressive (AR) and Moving-Average (MA) approaches. ARIMA is recommended to model time series data displaying nonstationary behaviours (Box and Jenkins 1970).

2. RESEARCH METHODOLOGY AND RESULTS

The term univariate time series refers to a time series that consists of single (scalar) observations recorded sequentially over equal time increments. Although a univariate time series data set is usually given as a single column of numbers, time is in fact an implicit variable in the time series. If the data are equi-spaced, the time variable, or index, does not need to be explicitly given. The time variable may sometimes be explicitly used for plotting the series. However, it is not used in the time series model itself.

2.1 Stationarity

A common assumption in many time series techniques is that the data are stationary. A stationary process has the property that the mean, variance and autocorrelation structure do not change over time. Stationarity can be defined in precise mathematical terms, but for our purpose we mean a flat looking series, without trend, constant variance over time, a constant autocorrelation structure over time and no periodic fluctuations.

For a formal definition of the concept, consider a stochastic process $\{X(t),\ t\in T\}$, where t represents time and belongs to a linear index set T. An index set T is said to be linear if for any *t* and *h* belonging to *T*, there sum $t+h$ also belongs to *T*. The index set *T* could be discrete or continuous, one-sided or two-sided. Examples of index sets for a discrete time stochastic process and for a continuous time process are given by $T=\{1,2,\}$ and $T=\{t,\ t\geq 0\}$, respectively. We can define stationary indifferent ways such as Strict stationary, weak stationary.

2.2 Transformations to Achieve Stationarity

If the time series is not stationary, we can often transform it to stationarity with one of the following techniques.

1. We can difference the data. That is, given the series Z_t, we create the new series
   $$Y_i = Z_i - Z_{i-1}$$
   The differenced data will contain one less point than the original data. Although you can difference the data more than once, one differene is usually sufficient.
2. If the data contain a trend, we can fit some type of curve to the data and then model the residuals from that fit. Since the purpose of the fit is to simply remove long term trend, a simple fit, such as a straight line, is typically used.
3. For non-constant variance, taking the logarithm or square root of the series may stabilize the variance. For negative data, you can add a suitable constant to make all the data positive before applying the transformation. This constant can then be subtracted from the model to obtain predicted (i.e., the fitted) values and forecasts for future points.

The above techniques are intended to generate series with constant location and scale. Although seasonality also violates stationarity, this is usually explicitly incorporated into the time series model.

2.3 Stationarity Check

A number of tests are proposed to test the presence of unit root in series But here we use Augmented Dickey Fuller Test.

2.3.1 Augmented Dickey Fuller Test

The basic assumption about the error term in the Dickey Fuller Test is that the error term is white noise i.e. have no autocorrelation. Due to the autocorrelation we cannot apply this test to test the hypothesis of unit root. In order to deal the problem of existence of serial correlation in the residual term it is proposed that lags of dependent are included as additional regressors in the model. Therefore by adding the lagged depended variable in the equation we estimate equation termed as Augmented dickey Fuller test (adf) regression. Before applying ADF the problem of autocorrelation is tested by Breusch – Godfrey Serial Correlation LM test. The choice of appropriate lag length is likely depend on the frequency of the data & autocorrelation of model. Different sequential rules for including lags of dependent variable in the regression. These include general to specific rule and specific to general.

2.4 Seasonality

Many time series display seasonality. By seasonality, we mean periodic fluctuations. For example, retail sales tend to peak for the Christmas season and then decline after the holidays. So time series of retail sales will typically show increasing sales from September through December and declining sales in January and February.

He following graphical techniques can be used to detect seasonality.

1. A run sequence plot will often show seasonality.
2. A seasonal subseries plot is a specialized technique for showing seasonality.
3. Multiple box plots can be used as an alternative to the seasonal subseries plot to detect seasonality.

The autocorrelation plot can help identify seasonality

2.5 Power Transformations and Forecasting

The efficacy of statistical models is often enhanced through the use of data transformation, analysis and forecasting using time series models are no exception. With an appropriate transformation on a time series, the model for the series may be simplified; the intervention effects may be better estimated; and the forecasts of future values may be improved. Most of the statistical methods assume that the variables are normally distributed. A data transformation is a useful tool to achieve Normality for the variables under study. However the mathematical modification of the data in this manner raises issues not only for the interpretation of the modeling results, but also the usefulness of forecasts based on the transformed data. In this chapter, we are particularly interested in the application of power transformations to improve forecasting accuracy when forecasts are retransformed back into original metric.

There are two primary issues in the application of power transformation. The first is the selection of an appropriate lambda value that will either improve the efficacy of the model or improve the accuracy of the forecasts. The second issue involves the correction of biases induced by the forecasts of the transformed series.

3. RESULTS AND DISCUSSION

The collected data is the Whole sale price of raw material from Jan 1980 to Oct 2014. In raw material we have cotton, tobacco, wool and skins.

I had collected the data from Statistical Bulletin by the State Bank of Pakistan and website of Federal Bureau of Statistics.

First of all we check what type of model we run i.e. include intercept term or not, include trend component in the model or not. First we run equation without intercept & trend.

$$\Delta Y_t = \rho Y_{t-1} + V_t$$

where Δ is first difference of the Y_t, t is time trend and Y_{t-1} is the lag of Y_t. V_t is the error term.

Table 1
Model 1

Variable	**Coefficient**	**Std. Error**	**t-Statistic**	**Prob.**
$Y(-1)$	0.001150	0.001153	0.997440	0.3289
R-squared	-0.001621	Mean dependent var		0.004614
Adjusted R-squared	-0.001621	S.D. dependent var		0.022232
S.E. of regression	0.022250	Akaike info criterion		-4.732144
Sum squared resid	0.011387	Schwarz criterion		-4.683058
Log likelihood	57.78573	Durbin-Watson stat		2.076366

As table show that the coefficient of Y_{t-1} is positive, so we have to include intercept in our model and estimate the equation by OLS

Now the equation is

$$\Delta\, Y_t = u + rY_{t-1} + V_t$$

Table 2
Model 2

Variable	**Coefficient**	**Std. Error**	**t-Statistic**	**Prob.**
C	0.367126	0.295913	1.240653	0.2278
$Y(-1)$	-0.092073	0.075150	-1.225201	0.2335
R-squared	0.063874	Mean dependent var		0.004614
Adjusted R-squared	0.021323	S.D. dependent var		0.022232
S.E. of regression	0.021994	Akaike info criterion		-4.716436
Sum squared resid	0.010642	Schwarz criterion		-4.618265
Log likelihood	58.59723	F-statistic		1.501117
Durbin-Watson stat	2.024179	Prob(F-statistic)		0.233456

As table show that the coefficient of Y_{t-1} is positive, so we have to include the intercept term. Before applying the unit root test we check that errors are white noise or not. For this we use Breusch-Godfrey Serial Correlation LM Test:

Ho: The residual term is white noise.
H1: The residual term is not white noise.

Table 3
Breusch-Godfrey Serial Correlation LM Test

F-statistic	0.006000	Probability		0.938990
Obs*R-squared	0.006855	Probability		0.934013
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.008854	0.323688	-0.027354	0.9784
Y(-1)	0.002249	0.082205	0.027358	0.9784
RESID(-1)	-0.018065	0.233220	-0.077461	0.9390
R-squared	0.000286	Mean dependent var		2.49E-17
Adjusted R-squared	-0.094925	S.D. dependent var		0.021511
S.E. of regression	0.022508	Akaike info criterion		-4.633388
Sum squared resid	0.010639	Schwarz criterion		-4.486132
Log likelihood	58.60066	F-statistic		0.003000
Durbin-Watson stat	1.986237	Prob(F-statistic)		0.997005

Table show that p-value associated with F-statistic is greater than α so we accept Ho and conclude that error term is white noise.

3.2 Testing For Unit Root

Now we apply ADF test to check the presence or absence of unit root.

Ho: Unit root exist.
H1: Unit root not exist.

Table 4
Unit Root

ADF Test Statistic	-1.225210	1% Critical Value*		-3.7497
		5% Critical Value		-2.9969
		10% Critical Value		-2.6381
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	-0.100166	0.081755	-1.225210	0.2347
D(Y(-1))	-0.007726	0.217170	-0.035575	0.9720
C	0.399438	0.322056	1.240277	0.2292
R-squared	0.071278	Mean dependent var		0.004708
Adjusted R-squared	-0.021594	S.D. dependent var		0.022727
S.E. of regression	0.022971	Akaike info criterion		-4.588045
Sum squared resid	0.010554	Schwarz criterion		-4.439937
Log likelihood	55.76252	F-statistic		0.767490
Durbin-Watson stat	2.005958	Prob(F-statistic)		0.477370

The value of ADF test statistic is less than the critical values in absolute terms so we conclude the presence of unit root & even after taking care of possible autocorrelation in error term, the series is non-stationary.

3.3 Testing for Stationarity on First Difference

We run the equation

$$\Delta Y_t = u + \rho \Delta Y_{t-1} + V_t$$

Before applying the ADF we test that regression is sufficient means errors are white noise. So we apply LM test

3.4 White Noise of the Error Term at First Difference

Ho: The residual term is white noise.
H1: The residual term is not white noise.

We use serial correlation LM test to check white noise of error term

Table 5
Breusch-Godfrey Serial Correlation LM Test

F-statistic	0.331347	Probability		0.565309
Obs*R-squared	0.334362	Probability		0.563102
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.032752	0.056942	0.575188	0.5656
D(Y(-1))	-5.797574	10.07191	-0.575618	0.5653
RESID(-1)	5.797946	10.07239	0.575628	0.5653
R-squared	0.001130	Mean dependent var		-9.38E-19
Adjusted R-squared	-0.005689	S.D. dependent var		0.037688
S.E. of regression	0.037795	Akaike info criterion		-3.703179
Sum squared resid	0.418547	Schwarz criterion		-3.665777
Log likelihood	551.0705	F-statistic		0.165674
Durbin-Watson stat	1.982468	Prob(F-statistic)		0.847402

As we see from the table that p-value associated with F-statistic is greater than α so we accept Ho and conclude that error term is white noise at first difference.

3.5 Testing for Unit Root at First Difference

Now we apply ADF test to check the presence or absence of unit root.

Ho: Unit root exist
H1: Unit root not exist.

Table 6
ADF Test

ADF Test Statistic	-11.05083	1% Critical Value*		-3.4543
		5% Critical Value		-2.8715
		10% Critical Value		-2.5720
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(Y(-1))	-0.926716	0.083859	-11.05083	0.0000
D(Y(-1),2)	-0.075442	0.059009	-1.278486	0.2021
C	0.005227	0.002257	2.316622	0.0212
R-squared	0.500005	Mean dependent var		-0.000274
Adjusted R-squared	0.496581	S.D. dependent var		0.053240
S.E. of regression	0.037775	Akaike info criterion		-3.704214
Sum squared resid	0.416672	Schwarz criterion		-3.666720
Log likelihood	549.3716	F-statistic		146.0031
Durbin-Watson stat	1.982733	Prob(F-statistic)		0.000000

The value of ADF test statistic is greater than the critical values in absolute terms so we reject the presence of unit root; hence the series is stationary at first difference

3.6 Seasonality Checking

We check seasonality by using HEGY test

First we run following regression equation

$$y_{8,t} = DS + DT + \pi_1\, y_{1,t-1} + \pi_2\, y_{2,,t-1} + \pi_3\, y_{3,t-1} + \pi_4\, y_{3,t-1} + \pi_5\, y_{4,t-1} + \pi_6\, y_{4,t-2} \\ + \pi_7\, y_{5,t-1} + \pi_8\, y_{5,t-2} + \pi_9\, y_{6,t-1} + \pi_{10}\, y_{6,t-1} + \pi_{11}\, y_{7,t-1} + \pi_{12}\, y_{7,t-1} + u_t$$

where DS and DT are deterministic season and trend. And the y_i's are defined as:

$$y_1 = y + y(-1) + y(-2) + y(-3) + y(-4) + y(-5) \\ + y(-6) + y(-7) + y(-8) + y(-9) + y(-10) + y(-11)$$

$$y_2 = -y + y(-1) - y(-2) + y(-3) - y(-4) + y(-5) \\ - y(-6) + y(-7) - y(-8) + y(-9) - y(-10) + y(-11)$$

$$y_3 = -y + y(-2) - y(-4) + y(-6) - y(-8) + y(-10)$$

$$y_4 = -y + 1.732 * y(-1) - 2 * y(-2) + 1.732 * y(-3) - y(-4) + y(-6) \\ - 1.732 * y(-7) + 2 * y(-8) - 1.732 * y(-9) + y(-10)$$

$$y_5 = -y - 1.732*y(-1) - 2 * y(-2) - 1.732 * y(-3) - y(-4) + y(-6) \\ + 1.732 * y(-7) + 2 * y(-8) + 1.732 * y(-9) + y(-10)$$

$$y_6 = -y + y(-1) - y(-3) + y(-4) - y(-6) + y(-7) - y(-9) + y(-10)$$

$$y_7 = -y - y(-1) + y(-3) + y(-4) - y(-6) - y(-7) + y(-9) + y(-10)$$

$$y_8 = y - y(-12)$$

Before checking existence of unit root we check whether the error term is white noise or not. As we have mentioned earlier that for this purpose we use three tests. Here our hypothesis is

Ho : ρ_k are all zero.
H1 : Some ρ_k are non-zero

Table 7
Seasonality

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.017251	0.021127	0.816527	0.4150
Y1(-1)	-6.66E-06	0.000335	-0.019893	0.9841
Y2(-1)	-0.134716	0.036387	-3.702326	0.0003
Y3(-2)	-0.153759	0.035523	-4.328457	0.0000
Y4(-1)	-0.341362	0.050737	-6.728138	0.0000
Y4(-2)	-0.363007	0.050684	-7.162170	0.0000
Y5(-1)	-0.009117	0.013976	-0.652336	0.5148
Y5(-2)	-0.025860	0.013474	-1.919261	0.0561
Y6(-1)	-0.309863	0.048172	-6.432439	0.0000
Y6(-2)	-0.283650	0.048099	-5.897150	0.0000
Y7(-1)	-0.153875	0.031623	-4.865902	0.0000
Y7(-2)	-0.103893	0.031584	-3.289444	0.0011
@SEAS(1)	-0.002373	0.011879	-0.199750	0.8418
@SEAS(2)	-0.012651	0.011782	-1.073787	0.2839
@SEAS(3)	-0.001462	0.011972	-0.122146	0.9029
@SEAS(4)	-0.002501	0.012098	-0.206750	0.8364
@SEAS(5)	0.006424	0.011865	0.541442	0.5887
@SEAS(6)	0.010878	0.012050	0.902801	0.3675
@SEAS(7)	-0.029863	0.011788	-2.533392	0.5119
@SEAS(8)	-0.005178	0.012171	-0.425467	0.6709
@SEAS(9)	-0.025267	0.011963	-2.112155	0.5356
@SEAS(10)	-0.006974	0.011859	-0.588080	0.5570
@SEAS(11)	0.001508	0.012036	0.125321	0.9004
R-squared	0.892496	Mean dependent var		0.080662
Adjusted R-squared	0.883184	S.D. dependent var		0.113845
S.E. of regression	0.038910	Akaike info criterion		-3.575737
Sum squared resid	0.384558	Schwarz criterion		-3.274826
Log likelihood	518.2396	F-statistic		95.84988
Durbin-Watson stat	1.829727	Prob(F-statistic)		0.000000

Box-Pierce Q-Statistic

Q statistics = 23.35775 < 50.46926 is χ^2 at DF 36

Therefore we accept Ho that all ρ_k are zero i.e. error terms are white noise.

Breusch-Godfrey Serial Correlation LM test

LM TEST FOR LAG 1

Table 8
Breusch-Godfrey Serial Correlation LM Test for Lag 1

F-statistic	0.758920	Probability	0.384496
Obs*R-squared	0.831705	Probability	0.361780

In Table, P-value = 0.384496 of F-statistic is greater than 0.05, also value of R-squared = 0.831705 is less than χ^2 = 3.84146 at 1 DF. Hence from here we accept our Ho that there is no serial correlation.

Table 9
LM Test for LAG 12

F-statistic	0.275011	Probability	0.96055
Obs*R-squared	3.03306	Probability	0.904738

In Table, P-value 0.96055 of F-Statistic is greater than 0.05 and value of R-squared = 3.033 is less than χ^2 = 21.0261 at 12 DF therefore we accept Ho and conclude that. there is no serial correlation.

After making the series of error term white noise our next step is to check whether unit root exist in the data or not.

- Ho : $\pi_1 = 0$
  H_1 : $\pi_1 \neq 0$

Since the calculated value of wald statisticis 7.19 which is greater is it's tabulated value which is τ = **4.18** that is it lies in the rejection region therefore we reject Ho. It implies that there is no unit root in the data.

- Ho : $\pi_2 = \pi_2 = \pi_3 = \ldots = \pi_{11} = 0$
  H_1 : at least one of πi's is non-zero

F-statistic	49.960	Probability	0.000000

Here calculated value of Wald F-Test is 49.960 which is greater than it's tabulated value which is 4.45 therefore we reject Ho and conclude that there is no seasonal unit root. Thus the series becomes stationary after taking one times first difference that is series is I(1).

3.7 Model Fitting

After making the series stationary our next goal is to fit a model on the series for the purpose of forecasting. It involves following steps.

Identification:

To get idea of the model whether it is AR or MA or the mixture of both we first examine ACF and PACF of the differenced series. From the ACF and PACF of differenced series we come to know that our model is ARMA. From ACF it is clear that non-seasonal order of MA should be 1 and there is no seasonality in the ACF since no spike is significant at 12^{th} and 24^{th} lag. Similarly from PACF it is clear that non-seasonal order of AR should be 2 and seasonal order should be 0.

Fitted Model

Best fitted model on the differenced series is as: ARIMA(2,1,2). Since we have only one model through which series of error term becomes white noise.

ARIMA Model: (2,1,2)
$X(t) = 1.180\ X(t-1) - .9146\ X(t-2) + Z(t) - 1.182\ Z(t-1) + .9997\ Z(t-2)$

WN Variance = .001103

AR Coefficients
1.179924 –.914567

Standard Error of AR Coefficients
.025630 .025630

MA Coefficients
–1.181880 .999734

Standard Error of MA Coefficients
.001462 .001462

(Residual SS) / N = .00110251

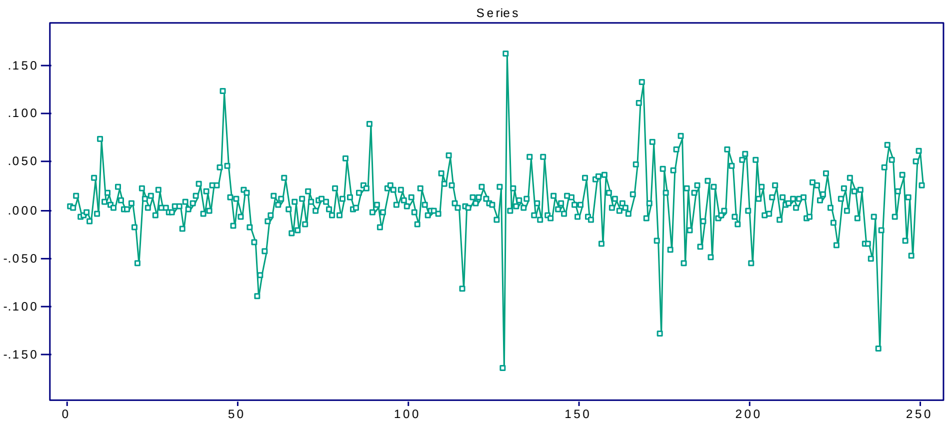
AICC = –.982270E + 03
BIC = –.987212E + 03

–2Log (Likelihood) = –.992515E + 03

Accuracy parameter = .100000E – 08

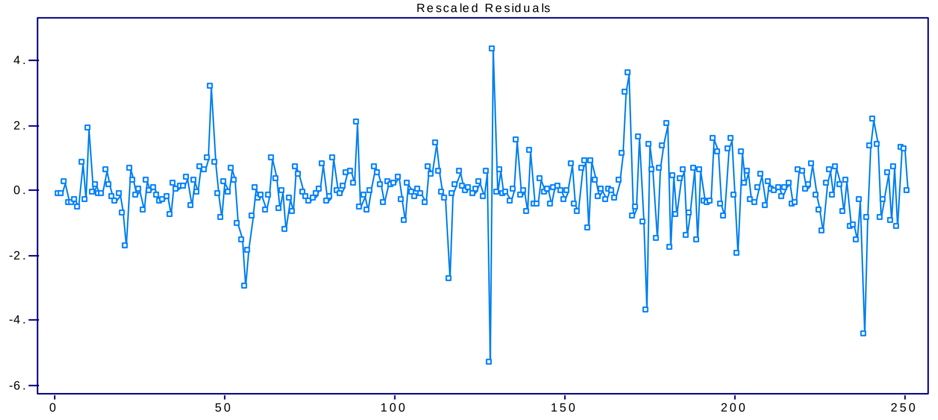
Number of iterations = 1

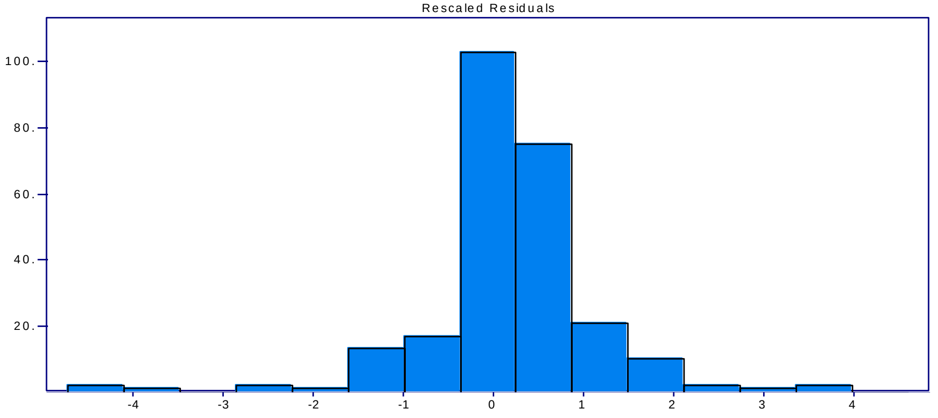
Number of function evaluations = 15494

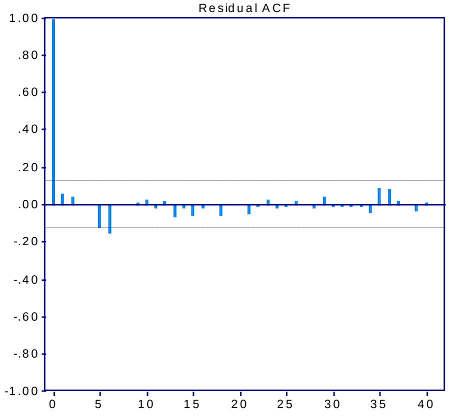


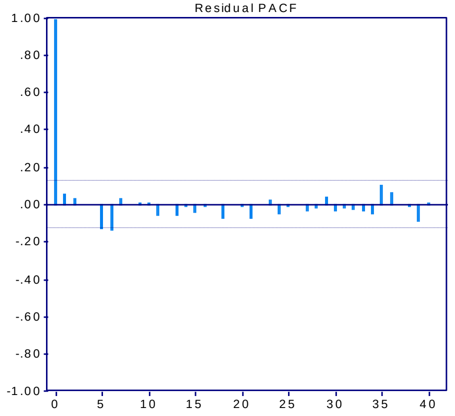
3.8 Diagnostic Checking

Since calculated values of chi-square at lags 12, 24, 36, 48 are all insignificant and AC's and PAC's of error terms are all within limits so we conclude that error terms are uncorrelated that the series of error terms is white noise.



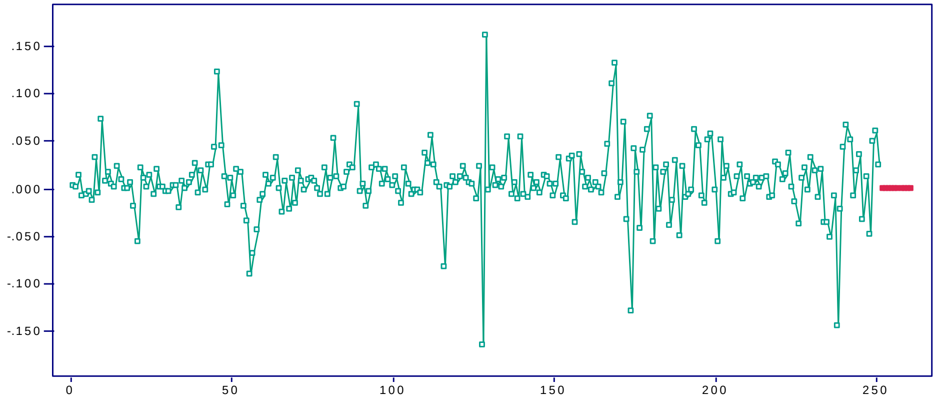






3.9 Forecasting

For the purpose of forecasting first of all we take inverse differencing and then forecast observation. In this manner we can transform the forecasted values to the original values. since the forecasted values through the fitted model is close to the original values so we can conclude that our fitted model is a good model.



Approximate 95 Percent		Prediction Bounds		
Step	Prediction	sqrt(MSE)	Lower	Upper
1	.00000	.03499	-.06857	.06857
2	.00000	.03499	-.06857	.06857
3	.00000	.03499	-.06857	.06857
4	.00000	.03499	-.06857	.06857
5	.00000	.03499	-.06857	.06857
6	.00000	.03499	-.06857	.06857
7	.00000	.03499	-.06857	.06857
8	.00000	.03499	-.06857	.06857
9	.00000	.03499	-.06857	.06857

CONCLUSION

The accurate estimation of concrete material prices is an essential practice, especially success and even viability. Factors affecting whole sale prices of raw material have been related in previous literatures to cost of the production process, raw material prices, and industry related factors. A Time Series Analysis was conducted on the material components of, in a univariate manner because other predictive techniques require future input variables that are not readily available to estimators. The analysis was differentiated for a stable economic period and a growth economic period. The outcomes of the Time Series model can be different in both periods and thus caution should be taken in the times of economic change or status switching. The results were verified using two distinct software for the Time Series Analysis. A reasonable degree of prediction accuracy was concluded for all materials. Time Series Analysis in general, although being a good prediction technique in stable economic and industry conditions, cannot predict sudden macroeconomic or other events.

REFERENCES

1. Abdelhamid, T.S. and Everett, J.G. (1999). Time series analysis for construction productivity experiments. *Journal of Construction Engineering and Management*, 125(2), 87-95.
2. Ashuri, B. and Lu, J. (2010). Time series analysis of ENR construction cost index. *Journal of Construction Engineering and Management*, 136(11), 1227-1237.
3. Bade, R. and Parkin, M. (2007). *Foundations of Macroeconomics*, Third Edition, Pearson Education, Inc., USA.
4. Box, G. and Jenkins, G. (1970). *Time series analysis: Forecasting and control*, San Francisco: Holden-Day.
5. Brockwell, P.J. and Davis, R.A. (2002). *Introduction to time series and forecasting*, second edition,
6. Farnum, N.R. and Stanton, L.W. (1989). *Quantitative Forecasting Methods*, first edition, PWS-KENT Publishing Company, Boston, USA.
7. Ghosh, S. and Jintanapakanont, J. (2004). Identifying and assessing the critical risk factors in an underground rail project in Thailand: a factor analysis approach. *International Journal of Project Management*, 22(8), 633-643.
8. Hagimura, S., Saitoh, T. and Yagihara, Y. (1988). Application of time series analysis and modern control theory to the cement plant. *Annals of the Institute of Statistical Mathematics*, 40(3), 419-438.
9. Brown, M.B. and Forsythe, A.B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69(346), 364-367.
10. Smith, H.C., Herbig, P., Milewicz, J. and Golden, J.E. (1996). Differences in forecasting behaviour between large and small firms. *Journal of Marketing Practice: Applied Marketing Science*, 2(1), 35-51.
11. Wilson, J.H. and Keating, B. (2007). *Business Forecasting with Accompanying Excel-Based ForecastX TM Software*, Fifth Edition, McGraw-Hill International Edition, Irwin.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 35-42

A RATIO TYPE ESTIMATOR FOR THE ESTIMATION OF UNKNOWN POPULATION VARIANCE USING AUXILIARY VARIABLE

Hira Asif and **Naila Amjad**
Lahore College for Women University, Lahore, Pakistan.
Email: hiramalik5551@gmail.com
nailaamjad2007@gmail.com

ABSTRACT

In this paper, we proposed a ratio type estimator for estimation of unknown population variance using mean and correlation coefficient along with third quartile of auxiliary variable. Expressions of bias and mean square error have been formed up to first order approximation. Further we derived the condition under which proposed estimator performs better than traditional ratio type variance estimator and existing ratio type variance estimators. Empirical study is carried out with the help of two natural populations to determine the results of proposed estimator over than existing estimators.

1. INTRODUCTION

Auxiliary information is being utilized by many investigators for improving the precision of estimate of population parameter. If prior information is available then we can use certain estimators such as ratio, product and linear regression estimators. Otherwise simplest estimator of population variance can be acquired by using simple random sampling without replacement. Auxiliary information is helpful not only at the stage of designing (for stratified, systematic or probability proportional to size sampling designs) but also in estimation stage. In sample surveys, it also helps in improving the sampling design for higher accuracy in the estimate of population parameter.

Ratio method is applied, when the study variable Y and auxiliary variable X has positive correlation between them and the line of regression of study variable Y on auxiliary variable X passes through origin. Product method is valuable in the study when variable Y and auxiliary variable X are negatively correlated with each other, while regression method is applicable when with linear relationship regression line does not pass through origin.

Consider a finite population $U = \{U_1, U_2, \ldots, U_N\}$ of N size. Let (Y, X) be the study variable and auxiliary variable with values (y_i, x_i) measured on U_i where i=1,2,…N given a vector $Y = \{Y_1, Y_2, \ldots Y_N\}$. The issue is to estimate population variance on random sample of n size, chosen from population U.

Estimating the finite population variance has great significance in various fields such as Industry, Agriculture, Medical and Biological Sciences. For example in the context of health; variations in body temperature, pulse beat and blood pressure are the basic guidelines to diagnosis where prescribed treatment is designed to control their variation (Subramani and Kumarapandiyan, 2012).

2. SOME EXISTING ESTIMATORS

Isaki (1983) developed a ratio type variance estimator for the population variance with auxiliary information

$$\mathbf{\hat{S}_R^2 = s_y^2 \frac{S_x^2}{s_x^2}} \tag{2.1}$$

$$B\left(\hat{S}_R^2\right) = \lambda S_y^2\left[\left(\beta_{2(x)} - 1\right) - (\lambda_{22} - 1)\right] \tag{2.1.1}$$

$$MSE\left(\hat{S}_R^2\right) = \lambda S_y^4\left[\left(\beta_{2(y)} - 1\right) + \left(\beta_{2(x)} - 1\right) - 2(\lambda_{22} - 1)\right] \tag{2.1.2}$$

where

$$\beta_{2(y)} = \frac{\mu_{40}}{\mu_{20}^2}, \beta_{2(x)} = \frac{\mu_{04}}{\mu_{02}^2}, \lambda_{22} = \frac{\mu_{22}}{\mu_{20}\mu_{02}}$$

and

$$\mu_{rs} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^r\,(x_i - \bar{X})^s.$$

Khan and Shabbir (2013) developed the ratio kind estimator on behalf of the population variance by quartiles of an ancillary variable

$$S_M^2 = s_y^2\left[\frac{S_x^2\rho + Q_3}{s_x^2\rho + Q_3}\right] \tag{2.2}$$

$$B(S_M^2) = \gamma S_y^2 R_P\left[R_P\left(\beta_{2(x)} - 1\right) - (\lambda_{22} - 1)\right] \tag{2.2.1}$$

$$MSE(S_M^2) = \gamma S_y^4\left[\left(\beta_{2(y)} - 1\right) + R_P^2\left(\beta_{2(x)} - 1\right) - 2R_P(\lambda_{22} - 1)\right] \tag{2.2.2}$$

where

$$R_P = \frac{S_x^2\rho}{S_x^2\rho + Q_3}, \beta_{2(y)} = \frac{\mu_{40}}{\mu_{20}^2}, \beta_{2(x)} = \frac{\mu_{04}}{\mu_{02}^2}$$

$$\lambda_{22} = \frac{\mu_{22}}{\mu_{20}\mu_{02}}$$

and

$$\mu_{rs} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^r\,(x_i - \bar{X})^s.$$

Upadhyaya and Singh (2001) used population mean as auxiliary variable for modified ratio type variance estimator

$$S_U^2 = s_y^2\left[\frac{\bar{X}}{\bar{x}}\right] \tag{2.3}$$

$$B(S_U^2) = \gamma S_y^2[C_x^2 - \lambda_{21}C_X] \tag{2.3.1}$$

$$MSE(S_U^2) = \gamma S_y^4\left[\left(\beta_{2(y)} - 1\right) + C_x^2 - 2\lambda_{21}C_X\right] \tag{2.3.2}$$

where

$$\lambda_{21} = \frac{\mu_{21}}{\mu_{20}\sqrt{\mu_{02}}}.$$

Subramani and Kumarapandiyan developed a class of modified ratio type variance estimators for estimating population variance

$$S_j^2 = s_y^2\left[\frac{\bar{X} + \omega_i}{\bar{x} + \omega_i}\right] \tag{2.4}$$

where ω_i are well-known parameters of supporting variables like kurtosis, skewness, measurement of variation, median, deciles, quartiles, association, also standard deviation etc.

$$B\left(S_j^2\right) = \gamma S_y^2\left[\theta_{pi}^2 C_x^2 - \theta_{pi}\lambda_{21}C_X\right] \tag{2.4.1}$$

$$MSE\left(S_j^2\right) = \gamma S_y^4\left[\left(\beta_{2(y)} - 1\right) + \theta_{pi}^2 C_x^2 - 2\theta_{pi}\lambda_{21}C_X\right] \tag{2.4.2}$$

where

$$\lambda_{21} = \frac{\mu_{21}}{\mu_{20}\sqrt{\mu_{02}}}, \qquad \theta_{pi} = \frac{\bar{X}}{\bar{X} + \omega_i}.$$

3. PROPOSED ESTIMATOR

We proposed a ratio type estimator for estimation of unknown population variance. The expressions of bias and mean square error have been formed up to first order approximation with constant is given by

$$S_P^2 = s_y^2\left[\frac{\bar{X}\rho + Q_3}{\bar{x}\rho + Q_3}\right] \tag{3.1}$$

$$B(S_P^2) = \gamma S_y^2(R^2C_x^2 - R\lambda_{21}C_x) \tag{3.1.1}$$

$$MSE(S_P^2) = S_y^4\gamma\left[\left(\beta_{2(y)} - 1\right) + R^2C_x^2 - 2R\lambda_{21}C_x\right] \tag{3.1.2}$$

where

$$R = \frac{\bar{X}\rho}{\bar{X}\rho + Q_3}.$$

4. EFFICIENCY OF PROPOSED ESTIMATOR

Below is the condition under which proposed estimator performing better than existing estimators.

The mean square error of proposed estimator is

$$\mathrm{MSE}(S_P^2) = \gamma S_y^4\left[\left(\beta_{2(y)} - 1\right) + R^2 C_x^2 - 2R\lambda_{21} C_X\right] \tag{4.1}$$

where

$$R = \frac{\bar{X}\rho}{\bar{X}\rho + Q_3}.$$

The mean square error of usual ratio type variance estimator

$$\mathrm{MSE}(S_R^2) = \gamma S_y^4\left[\left(\beta_{2(y)} - 1\right) + \left(\beta_{2(x)} - 1\right) - 2(\lambda_{22} - 1)\right] \tag{4.2}$$

The mean square error of ratio type variance estimator by Khan and Shabbir, (2013)

$$\mathrm{MSE}(S_M^2) = \gamma S_y^4\left[\left(\beta_{2(y)} - 1\right) + R_P^2\left(\beta_{2(x)} - 1\right) - 2R_P(\lambda_{22} - 1)\right] \tag{4.3}$$

where

$$R_P = \frac{S_x^2\rho}{S_x^2\rho + Q_3}.$$

The mean square error of modified ratio type variance estimator by Upadhaya and Singh (2001)

$$\mathrm{MSE}(S_U^2) = \gamma S_y^4\left[\left(\beta_{2(y)} - 1\right) + C_x^2 - 2\lambda_{21} C_X\right] \tag{4.4}$$

The mean square error of modified ratio type variance estimator of population variance by Subramani and Kumarapandiyan (2015).

$$\mathrm{MSE}\left(S_j^2\right) = \gamma S_y^4\left[\left(\beta_{2(y)} - 1\right) + \theta_{pi}^2 C_x^2 - 2\theta_{pi}\lambda_{21} C_X\right] \tag{4.5}$$

where

$$\theta_{pi} = \frac{\overline{X}}{\overline{X} + \omega_i}$$

From expression (4.1) and (4.2)

$$MSE(S_P^2) \leq MSE(S_R^2)\ if\ R = \frac{\lambda_{21} + \left[\left(\beta_{2(x)} - 1\right) - 2(\lambda_{22} - 1) + \lambda_{21}^2\right]^{\frac{1}{2}}}{C_x}$$

From expression (4.1) and (4.3)

$$MSE(S_P^2) \leq MSE(S_M^2)\ if\ R \leq \frac{\lambda_{21} + \left[\left(\beta_{2(x)} - 1\right) - 2(\lambda_{22} - 1) + \lambda_{21}^2\right]^{1/2}}{C_x}$$

From expression (4.1) and (4.4)

$$MSE(S_P^2) \leq MSE(S_U^2)\ either\ \mathrm{R} \leq \frac{\lambda_{21} + \left[R_P^2\left(\beta_{2(x)} - 1\right) - 2R_P(\lambda_{22} - 1) + \lambda_{21}^2\right]^{1/2}}{C_x}$$

From expression (4.1) and (4.5)

$$MSE(S_P^2) \leq MSE\left(S_j^2\right)\ either\ 2\frac{\lambda_{21}}{C_x} - \theta_{pi} \leq R_{PI} \leq \theta_{pi}\ or\ \theta_{pi} \leq R_{PI} \leq 2\frac{\lambda_{21}}{C_x} - \theta_{pi}$$

5. EMPIRICAL STUDY

Following two data sets are taken from Singh and Chaudhary (1986) to compare the efficiency of proposed estimator with existing estimators.

Populations	**Population 1**	**Population 2**
N	34	34
n	20	20
ρ	0.4491	0.4453
$\bar{Y}$	85.6412	85.6412
$\bar{X}$	20.8884	19.9441
$\beta_{2(X)}$	2.9123	3.7257
C_Y	0.8561	0.8561
C_X	0.7205	0.7532
S_Y	73.3141	73.3141
S_X	15.0506	15.0215
$\beta_{2(Y)}$	13.3667	3.3667
Q_3	25.475	27.8
λ_{21}	-0.3105	-0.2946
λ_{22}	1.1525	1.2244

Bias and MSE of Two Populations

Notation	Estimators	Pop 1 Bias	Pop 2 Bias	Pop 1 MSE	Pop 2 MSE
$\mathrm{S}_{\mathrm{R}}^2$	$\mathrm{S}_{\mathrm{R}}^2 = \mathrm{s}_{\mathrm{y}}^2 \frac{\mathrm{S}_{\mathrm{x}}^2}{\mathrm{s}_{\mathrm{x}}^2}$	9180.889	13048.89	3918393	4106119
S_M^2	$S_M^2 = s_y^2 \left[\frac{S_x^2\,\rho + Q_3}{s_x^2\,\rho + Q_3}\right]$	5744.472	7807.26	3742249	3838021
$\mathrm{S}_{\mathrm{U}}^2$	$\mathrm{S}_{\mathrm{U}}^2 = \mathrm{s}_{\mathrm{y}}^2 \left[\frac{\bar{\mathrm{X}}}{\bar{\mathrm{x}}}\right]$	3875.273	4117.17	3738698	3751189
$\mathrm{S}_{\mathrm{j}}^2$	$S_j^2 = s_y^2 \left[\frac{\bar{X} + Q_1}{\bar{x} + Q_1}\right]$	44.336	44.39	7685669	7682366
$\mathrm{S}_{\mathrm{j}}^2$	$S_j^2 = s_y^2 \left[\frac{\bar{X} + Q_r}{\bar{x} + Q_r}\right]$	32.370	30.41	7604903	7588714
$\mathrm{S}_{\mathrm{j}}^2$	$S_j^2 = s_y^2 \left[\frac{\bar{X} + Q_d}{\bar{x} + Q_d}\right]$	47.868	46.89	7709091	7698856
$\mathrm{S}_{\mathrm{j}}^2$	$S_j^2 = s_y^2 \left[\frac{\bar{X} + Q_a}{\bar{x} + Q_a}\right]$	30.541	29.20	7592328	7580457
$\mathrm{S}_{\mathrm{j}}^2$	$S_j^2 = s_y^2 \left[\frac{\bar{X} + Q_3}{\bar{x} + Q_3}\right]$	22.814	21.21	7538246	7524817
$\underline{\mathrm{S}_{\mathrm{P}}^2}$	$S_P^2 = s_y^2 \left[\frac{\bar{X}\,\rho + Q_3}{\bar{x}\,\rho + Q_3}\right]$	*10.824*	*9.63*	*7449660*	*7439361*

6. CONCLUSION

Survey Statisticians always look for of competent estimators. The basic objective of this estimator is to provide the efficient estimator with auxiliary information. In the current study auxiliary variable is utilized for the estimation of population variance. Auxiliary information is being utilized by many investigators for improving the precision of estimate of population parameter. Ratio method is applicable when the study variable and auxiliary variable has positive correlation between them and regression line of study variable on auxiliary variable passes over origin. Theoretically and empirically it is shown that the proposed estimator is more precise than the some existing estimators. Numerical study is carried out in two natural populations, which easily shown that mean square error of proposed estimator is less than existing estimator. It is therefore found that proposed estimator perform better than existing estimators.

REFERENCES

1. Al-Jararha, J. and Al-Haj Ebrahem, M. (2012). A ratio estimator under general sampling design. *Austrian Journal of Statistics*, 41(2), 105-115.
2. Asghar, A., Sanaullah, A. and Hanif, M. (2014). Generalized Exponential Type Estimator forPopulation Variance in Survey Sampling. *Revista Colombiana de Estadística*, 37(1), 213-224.
3. Cochran, W.G. (1977). *Sampling techniques*, Third Edition, Wiley Eastern Limited.
4. Gupta, S. and Shabbir, J. (2008). Variance estimation in simple random sampling using auxiliary information. *Hacettepe Journal of Mathematics and Statistics*, 37, 57-67.
5. Isaki, C.T. (1983). Variance estimation using auxiliary information. *Journal of the American Statistical Association*, 78, 117-123.
6. Kadilar, C. and Cingi, H. (2006a). Improvement in variance estimation using auxiliary Information. *Hacettepe Journal of Mathematics and Statistics*, 35(1), 111-115.
7. Khan, M. and Shabbir, J. (2013). A Ratio Type Estimator of Population Variance using Quartiles of an Auxiliary Variable. *Journal of Statistics Applications & Probability*, 2(3), 319-325.
8. Misra, P. and Singh, R.K. (2015). Estimators for Finite Population Variance Using Mean and Variance of Auxiliary Variable. *International Journal of Probability and Statistics*, 4(1), 1-11.
9. Murthy, M.N. (1967). *Sampling theory and methods*. Statistical Publishing Society, Calcutta, India.
10. Olufadi, Y. and Kadilar, C. (2014). A Study on the Chain Ratio-Type Estimator of Finite Population Variance. *Journal of Probability and Statistics*, 2014, Article ID 723982, 5 pages, http://dx.doi.org/10.1155/2014/723982.
11. Panse, V.G. and Sukhatme, P.V. (1954). *Statistical methods for agriculture workers*. Indian Agricultural Research Council, New Delhi, India.
12. Sharma, P. and Singh, R. (2014). Improved ratio type estimator using two auxiliary variables under second order approximation. *Mathematical Journal of Interdisciplinary Sciences*, 2(2), 179-190.
13. Shittu, O.I., and Adepoju, K.A. (2014). On the Efficiency of Modified Ratio Estimator Based on Linear Combination of Kurtosis, Median and Quartile Deviation. *International Journal of Modern Mathematical Sciences*, 11(2), 103-107.
14. Shukla, A.K., Misra, S., Mishra, S.S. and Yadav, S.K. (2015). Optimal Search of Developed Class of Modified Ratio Estimators for Estimation of Population Variance. *American Journal of Operational Research*, 5(4), 82-95.
15. Singh, D. and Chaudhary, F.S. (1986). *Theory and analysis of sample survey designs*. New Age International Publisher.
16. Singh, H.P. and Solanki, R.S. (2013). A new procedure for variance estimation in simple random sampling using auxiliary information. *Statistical Papers*, 54, 479-497.
17. Singh, H.P., Tailor, R. and Tailor, R. (2010). On ratio and product methods with certain known population parameters of auxiliary variable in sample surveys. *SORT: Statistics and Operations Research Transactions*, 34(2), 157-180.

18. Singh, P.H., Pal, S.K. and Solanki, R.S. (2013). Improved Estimation of Finite Population Variance Using Quartiles. *Journal of the Turkish Statistical Association*, 6(3), 116-121.
19. Subramani, J. and Kumarapandiyan, G. (2012a). Variance estimation using median of the auxiliary variable. *International Journal of Probability and Statistics*, 1(3), 36-40.
20. Subramani, J. and Kumarapandiyan, G. (2012b). Variance estimation using quartiles and their functions of an auxiliary variable. *International Journal of Statistics and Applications*, 2(5), 67-42.
21. Subramani, J. and Kumarapandiyan, G. (2012c). Estimation of variance using deciles of an auxiliary variable. *Proceedings of International Conference on Frontiers of Statistics and Its Applications*, Bonfring Publisher, 143-149.
22. Subramani, J. and Kumarapandiyan, G. (2013). Estimation of variance using known coefficient of variation and median of an auxiliary variable. *Journal of Modern Applied Statistical Methods*, 12(1), 58-643.
23. Subramani, J. and Kumarapandiyan, G. (2015). A Class of Modified Ratio Estimators for Estimation of Population Variance. *Journal of Applied Mathematics, Statistics and Informatics*, 11(1), 91-114.
24. Subramani. J. (2015). Generalized Modified Ratio Type Estimator for Estimation of Population Variance. *Sri Lankan Journal of Applied Statistics*, 16(1), 69-90.
25. Upadhaya, L.N. and Singh, H.P. (2001). Estimation of population standard deviation using auxiliary information. *American Journal of Mathematics and Management Sciences*, 21(3-4), 345- 358.
26. Wolter, K.M. (1985). *Introduction to Variance Estimation*. Springer-Verlag.
27. Yadav, S.K. and Kadilar, C. (2013). Improved Exponential type ratio estimator of population variance. *Colombian Journal of Statistics*, 36(1), 145-152.
28. Yadav, S.K., Mishra, S.S., Shukla, A.K. and Tiwari, V. (2015). Improvement of Estimators for Population Variance using Correlation Coefficient and Quartiles of the Auxiliary Variable. *Journal of Statistics Applications & Probability*, 4(2), 259-263.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 43-50

UTILIZATION OF SIA LOG-SYMMETRIC DISTRIBUTIONS FOR ACHIEVING BETTER FIT

Saleha Naghmi Habibullah[1] and **Kessica Xavier**[2]
Department of Statistics, Kinnaird College for Women, Lahore, Pakistan
Email: [1]saleha.habibullah@kinnaird.edu.pk
[2]kessica.xavier@gmail.com

ABSTRACT

The class of SIA log-symmetric distributions is special in that it facilitates the construction of estimators that are more efficient than the corresponding moment estimators --- a phenomenon which, more often than not, yields better-fitting models for real life data-sets as compared with models obtained through moments estimators. Such estimators have recently been given the title 'SIA-estimators' and the utilization of an SIA log-symmetric distribution in conjunction with an SIA estimator of the distribution parameter(s) can be called 'the SIA methodology'. In this paper, we apply one of the recently developed SIA log-symmetric distributions to a data-set taken from the literature for which the exponential distribution had been considered initially and the gamma distribution subsequently. The chi square test of goodness of fit testifies to the usefulness of the SIA-methodology.

KEYWORDS

SIA log-symmetric distributions, SIA-estimators, SIA-methodology, 'SIA lognormal power distribution', better-fitting model.

1. INTRODUCTION AND LITERATURE REVIEW

There exists an interesting class of probability distributions of the non-negative continuous random variable X for each of which X/median is distributed in exactly the same manner as median/X. This phenomenon has been referred to as self-inversion (Habibullah and Saunders, 2011) and the class of such distributions includes a number of well-known density functions including the lognormal, log-logistic, log-Laplace, log-Cauchy and Birnbaum Saunders distributions. Habibullah and Fatima (2015) proposed that the median of such a distribution be denoted by A, and the distribution be referred to as being 'Self-Inverse at A' ('SIA'). However, in view of the fact that such distributions had already been referred to as being 'log-symmetric' by Jones (2008), Habibullah (2017) proposed that the two nomenclatures be merged and this class of distributions be referred to as the class of 'SIA log-symmetric distributions'.

Estimation of parameters is one of the first steps required in the process of model-fitting and, during the past five years, a number of papers have emerged showing that the self-inversion property facilitates the construction of estimators of distribution parameters that are *more efficient* than the ordinary moment-estimators. (See Fatima and Habibullah (2013a,b), Habibullah and Fatima (2014a,b) and Habibullah (2015).) This remarkable property makes this class of distributions attractive and somewhat special. These estimators have been given the title ‘SIA-estimators’ (Xavier and Habibullah, 2016; Ali and Habibullah, 2016) and the utilization of an SIA log-symmetric distribution for modeling a real-life data-set that is exhibiting a positively skewed histogram on the positive half of the X-axis in conjunction with the utilization of an SIA-estimator of the distribution parameter(s) can be referred to as ‘the SIA-methodology’.

In this paper, we apply the Lognormal-Power distribution developed by (Samuel and Habibullah (2016) to a data-set on failure-times of transmitter tubes for which the exponential distribution was considered by Davis (1952) and the gamma distribution was suggested by Greenwood and Durand (1960). We apply the chosen SIA log-symmetric distribution to these failure-times using (i) the method of moments (‘non-SIA’) estimators, and (ii) the SIA-estimators of μ and σ, setting r =0.97 in each case. As well, we fit the gamma distribution to this failure-times using moment estimators of θ and k. The value of the chi-square statistic turns out to be the smallest in the case of the SIA lognormal power distribution using the SIA-estimators, second-smallest in the case of the SIA lognormal power distribution using the Non-SIA estimators and the largest in the case of the gamma distribution. This demonstrates the usefulness of the SIA-methodology.

2. THE ‘SIA-LOGNORMAL POWER DISTRIBUTION’

The widely known lognormal distribution is given by

$$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2}, \qquad 0<x<\infty,\ \mu>0,\sigma>0$$

As ‘A’, the median of the distribution is given by e^{μ}, therefore, the pd$f$ of the distribution can be re-written as

$$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{\ln^2\left(\frac{x}{\mathrm{A}}\right)}{2\sigma^2}},\ 0<x<\infty,\ \mathrm{A}>0,\sigma>0 \tag{2.1}$$

İt is easy to show that density function (2.1) is a member of the class of SIA log-symmetric distributions.

Using the power transformation Y = X^r, Samuel and Habibullah (2016) developed the ‘SIA-Lognormal Power distribution’ and derived its fundamental properties. The PDF of this distribution is given by

$$f(x)=\frac{1}{xr\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\ln x-r\mu}{r\sigma}\right)^2},\ 0<x<\infty,\mu>0,\sigma>0,r>0 \tag{2.2}$$

İt is interesting and important to note that:

i) The ‘SIA-lognormal power distribution’ given by (2.2) is exactly the same as the lognormal distribution with paramters rμ and rσ.

ii) The 'SIA lognormal power distribution' is self-inverse at $e^{r\mu} = \left(e^{\mu}\right)^r = \mathrm{A}^r$ (not at A).

iii) The parameter r can be thought of as a 'tuning parameter' (one that enables us to improve the fit of the distribution to any given data-set).

3. FAILURE-TIMES DATA

We pick up a data-set pertaining to failure-times (in hours) of 188 transmitter tubes given in Greenwood and Durand (1960). The data is reproduced in Table 3.1 whereas the histogram of the data-set is given in Figure 3.1. It is obvious from the histogram that the empiriucal distribution is moderately positively skewed.

Table 3.1
Failure-data for 188 Transmitter Tubes

Hours to Failure	Observed Frequency	Hours to Failure	Observed Frequency	Hours to Failure	Observed Frequency
0-5	15	50-55	4	100-105	3
5-10	18	55-60	7	105-110	4
10-15	24	60-65	2	110-115	1
15-20	30	65-70	1	115-120	2
20-25	22	70-75	3	120-125	2
25-30	15	75-80	1	125-130	0
30-35	10	80-85	1	130-135	0
35-40	7	85-90	2	135-140	0
40-45	6	90-95	2	140-145	0
45-50	4	95-100	1	145-150	1

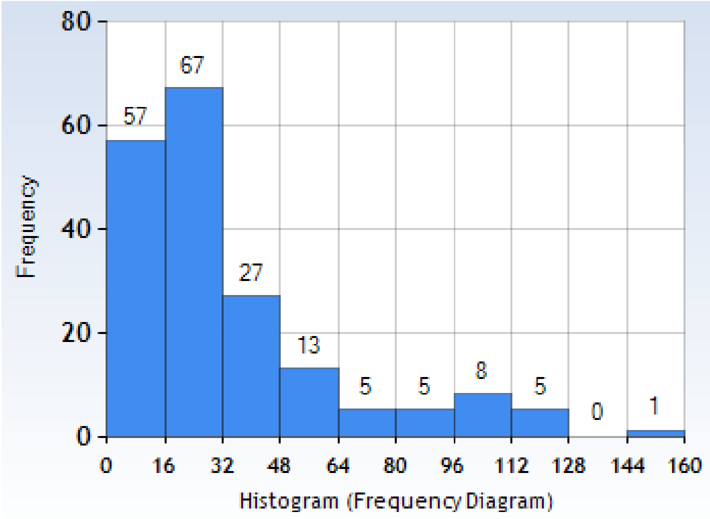


Fig. 3.1: Histogram of the Failure-Times Data Constructed through Software Freely available on the website www.socscistatistics.com/descriptive/histograms/

4. MODEL-FITTING

We fit the 'SIA Lognormal Power distribution' to the failure-time data and compare it with the gamma distribution suggested by Greenwood and Durand (1960) with reference to goodness of fit.

4.1 Fitting the 'SIA Lognormal Power Distribution' using Non-SIA Estimators of Distribution Parameters

For the SIA-lognormal power distribution, the 'Non-SIA' estimators (i.e. the ordinary method of moments estimators) of μ and σ are given by

$$\hat{\mu}_{Non-SIA} = \frac{1}{r}\ln\left(\hat{A}_{Non-SIA}\right) \tag{4.1}$$

and

$$\hat{\sigma}_{Non-SIA} = \frac{1}{r}\sqrt{2\ln\left(\frac{\bar{x}}{\hat{A}_{Non-SIA}}\right)} \tag{4.2}$$

where $\bar{x}$ is the simple mean and $\hat{A}_{Non-SIA}$ is the median of the data.

The mean of the data-set is 32.21 and the median is 21.59. Using these values and letting r = 0.97, we obtain $\hat{\mu}_{Non-SIA} = 3.167$ and $\hat{\sigma}_{Non-SIA} = 0.922$ which yield expected frequencies given in column no. 3 of Table 4.1.

Table 4.1
Observed Frequencies and Expected Frequencies using Three Different Models

x	f_0	Expected frequencies based on the SIA Lognormal Power Distribution		Expected frequencies based on the Gamma distribution
		$f_{e(Non\text{-}SIA)}$	$f_{e(SIA)}$	$f_{e\ (gamma)}$
0-5	15	08.56	12.56	21.97
5-10	18	24.08	27.27	22.52
10-15	24	25.37	24.95	20.39
15-20	30	21.83	20.98	17.95
20-25	22	18.66	17.21	15.60
25-30	15	14.06	13.19	13.45
30-35	10	12.04	11.15	11.54
35-40	07	09.95	08.49	9.86
40-45	06	07.38	07.28	8.41
45-50	04	06.79	06.16	7.15
50-55	04	05.18	04.64	6.07
55-65	09	08.19	07.77	9.51
65-80	05	08.35	07.81	9.45
80-100	06	06.40	06.21	7.02
100-150	13	11.16	12.32	7.10
Total	**188**	**188**	**188**	**188**

4.2 Fitting the 'SIA Lognormal Power Distribution' using SIA-Estimators of Distribution Parameters

Birnbaum and Saunders (1969) proposed the 'mean-mean' defined as the geometric mean of the arithmetic and the harmonic means of the data as an estimator of β, the median of the Fatigue-Life distribution. It is easy to show that, for any SIA log-symmetric distribution, the square root of the product of the arithmetic and harmonic means of the distribution is equal to the distribution median. As such, the 'mean-mean' proposed by Birnbaum and Saunders (1969) can be regarded as an SIA estimator of the median of every SIA log-symmetric distribution. We utilize the mean-mean to estimate the median of the 'SIA Lognormal Power distribution' and call it $\hat{A}_{SIA}$.

Since the mean of the failure-times data is 32.21 and the harmonic mean is 12.74, we have

$$\hat{A}_{SIA} = \sqrt{AM \times HM} = \sqrt{32.21 \times 12.74} = 20.26 \tag{4.3}$$

Using the value of $\hat{A}_{SIA}$ in (4.3) and letting r = 0.97, we obtain

$$\hat{\mu}_{SIA} = \frac{1}{r}\ln\left(\hat{A}_{SIA}\right) = 3.10 \tag{4.4}$$

and

$$\hat{\sigma}_{SIA} = \frac{1}{r}\sqrt{2\ln\left(\frac{\bar{x}}{\hat{A}_{SIA}}\right)} = 0.99 \tag{4.5}$$

The values of $\hat{\mu}_{SIA}$ and $\hat{\sigma}_{SIA}$ obtained in (4.4) and (4.5) yield expected frequencies given in column no. 4 of Table 4.1.

4.3 Fitting The Gamma Distribution using Moment Estimators of Distribution Parameters

The pdf of the well-known gamma distribution is given by

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}, \qquad 0 < x < \infty, k > 0, \theta > 0$$

For this distribution, the ordinary method of moments estimators of k and θ are given by

$$\hat{k} = \frac{\bar{x}^2}{s^2} \tag{4.6}$$

and

$$\hat{\theta} = \frac{s^2}{\bar{x}} \tag{4.7}$$

The variance of the failure-times data is 896.8557. Hence, using the mean and variance of the data-set, we obtain $\hat{k} = 1.15$ and $\hat{\theta} = 28.00$ which yield expected frequencies given in column no. 5 of Table 4.1.

4.4 Comparison through the Chi-Square Statistic

Using the chi-square goodness of fit at 12 degrees of freedom taking α = 0.05 the tabulated value of χ^2 is 21.03. As far as the computed values of χ^2 are concerned, in the case of the gamma distribution, we have χ^2 = 25.64, in the case of the 'SIA Lognormal Power distribution' using Non-SIA estimators of distribution parameters, χ^2 = 14.83 and in the case of the 'SIA Lognormal Power distribution' using SIA-estimators, we obtain χ^2 = 11.82. Thus we can see that the 'SIA Lognormal Power distribution' fitted through Non-SIA estimators is a much better fit than the gamma distribution suggested by

Greenwood and Durand (1960) whereas the 'SIA Lognormal Power distribution' fitted through SIA-estimators is an *even better* fit than the 'SIA Lognormal Power distribution' fitted through Non-SIA estimators. The 'SIA Lognormal Power distribution' fitted through SIA-estimators of μ and σ can now be used to compute probabilities of interest such as the probability of failure of the transmitter tube within the warranty period and the reliability of the equipment at a relatively advanced age.

5. CONCLUSION

SIA log-symmetric distributions facilitate the construction of SIA-estimators of distribution parameters due to which we obtain better-fitting models for real-life data-sets as compared with models obtained through estimators obtained by the method of moments. Utilization of an SIA log-symmetric distribution in conjunction with an SIA estimator of the distribution parameter(s) can be called 'the SIA methodology'. In this paper, we have demonstrated the usefulness of the SIA methodology by applying the 'SIA Lognormal Power distribution' to a data-set on failure-times of transmitter tubes for which the exponential and gamma distribution were considered previously.

REFERENCES

1. Ali, A. and Habibullah, S.N. (2016). On A Decile-Based SIA-Estimator of the Scale Parameter of the Log-Cauchy Distribution, *Proceedings of the 14th International Conference on Statistical Sciences*, Karachi, Pakistan, 29, 427-434.
2. Birnbaum, Z.W. and Saunders, S.C. (1969). Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability*, 6, 328-347.
3. Davis, D.J. (1952). An analysis of some failure data. *J. Amer. Statist. Assoc*., 47, 113-150.
4. Fatima, S.S. and Habibullah, S.N. (2013a). Self-Inversion-Based Modifications of L-Estimators of Central Tendency for Probability Distributions in the Field of Reliability and Safety, *Proceedings of the International Conference on Safety, Construction Engineering and Project Management (ICSCEPM)* (Islamabad, Pakistan, (pp. 19-21).
5. Fatima, S.S. and Habibullah, S.N. (2013b). On Modifications of L-Estimators of Dispersion in the Case of Self-Inverse Distributions, *Proceedings of 11th International Conference on Statistical Sciences: Social Accountability, Global Economics and Human Resource Development with Special Reference to Pakistan*, Dera Ghazi Khan, Pakistan, 25, 1-8.
6. Greenwood, J.A. and Durand, D. (1960). Aids for Fitting the Gamma Distribution by Maximum Likelihood. *Technometrics*, 2(1), 55-65.
7. Habibullah, S.N. (2015). On a Newly Developed Methodology For More Accurate Modeling, Invited Paper, *Proceedings of the First International Multi-Disciplinary Conference (IMDC 2015),* (University of Lahore Gujrat Campus, Gujrat, Pakistan, Nov. 23-24, 2015).
8. Habibullah, S.N. (2017). On the Power Generalization of Log-Symmetric SIA Distributions. *Journal of ISOSS,* 3(1), 45-54.

9. Habibullah, S.N. and Fatima, S.S. (2014a). SIU-Based Modification in Kelley's Measure of Skewness to Achieve Gains in Efficiency, *Proceedings of the Second ISM International Statistical Conference 2014 with Applications in Sciences and Engineering (ISM II)* organized by Faculty of Industrial Sciences and Technology, Universiti Malaysia Pahang, Kuantan, Malaysia on Aug 12-14, 2014.
10. Habibullah, S.N. and Fatima, S.S. (2014b). Self-Inversion-Based Modification of Crow and Siddiqui's Coefficient of Kurtosis to Achieve Gains in Efficiency, *Proceedings of the Thirteenth Islamic Countries Conference on Statistical Sciences (ICCS-13)*, Bogor, Indonesia, 27, 1-8.
11. Habibullah, S.N. and Fatima, S.S. (2015). On a Newly Developed Estimator for More Accurate Modeling with an Application to Civil Engineering, *Proceedings of the 12th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP12)* organized by CERRA (Vancouver, BC, Canada, July 12-15, 2015).
12. Habibullah, S.N. and Saunders, S.C. (2011). A Role for Self-Inversion, *Proceedings of International Conference on Advanced Modeling and Simulation (ICAMS, Nov 28-30, 2011)* published by Department of Mechanical Engineering, College of Electrical and Mechanical Engineering, National University of Science and Technology (NUST), Islamabad, Pakistan.
13. Jones, M.C. (2008). On Reciprocal Symmetry. *Journal of Statistical Planning and Inference*, 138(10), 3039-3043.
14. Samuel, S. and Habibullah, S.N. (2016). SIA-Lognormal Power Distribution, *Proceedings of the 2nd International Multi-Disciplinary Conference (IMDC 2016)*, December 19-20, 2016, University of Lahore, Gujrat Campus, Gujrat, Pakistan.
15. Xavier, K. and Habibullah, S.N. (2016). On A Decile-Based SIA-Estimator of the Shape Parameter of the Log-Logistic Distribution, Proceedings of the *14th International Conference on Statistical Sciences*, Karachi, Pakistan, 29, 419-426.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 51-55

THE EFFECT OF FAST FOOD ON STUDENTS OF SHAH ABDUL LATIF UNIVERSITY KHAIRPUR

Shoaib Ali Shah[1], **Karamullah Memon**[2] and **Faisal Afzal Siddiqui**[3]
Department of Statistics, Shah Abdul Latif University, Khairpur, Pakistan
Email: [1]shoaibalishah345@gmail.com
[2]karamullah.memon@gmail.com
[3]brc.khi@gmail.com

ABSTRACT

Fast food costs relatively title and tastes good but the negative effects on physical health last much longer than these immediate concerns with the high calorie meals come more fat, cholesterol, salt and sugar and therefore fewer vitamins, minerals and other nutrients than in healthier foods.

The objective of this study to know about the effect of fast food consumption on the health of university going students (20-22 year). Total of 100 university going student were selected from different department students of SALU, Khairpur. The study was carried out by using the following tools to analyze the effect of fast food consumption health of respondents. Self designed and pre tested questionnaire was used in the study.

SPSS version 21 was used to draw meaningful inferences from the collected raw data. The result found that among 20 year 100% respondents were found obese level 1. Among 20-22 year 25.0% of respondents were found obese level 3. Among 22-24 year 56.0% of respondents were found obese level 3 among 24-26 year 13.0% respondents.

KEY WORDS

University goes students, effect of fast food consumption, health status.

INTRODUCTION

Fast food in emerging trend among the younger generation. The ready availability taste low cast marketing strategies and peer pressure make them popular with students fast food restaurants are primed to maximize the speed efficiency and conformity, (Arya and Mishra, 2013). The menu is kept limited and standardized essentially to minimize the waiting time so that the stompers eat quickly and leave (Akbay, Tiryaki and Gul 2007). This perspective delineates the emerging the fast food culture in Pakistan its impact on students and strategies to counter it fast foods are widely available in university through variety of outlets. cafeterias at the university offer sods, cold drinks, chips and many other foods of low nutritional value sale of junk foods in university cafeteria often competes with more nutritious university lunch schemes (Poti, Duffey and Popkin 2013).

On a typical day that fast food is eaten student consume substantially more total energy and have worse dietary quality compeered with a typical day without fast food there are much greater side effects of fast food and people un aware of its ill consequences (Rosenheck, 2008). It can tend to many detrimental diseases pronged use of fast food is also a cause of health problems obesity indigestion problem and high cholesterol fast food companies are targeting kids and youngster through great promotion strategies delicious recopies and attractive advertisement (Bowman et al. 2004).

There are much greater sued effects of fast food and people are unaware of its ill consequences chubbiness increase in cholesterol levels dietary deficiencies cardiac dies orders loss of muscle mass depression. Sexual dysfunction, asthma, strokes (Paeratakul et al., 2003).

Fast foods have high level of fast and sugars that are not only unhealthy but addictive and that creates a vicious cycle making it hard for students choose healthy food high content of trans fat in commercially available fast foods predispose students to risk of future heart diseases (Yahya, Zafar and Shafiq, 2013).

OBJECTIVES

1. To study the effect of fast food consumption on the health of university going students.
2. To know about the health status of university going students.

Design

A study was conducted recruiting university going students. University were selected from Pakistan Sindh District Khairpur.

A total of 100 university going students were selected randomly. A total of 100 students were selected from different department of university of (SALU) District Khairpur. The samples were selected randomly from different department.

Tool

Questionnaire method:- The schedule was used to collect the information on general profile and use of BMI(Body Mass index)of university going students. Pilot study was conducted to assess reliability and validity of the questionnaire. The self-made questionnaire was developed.

Statistical Analysis

SPSS version 21 was used to draw meaningful inferences from the collected raw data.

RESULT AND DISCUSSION

The data on sample characteristics were analyzed using descriptive statistics and presented in term of frequency, percentage. The data obtained from sample are presented in term of Age.

Table 1
Distribution of Respondent According to Sex
Statistics
Gender

N	Valid	100
	Missing	0

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	**Male**	72	72.0	72.0	72.0
	Female	28	28.0	28.0	100.0
	Total	**100**	**100.0**	**100.0**	

The result finished in Table 1, indicates the majorities of male 72.0% of the respondent sample were male and majorities of female 28.0% respondent sample were female.

Crosstabs

Case Processing Summary

	Cases					
	Valid		**Missing**		**Total**	
	N	**Percent**	**N**	**Percent**	**N**	**Percent**
Age * your weight	100	100.0%	0	0.0%	100	100.0%

Age * your weight Cross tabulation

			Your weight								Total
			58kg	50kg	65kg	72kg	60kg	75kg	55kg	49kg	
Age	20 year	Count	0	1	1	2	1	0	1	0	6
		Expected Count	.8	.4	1.5	.8	1.1	.7	.6	.2	6.0
	20-22 year	Count	3	1	5	2	2	5	6	1	25
		Expected Count	3.3	1.5	6.3	3.3	4.5	2.8	2.5	1.0	25.0
	22-24 year	Count	8	1	14	9	14	6	1	3	56
		Expected Count	7.3	3.4	14.0	7.3	10.1	6.2	5.6	2.2	56.0
	24-26 year	Count	2	3	5	0	1	0	2	0	13
		Expected Count	1.7	.8	3.3	1.7	2.3	1.4	1.3	.5	13.0
Total		Count	13	6	25	13	18	11	10	4	100
		Expected Count	13.0	6.0	25.0	13.0	18.0	11.0	10.0	4.0	100.0

Shows the (body mass index) of university going students by standard measure values among 20 year of age category only 6 (100%) respondent was found belonging to obese level 2 whereas normal, overweight, obese level land obese level 3 categories were found nil among 20-22 year of category 2(25.0%) respondent were found level 1 obse, 2(25.0%) respondent were found level 2 obese and 3 22-24 year (56.0%) respondents were found level 3 obese whereas normal and overweight categories were found nil among the year of age category 24-26 year(13.0%) respondents were found level over weight categories were found among the 22-24 year of age category(56.0%) respondents were found level over weight.

Chi-Square Tests

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	34.014[a]	21	.036
Likelihood Ratio	36.821	21	.018
Linear-by-Linear Association	3.520	1	.061
N of Valid Cases	100		

a. 25 cells (78.1%) have expected count less than 5. The minimum expected count is .24.

DISCUSSION

Fast food products are often rumored to case variety of health problems ingredients that are artificial, high in sugar or fat are not meant to be eaten on a regular basis consuming these ingredients regularly can lead to health issues including heart disease diabetes and cancers. Fast food is convenient and tasty though it prepared with low nourishing or unhealthy ingredients. In this study among 20 year (6.0%) of respondents were found obese level 3. Among 20-22 year (25.0%) respondents were found obese level 3 among 22-24 year (56.0%) respondents were found obese level 2 among the 24-26 year 13.0%) respondents were found other level 3. Another finding related to this study.

CONCLUSION

Food can be appealing for a variety of reasons, including convenience, price and test for student who do not always understand the health consequences of their eating habits fast food may appear especially appetizing. However, regularly consuming fattening food can be addictive for student and lead to complications like obesity, chronic illness, low self-esteem and even depression, as well as affecting how they perform in university and extracurricular activities. This study found that among 20-22 year (25.0%) students were found obese level 3 rather than other age groups.

REFERENCES

1. Arya G. and Mishra, S. (2013). Effect of junk food & beverages on Adolescent's health. *IOSR Journal of Nursing and Health Science*, 1, 26-32.
2. Akbay, C., Tiryaki, G.Y. and Gul, A. (2007). Consumer characteristics influencing fast food consumption in Turkey. *Food control*, 18(8), 904-913.
3. Poti, J.M., Duffey, K.J. and Popkin, B.M. (2013). The association of fast food consumption with poor dietary outcomes and obesity among children: Is it the fast food or the remainder of the diet? *The American Journal of Clinical Nutrition*, 99(1), 162-171.
4. Rosenheck, R. (2008). Fast Food Consumption and Increased Caloric Intake. *The International Association for the Study of Obesity*, 21, 245-253.
5. Bowman, S.A., Gortmaker, S.L., Ebbeling, C.B., Pereira, M.A. and Ludwig, D.S. (2004). Effects of fast-food consumption on energy intake and diet quality among children in a national household survey. Pediatrics, 113(1), 112-118.
6. Paeratakul, S., Ferdinand, D.P., Champagne, C.M., Ryan, D.H. and Bray, G.A. (2003). Fast-food consumption among US adults and children: dietary and nutrient intake profile. Journal of the American dietetic Association, 103(10), 1332-1338.
7. Yahya, F., Zafar, R. and Shafiq, S. (2013). Trend of fast food consumption and its effect on Pakistani society. Food Science and Quality Management, 11, 1-7.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 57-66

EFFECT OF FREE WIFI ON THE ACADEMIC PERFORMANCE OF STUDENTS AT SHAH ABDUL LATIF UNIVERSITY (SALU), KHAIRPUR

Muhammad Yasin Pitafi[1] and **Faisal Afzal Siddiqui**[2]
Department of Statistics, Shah Abdul Latif University, Khairpur, Pakistan
Email: [1]myasinpitafi@gmail.com
[2]brc.khi@gmail.com

ABSTRACT

Internet is one of the largest networking which computers connected to share the information in all over the world due to this connection world known as the global village means we can talk or share information in seconds at one corner to another corner of the world. But recent advancement in the information technology is wireless technology which means that the computers (Electronic Devices) share the information without any wire. Wireless Fidelity (WIFI) is one of the wireless technology which is the common term in the process of connection Local Area Network (LAN) with high frequency wireless. Private and Public institutions, Businesses, Organisations, Government agencies and Other corporate have investing on the large scale and growing the wireless technology operations. Educational Institutions providing the free WIFI to promote education and to increase opportunities to achieve their teaching and learning objectives in education and research through better WIFI speed of uploading and downloading. This study is conducted at Shah Abdul Latif University Khairpur (SALU), with the aim to determine the "Effect of Free WIFI on the academic performance of students at SALU Khairpur". This study also determines how many students using this technology for academic problems and how many students use only for personal matters or for the social problems. In this study primary data has been used and data collected through questionnaire method. The survey conducted by the 175 BS students as a sample size. Conclusion was drawn with the help of Tables generated through SPSS, these tables indicating the Effect of Free WIFI on the academic performance of students at SALU Khairpur.

KEYWORDS

Internet, WIFI technology, Social life, Academic performance, Students, Khairpur, Pakistan.

INTRODUCTION

The word "Internet" is the combination of two words Inter & Net, Inter means place (Between or among the people or places) and net means Strong thread or meshed fabric.

Internet… a word that dominates most computer-related conversations. For good reason, too: Since its appearance in the 90s, it has become an integral part of our lives.

But what exactly is the Internet? It's a global network of millions of computers that exchange information. It's the largest network that exists today and connects private, government and academic networks. On the Internet you can find a huge amount of information, photos, videos and music.

Over the past two decades there has been a very interesting transformation in the area of communication and how people are sharing different forms of information and ideas. Previously access to Internet in sharing of data information and communication were done by the use of wired technologies, but recently things have changed with the involvement of wireless (without wire) communication technology (WCT) (Negus and Petrick, 2009).

The Internet is a relatively new channel for scholarly resources, and contains vast quantities of information that vary a great deal regarding its contents, aim, target group, reliability etc. Hence, it is important that the end-user is aware of the diverse information available on the Internet, and educated in the criteria by which the information content should be assessed (Chapman, 2002).

Students have only recently received the opportunity to use the Internet to seek and obtain scholarly material and consequently, knowledge on how effectively they make use of this channel is limited. Students' information seeking culminates as they work on their theses. Many studies have been conducted regarding the type of information the end users seek and obtain on the Internet and in which circumstances they prefer electronic sources to paper sources (Tenopir et al., 2003).

Internet use has become very popular in many areas as well as in education in recent years. Accordingly, Internet access in schools has increased greatly over the last 20 years (Berson, 2000).

WI-FI is a technology that allows electronic devices to connect to Local Area Network (LAN) and be able to access Internet and share connections (O'Leary and O'Leary, 2011).

Wi-Fi allows networking of computers and digital devices without the need for wires. Data is transferred over radio frequencies, allowing Wi-Fi capable devices to receive and transmit data when they are in range of a Wi-Fi network.

The widespread use of the technology and its availability in both residential homes and public places including parks, gathering spots, and coffee shops have made it one of the most popular data transmission technologies available today (Ed Oswald, 2004).

Therefore, it is important for researchers to focus on the wireless technology. The purpose of this study is to know how the technology of (WI-FI) impact on academic performance of students at SALU Khairpur. In developed countries mostly institutions providing the opportunity of Local Area Network (LAN) or the Facility of WI-FI through this networking students can connect their devices through WI-FI in the seek to research of knowledge. But now under developing countries such as Pakistan also providing this facility of WI-FI in the institutions to promote the education. In 2nd Semester of 2018 Shah Abdul Latif University Khairpur start WI-FI facility to increase the researching activities and academic performance of students.

METHODOLOGY

According to Oates (2008), quantitative research is a type of research that involves the use of numerical expressions in terms of data generation. This research method is also related to positivism, realism and interpretivism research design. In addition, Oates (2008), believes that Questionnaire is a commonly used technique of collecting quantitative data for academic research.

This study was conducted at Shah Abdul Latif University (SALU) Khairpur with the 175 participants as a sample size. From the entire Questionnaires distributed, 100% were returned and analyzed after a number of follow-up were made. Random Sampling technique used for the collection of Data, SPSS version 16 used for analyze the data, Using SPSS Frequencies had made and after that Results of Tables discuss in MS Word. On these results conclusion was drawn.

LITERATURE REVIEW

This section determine the previous work of the researcher which is done on the WIFI and other theories that are related to WIFI.

A study conducted by Wakefield [14] in 2008 found that 90% of students believe that WIFI is as important as education as a traditional classroom or typical computer.

WIFI is very important for university students living on campus and indirectly affecting students. Wakefield survey found that 48% of students would choose WIFI rather than other facilities and surveys from International Association of privacy professionals [13] found that WIFI access on their campus sites helped them get better grades.

The study was also conducted by sturgeon, Allison and Miller [10] to determine whether the use of WIFI labs had a significant impact on student performance or not. This study is to identify whether the facilities provided are affecting student performance. From the study, students' performance increased by 84% compared to 64% for the previous year without the use of WIFI. There was an average increase in student outcomes from 56% in 2003 to 71% in 2006 following the provision of such technology facilities. While this study showing that mostly students using WIFI having best grades that means effect of free WIFI on the academic performance, more detail of research and analysis should be continued in Results and Discussion to identify the Effect of free WIFI on the academic performance of students.

RESULTS AND DISCUSSION

After the collection of data, data converted into Tables for analysis.

Table 1
Access of WIFI

	Frequency	Percent	Valid Percent	Cumulative Percent
Yes	99	56.6	56.6	56.6
No	76	43.4	43.4	100.0
Total	**175**	**100.0**	**100.0**	

This question ask to students to know about the access of WIFI. Table 1 indicate that, 56.6% students are using WIFI, 43.4% says that they are not using. Above results showing that majority of the students are using WIFI about 57% improving their academic performance.

Table 2
Using of WIFI

	Frequency	Percent	Valid Percent	Cumulative Percent
0-5 h	163	93.1	93.1	93.1
5-10 h	6	3.4	3.4	96.5
10-15 h	2	2.3	2.3	98.9
15-20 h	4	1.1	1.1	100.0
Total	**175**	**100.0**	**100.0**	

Table 2 showing that respondents respond that 93.1% are using WIFI 0-5 hours, 3.4% using 5-10 hours, 2.3% using 10-15 hours and 1.1% using 15-20 hours. Majority of students using 0-5 hours WIFI in a day for academic and social work which is normal.

Table 3
Purpose of using WIFI

	Frequency	Percent	Valid Percent	Cumulative Percent
Assignment	36	20.6	20.6	20.6
other than Assignment	19	10.9	10.9	31.4
Books/Journals	63	36.0	36.0	67.4
App/Music	27	15.4	15.4	82.9
All	30	17.1	17.1	100.0
Total	**175**	**100.0**	**100.0**	

This Question seeks to understand the purpose of WIFI. The indication from Table 3 above shows that the students are using WIFI as, 20.6% for Assignments, 10.9% for Other than the Assignments, 36% for downloading of Books and Journals, 15.4% for downloading the Applications and Music and 17.1% for All (Assignments, Other than assignments, Books/journals/music). Finding proves that the 36% students are using WIFI for the downloading of Books/Journals and 20.6% for the Assignments. Means that the students are using WIFI for the academic work.

Table 4
Best for Grades

	Frequency	Percent	Valid Percent	Cumulative Percent
Books	68	38.9	38.9	38.9
Internet	75	42.9	42.9	81.7
Both	32	18.3	18.3	100.0
Total	**175**	**100.0**	**100.0**	

This Question seeks to know about the interest of students which is best for grades either books or net? Table 4 indicate that students respond best for study or Grades 38.9% says Books, 42.9% says Net and 18.3% respondents respond both (Books and Net) are best for Grades. Above table showing both Books & net are best for the Grades, But NET having more importance according to respondents.

Table 5
Opinions about the using of WIFI

	Frequency	Percent	Valid Percent	Cumulative Percent
Positive	109	62.3	62.3	62.3
Negative	6	3.4	3.4	65.7
Both	60	34.3	34.3	100.0
Total	**175**	**100.0**	**100.0**	

Above Table 5 showing the behavior or the opinion of students at SALU Khairpur, table indicate that 62.3% students are using WIFI positively, 3.4% Negatively and 34.3% students are using both (Positively and Negatively). Therefore, this study out that more students are using WIFI positively, which will be good for the students' academic performance.

Table 6
Useful of WIFI

	Frequency	Percent	Valid Percent	Cumulative Percent
Useful of WIFI very extent	85	48.6	48.6	48.6
Some	65	37.1	37.1	85.7
Undecided	11	6.3	6.3	92.0
Not	14	8.0	8.0	100.0
Total	**175**	**100.0**	**100.0**	

In above table students respond as WIFI is 48.6% says useful, 37.1% says some useful, 6.3% Undecided and 8% not useful. Therefore 48.6% students says that WIFI is useful for Academic performance and 37.1% says some useful according to these results we can conclude that WIFI is useful.

Table 7
Time to Study in a Day

	Frequency	Percent	Valid Percent	Cumulative Percent
0-3	113	64.6	64.6	64.6
3-6	54	30.9	30.9	95.4
6-10	8	4.6	4.6	100.0
Total	**175**	**100.0**	**100.0**	

How much time you have gave to your study in a day? From table 64.6% students gave 0-3 hours, 30.9% students gave 3-6 hours, 4.6% students gave 6-10 hours. Mostly students giving 0-3 hours our studies which is average.

Table 8
CGPA of the Students

	Frequency	Percent	Valid Percent	Cumulative Percent
0-1	2	1.1	1.1	1.1
1-2	13	7.4	7.4	8.6
2-3	51	29.1	29.1	37.7
3-4	109	62.3	62.3	100.0
Total	**175**	**100.0**	**100.0**	

Above table showing the results as 1.1% having 0-1 CGPA, 7.4% having 1-2 CGPA, 29.1% having 2-3 CGPA and 62.3% having 3-4 CGPA. Means that mostly students are using the WIFI and also having greatest CGPA.

Table 9
Spending time for Social Life

	Frequency	Percent	Valid Percent	Cumulative Percent
0-3	125	71.4	71.4	71.4
3-6	29	16.6	16.6	88.0
6-9	13	7.4	7.4	95.4
9-12	8	4.6	4.6	100.0
Total	**175**	**100.0**	**100.0**	

Above table indicate the social life behavior of students it is clearly indicate that majority of the student give less time to its social life. Table results are as 71.4% students are giving 0-3 hours, 16.6% students are giving 3-6 hours, 7.4% students are giving 6-9 hours and 4.6% students are giving 9-12 hours.

Table 9
WIFI Improves Communication

	Frequency	Percent	Valid Percent	Cumulative Percent
Strongly disagree	16	9.1	9.1	9.1
Disagree	33	18.9	18.9	28.0
Not sure	28	16.0	16.0	44.0
Agree	67	38.3	38.3	82.3
Strongly agree	31	17.7	17.7	100.0
Total	**175**	**100.0**	**100.0**	

To understand whether WIFI improves communication between students and lecturers, this question was asked (Do WIFI improves your communication with classmates and lecturers?). Table 9 indicate that, 9.1% of participates Strongly disagree, 18.9% Disagree, 16% Not sure, 38.3% Agree and 17.7% respondents are Strongly Agree. From above mention table we can say that majority students are Agreed that the use of WIFI improves how students communicate and interact with each other and their lecturers.

Table 10
Awareness of WIFI

	Frequency	Percent	Valid Percent	Cumulative Percent
Strongly disagree	14	8.0	8.0	8.0
Disagree	10	5.7	5.7	13.7
Not sure	48	27.4	27.4	41.1
Agree	79	45.1	45.1	86.3
Strongly agree	24	13.7	13.7	100.0
Total	**175**	**100.0**	**100.0**	

One might assume that most students are familiar with WIFI based on the use rate, but it must not be ignored that some students are not familiar with WIFI based on their background and other factors; hence it is important to understand if participants truly familiarize themselves with WIFI. Then, Table 10 shows a question (Would you say that most students are familiar with WIFI?) The findings show that 8% Strongly Agree, 5.7% Disagree, 27.4% Not sure, 45.1% are Agree and 13.7% are Strongly Agree. Most students are familiar with WIFI.

Table 11
WIFI Improve Academic Performance

	Frequency	Percent	Valid Percent	Cumulative Percent
Strongly disagree	13	7.4	7.4	7.4
Disagree	19	10.9	10.9	18.3
Not sure	26	14.9	14.9	33.1
Agree	90	51.4	51.4	84.6
Strongly agree	27	15.4	15.4	100.0
Total	**175**	**100.0**	**100.0**	

This question tries to find out if participants can study well because of migration to WIFI. Table 11 shows that 7.4% of the participants Strongly disagree, 10.9% Disagree, 14.9% Not sure, 51.4% Agree and 15.4% are Strongly agree. Above table shows that about 67% students are agree WIFI improve the academic performance.

Table 12
Location to Access for WIFI

	Frequency	Percent	Valid Percent	Cumulative Percent
Class room	47	26.9	26.9	26.9
Computer lab	22	12.6	12.6	39.4
Central Library	66	37.7	37.7	77.1
Student Cafeteria	28	16.0	16.0	93.1
Hostel	4	2.3	2.3	95.4
All	8	4.6	4.6	100.0
Total	**175**	**100.0**	**100.0**	

The desired location of accessing WIFI is aimed at finding out the location that most students have access to WIFI. The options given are based on locations that most students are always spend time, such as, Class room, computer lab, Library, student cafeteria and

All mentions. In above Table 12, The response indicate that 26.9% of the student prefer accessing WIFI at Class room, 12.6% Computer lab, 37.7% Central library, 16% Students cafeteria, 2.3% Hostel and 4.6% at all above mention (class room, Computer Lab, Central Library, Students cafeteria and Hostel).Majority students said that they use WIFI at Central Library and at Class Room than other places.

Table 13
Frequency Usage of WIFI by Students at SALU Khairpur

	Frequency	Percent	Valid Percent	Cumulative Percent
Hourly	53	30.3	30.3	30.3
Daily	85	48.6	48.6	78.9
Weekly	28	16.0	16.0	94.9
Monthly	9	5.1	5.1	100.0
Total	**175**	**100.0**	**100.0**	

How often do you use WIFI? This question aimed to find out how often students get connected to WIFI. The findings shows that 30.3% students are connect Hourly to WIFI, 48.6% Daily, 16% Weekly and 5.1% students are connect Monthly. Therefore, this study out that majority students are connect to WIFI daily.

Table 14
Preferred Device use to Access WIFI by Students at SALU Khairpur

	Frequency	Percent	Valid Percent	Cumulative Percent
Smartphone	147	84.0	84.0	84.0
Tablet	1	.6	.6	84.6
Laptop	20	11.4	11.4	96.0
Other device	7	4.0	4.0	100.0
Total	**175**	**100.0**	**100.0**	

There are different kinds of devices that can be students used to access the WIFI at University. This question aimed to know the most preferred device students use when accessing WIFI in the University. The findings in Table 14 shows that 84% Students access WIFI via Smartphone (Mobile), 0.6% through Tablet PC, 11.4% through Laptop and 4% through other devices. From above table majority students are using the Smartphones to access the WIFI at SALU Khairpur.

Table 15
Quality of WIFI at SALU Khairpur

	Frequency	Percent	Valid Percent	Cumulative Percent
Very low	31	17.7	17.7	17.7
Low	39	22.3	22.3	40.0
Fair	68	38.9	38.9	78.9
High	35	20.0	20.0	98.9
Very high	2	1.1	1.1	100.0
Total	**175**	**100.0**	**100.0**	

The quality of WIFI service provided to students also needs to be tested. By quality the study regards factors like accessibility, speed, security and reliability. The question was asked to students "How would you rank the quality levels of WIFI in SALU Khairpur?" The response of this question by students in above Table 15 as 17.7% ranked WIFI Very low, 22.3% Low, 38.9% Fair, 20% High and 1.1% students say it is Very high. From above table we can see that majority students say that quality of Free WIFI at SALU Khairpur is Fair.

Table 16
Contacting to Professors for Study Problems Via WIFI at SALU Khairpur

	Frequency	Percent	Valid Percent	Cumulative Percent
Yes	41	23.4	23.4	23.4
No	80	45.7	45.7	69.1
Sometimes	54	30.9	30.9	100.0
Total	**175**	**100.0**	**100.0**	

Without Professors/Teachers students can't get the knowledge or understand any topic in which they are facing the problem. This question aim asked to students to know about, Is they are contacting to Lecturers or not? The question was asked as "Did you contact your teachers for study problems via WIFI at SALU Khairpur?" The students respond as 23.4% students respond as YES, 45.7% NO and 30.9% say we contact SOMETIMES.

Table 17
Types of Files/Lectures Downloaded through WIFI at SALU Khairpur

	Frequency	Percent	Valid Percent	Cumulative Percent
Videos	35	20.0	20.0	20.0
Slides	27	15.4	15.4	35.4
Pdf files	70	40.0	40.0	75.4
word doc	13	7.4	7.4	82.9
Other	14	8.0	8.0	90.9
All above	16	9.1	9.1	100.0
Total	**175**	**100.0**	**100.0**	

The desired Files downloaded by students the above Table 17 indicate that 20% students respond as they download Videos, 15.4% Slides (ppt), 40% PDF Files, 7.4% word Document (*.docx), 8% Other files and 9.1% students say that they download All above files (such as Videos, Slides, PDF files etc.). From above table we can see that majority students download PDF files which means they like to read books but probably they don't have approach to purchase in hardcopy that's why they download books or research material in softcopy.

CONCLUSION

The conclusion drawn from the study is that the use of Free WIFI at Shah Abdul Latif University (SALU) Khairpur is a major factor which is affecting the academic performance of students. The students achieve good CGPA, who used WIFI for academic purposes. Study also find that in order of speed students ranked for WIFI as "Fair", Mostly students Prefer the Net for study problems and they download Books/Journals for study and research also found that WIFI improves communication between Friends for study problems. Majority students about 84% connect with WIFI Via Smartphone, Students spending less time for their social life and they are using WIFI positively for study, which means they desire to increase their knowledge and academic record.

REFERENCES

1. Negus, K.J. and Petrick, A. (2009). History of wireless local area networks (WLANs) in the unlicensed bands, *Info*, 11(5), 36-56. DOI: 10.1108/14636690910989324.
2. O'Leary, T. and O'Leary, L. (2011). *Computer essentials*. New York: McGraw-Hill Education, International edition.
3. Tenopir, C., King, D.W., Boyce, P., Grayson, M., Zhang, Y. and Ebuen, M. (2003). Patterns of journal use by scientists through three evolutionary phases. *D-lib Magazine*, 9(5), 1082-9873.
4. Chapman, L. (2002). Russian roulette or Pandora's box: use of Internet as a research tool. Paper presented at VALVA 2002. *11th Biennial conference and Exhibition*, 6-8 Melbourne. Victoria, Australia.
5. Berson, M.J. (1996). Effectiveness of computer technology in the social studies: A review of the literature. *Journal of Research on Computing in Education*, 2(4), 486-499.
6. Oates, B.J. (2008). *Researching information systems and computing*, London: SAGE Publications.
7. Asdaque, M.M., Khan, M. and Rizvi, S. (2010). Effect of internet on the academic performance and social life of university students in Pakistan. *Journal of Education and Sociology*, 4, 21-27.
8. Moate, K.M., Chukwuere, J.E. and Mavhungu, M.B. (2017). The impact of wireless fidelity on students? Academic performance in a developing economy. In Proceedings of *International Academic Conferences* (No. 4607490), International Institute of Social and Economic Sciences.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 67-74

THE IMPACT OF USING MOBILE PHONE ON THE ACADEMIC PERFORMANCE OF STUDENTS

Rukhsana Mangsi[1], Karamullah Memon[2] and **Faisal Afzal Siddiqui[3]**
Department of Statistics, Shah Abdul Latif University, Khairpur, Pakistan
Email: [1]balouch2100@gmail.com
[2]karamullah.memon@gmail.com
[3]brc.khi@gmail.com

ABSTRACT

This paper concentrates on the impact of mobile phones on university students with specific conclusion to SALU university of Khairpur. Through survey data collected basis 200 sample haggard from different departments of University of Khairpur for this objective. The study used to describe and quantity materials for analysis. The results show the mostly university students purchasing expensive mobile phones and they use in examination hall as unfair sources, and also they use the dictionary and other illuminating functions. Parents of female students comfortable likely and give this technology their daughter to easily contact to each other. According to thinking of students say this technology is not just increase our knowledge but also improve the quantity of education level. Students of university use this new technology and get more and more benefit as maintain high level of education.

KEY WORDS

Cell phone, Social media, SPSS, Intellectual Disabilities, University students, Internet Addiction, Learning.

INTRODUCTION

Cell phone is an apparatus which has influenced our social contacts, studies/education framework, safety/security, business exercises and numerous different matters of life (Ling, 2003). It is one of those purchasers' merchandise which made its market quickly and supplanted different types of interchanges. Being generally utilized the world over similarly by rich and poor, it was formally presented in Pakistan in mid 1990s and was taken as materialistic trifle however it is currently a desperate need of greater part of Population. It was at first extremely costly with constrained rivalry in the market yet last on the opposition of the mobile phone creating organizations and specialist organizations have made it exceptionally shoddy for the shoppers. Its dispersion is faster among the arrangement of late advances with sound consequences for social settings and sweeping implications (Townsend, 2002). In Pakistan its supporters crossed the digit of 121.13 million in March, 2013. It contains 98.40 percent paid ahead of time and 1.60 percent postpaid supporters (Government of Pakistan, 2013). In Pakistan, each out of each 100 individuals, 59.1 have versatile or settled telephone benefit.

Cell phone is valuable for regular man for simple social contacts and social issues. It is a device for enthusiastic contact among individuals and individuals depend more on cell phone more than any other time in recent memory (Yang and Lay, 2012). It is useful for representative with the end goal to do numerous business bargains. The across the board utilization of cell phones has re-formed, re-sorted out and changed different social features (Ravichandran, 2009). It is utilized by individuals of all age gatherings however individuals of age aggregate 60 years an above are observed to be in less utilization of cell and the individuals who are under 25 years age every now and again utilize this quickly developing innovation (Wajcman et al., 2007). This might be because of the way that plans, highlights and functionalities of the cell phones changed significantly lately (Head and Ziolkowski, 2012). Research in the cell phone and its effects isn't exceptionally basic when contrasted with other innovative issues (Aoki and Downers, 2004). It is a hard undertaking to assess the social and financial results of cell phones (McGeehan, 2005).

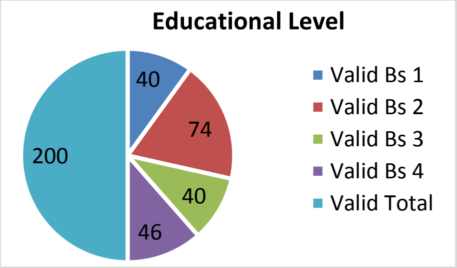
Cell phones are extremely regular specialized gadgets among the college understudies. Relatively every understudy of a college has at least one cell phone. Usually marvels among the young people (Campbell, 2006). It is influencing their social conduct, wellbeing and spending plan (Ravichandran, 2009). Anyway Ishii (2011) rejected the theory of unfriendly impacts of cell phones on young people in Japan. In any case, Jamal et al (2012) concur with the unfavorable impacts of cell phones on female understudies of Saudi Arabia. The utilization of cell phones is expanding cost of training as in china around 22% of the college understudies change their portable sets every year and 78% supplant it after each a few years. The cell phone arrange covers 92% territory of Pakistan (Govt. of Pakistan, 2012-13). This paper endeavors to analyze the impacts of cells on college understudies at Peshawar by accepting University of Peshawar as study case.

METHODOLOGY

The preliminary data get through to questionnaire papers. For survey researching 200 sample of students selected from different department in SALU of Khairpur. Random sampling technique was adopted to sample the students. After the collection of the data or information, the results were analyzed on the SPSS and result show in tables and corresponding their frequencies and percentages.

RESULTS & DISCUSSION

Frequencies



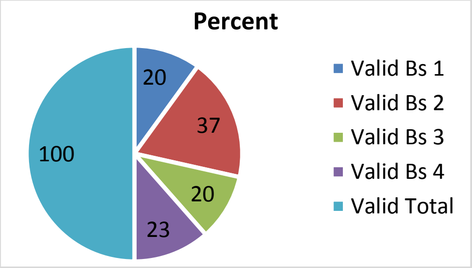


Table 1
Educational Level of Respondents.

Educational Level					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Bs 1	40	20.0	20.0	20.0
	Bs 2	74	37.0	37.0	57.0
	Bs 3	40	20.0	20.0	77.0
	Bs 4	46	23.0	23.0	100.0
	Total	**200**	**100.0**	**100.0**	

Main objective in this paper focusing on the benefits of mobile phone on the university students in Khairpur by get choice the University of Khairpur as a case study. The collection preliminary information for this objective from 200 students of separate level enrolled in university of SALU Khairpur. Above table showing the results with graphically all the students Education Level.

Education level of BS part-I is 20%
Education Level of BS part-II is 37%
Education Level of BS part-III is 20%
Education level of BS part-IV is 23%

Mobile phone is playing very important role in our society because by this we can contact anywhere in all over the world in some seconds as well as students studies. 99% students liking and taking mobile phone while 1% had not liking/taking/no phone present.

Table 2
What Type of Activity you do on your Phone during your Leisure or Free Time?

Frequencies

Phone Activity					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Face Book/Capturing Photos	41	20.5	20.5	20.5
	Gaming	57	28.5	28.5	49.0
	Music/WhatsApp	51	25.5	25.5	74.5
	Study	34	17.0	17.0	91.5
	Videos Movies	17	8.5	8.5	100.0
	Total	**200**	**100.0**	**100.0**	

The student’s daily routine activities more working either it study or some occasion/event. But in daily activities of students what time to save/receive for cellular phone in leisure time .20% students doing in free time of using face book and capturing photos, 28% students phone using for gaming, 25% students listening music and using WhatsApp, 17% students phone using for study and 8% students use on mobile phone for seeing the videos movies.

"According to students we make the some statements/questions and given some options. There are five options showing as 1 Strongly Disagree, 2 Disagree, 3 Neutral, 4 Agree and 5 Strongly Agree".

1st Statement:

The students send missed calls to class fellows to disturb during attending classes.

Frequencies

Miscall during Classes						
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Strongly Disagree	53	26.5	26.5	26.5
	Disagree	66	33.0	33.0	59.5
	Neutral	35	17.5	17.5	77.0
	Agree	42	21.0	21.0	98.0
	Strongly Agree	4	2.0	2.0	100.0
	Total	**200**	**100.0**	**100.0**	

The students show own view, according to 200 sample of university students presenting the 53 strongly disagree, 66 disagree, 34 neutral, 42 agree and 4 strongly agree. Show the mostly students Disagree this means many students does not like the disturbing during classes time.

2nd Statement is:

The student does not feel hesitation in sending interesting/funny SMSs to his/her teacher.

Frequency

Students End Funny Message Teacher					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Strongly Disagree	36	18.0	18.0	18.0
	disagree	68	34.0	34.0	52.0
	Neutral	43	21.5	21.5	73.5
	Agree	48	24.0	24.0	97.5
	Strongly Agree	5	2.5	2.5	100.0
	Total	**200**	**100.0**	**100.0**	

Some relationships among students and teacher like good, this reasons some students freely send funny SMSs without any hesitation. According to above table 36 strongly disagree, 68 disagree, 43 neutral, 48 agree and 5 strongly agree. 68 students disagree, this means many students show that the some space students and teacher relationship and not sending the any funny SMSs.

3rd Statement is:

The students feel proud of having costly Mobile phone.

Frequencies

Proud Costly Phone Use					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Strongly Disagree	22	11.0	11.0	11.0
	Disagree	42	21.0	21.0	32.0
	Neutral	30	15.0	15.0	47.0
	Agree	69	34.5	34.5	81.5
	Strongly Agree	37	18.5	18.5	100.0
	Total	**200**	**100.0**	**100.0**	

Base on 200 students of university the 69 students agree the purchasing costly cellular phone with feel proudly and 30 neutral, 42 disagree, 22 strongly disagree and 37 strongly agree. That mean mostly students likely the costly phone.

4th Statement is:

The students use Mobile phone in the examination Hall as a source of unfair means.

Frequencies

Phone Use Examination Hall					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Strongly Disagree	35	17.5	17.5	17.5
	Disagree	55	27.5	27.5	45.0
	Neutral	25	12.5	12.5	57.5
	Agree	66	33.0	33.0	90.5
	strongly Agree	19	9.5	9.5	100.0
	Total	**200**	**100.0**	**100.0**	

Mobile phone is very useful now that a time, depending user either is use positive way or negative. According to table 35 strongly disagree, 55 disagree, 25 neutral, 66 agree and 19 strongly agree. Analysis the above table showing majority (66) university students agree the using the mobile phone in examination hall as an unfair sources.

5th Statement:

The parents of female students feel easy due to Mobile Phone because their daughters can contact them easily.

Frequencies

Parents Contact their Daughter Easily					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Strongly Disagree	11	5.5	5.5	5.5
	Disagree	14	7.0	7.0	12.5
	Neutral	16	8.0	8.0	20.5
	Agree	106	53.0	53.0	73.5
	Strongly Agree	53	26.5	26.5	100.0
	Total	**200**	**100.0**	**100.0**	

Many families not allow cell phone of the female students but actually is very helpful device of female students because then reach university they faces many issues that time female students contact their parents easily. According to above table show majority (106) students agree this thing.

6th Statement:

Students use dictionary/thesaurus/calculate of Mobile.

Frequencies

Use Dictionary of Phone					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Strongly Disagree	12	6.0	6.0	6.0
	Disagree	17	8.5	8.5	14.5
	Neutral	21	10.5	10.5	25.0
	Agree	101	50.5	50.5	75.5
	Strongly Agree	49	24.5	24.5	100.0
	Total	**200**	**100.0**	**100.0**	

Now a time newly brand of cellular phone very useful one benefits is that dictionary/calculator facility available. 12 strongly disagree, 17 disagree, 21 neutral, 101 agree and 49 strongly agree that 101 students agree and using the this facility of mobile phone.

7th Statement is:

The mobile phone is the wastage of time for students.

Frequencies

Phone wastage of time					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Strongly Disagree	14	7.0	7.0	7.0
	Disagree	58	29.0	29.0	36.0
	Neutral	61	30.5	30.5	66.5
	Agree	53	26.5	26.5	93.0
	Strongly Agree	14	7.0	7.0	100.0
	Total	**200**	**100.0**	**100.0**	

Often and mostly students use phone entertainment, listening music, watching movies etc. they waste of that time. Strongly disagree 14,58 disagree, 61 neutral, 53 agree and 14 strongly agree total out of 200 sample. The high proportion show the neutral basis of 200 sample it means the half students agree and half students does not agree this statement.

8th Statement is:

The students' academic performance has been increased due to this technology.

Frequencies

Students Study Increase by Phone					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Strongly Disagree	6	3.0	3.0	3.0
	Disagree	42	21.0	21.0	24.0
	Neutral	57	28.5	28.5	52.5
	Agree	75	37.5	37.5	90.0
	Strongly Agree	20	10.0	10.0	100.0
	Total	**200**	**100.0**	**100.0**	

Now a day cell phone technology is very helpful and majority persons get good benefit this things either is study level or other field work. Above table the high majority (75) express that mostly university students increasing education level through this technology.

9th Statement is:

The mobile phone has helped to improve the level of the quality of education

Frequencies

Phone Help Improve Study Level					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Strongly Disagree	6	3.0	3.0	3.0
	Disagree	20	10.0	10.0	13.0
	Neutral	42	21.0	21.0	34.0
	Agree	94	47.0	47.0	81.0
	Strongly Agree	38	19.0	19.0	100.0
	Total	**200**	**100.0**	**100.0**	

The cellular phone helped the students easily utilized the quality of knowledge and get more thing about education. The table show 94 of respondents were opinion that use cellular phone has improved the education level. While other hand show 6 strongly disagree, 20 disagree, 42 neutral and only 38 strongly agree.

CONCLUSION

According to the results mostly university students using the mobile phone for social contact. After analysis the results and survey questionnaire basis of 200% sample of students reaching this point of views the 66% university students does not like disturbing during attending classes, mostly students focusing study decrease then ring /vibration phone during classes so does not like this thing, 68% students disagree about the sending funny and interesting massages without any hesitation they views show that the some space between students and teacher with respect. Majority students like the costly mobile phone with feel proudly when purchasing costly phones. 66% students unfair sources using mobile phones on the examination hall and high percent agree this thing it means negative effects on the university students studies. The benefit one thing is that the female students using mobile phones they easily to contact with parents. It is use for dictionary, calculator plus internet. 94% say students mobile phone helpful for study level and also increase the self-confidence when newly document made own self and 61% students wastage of time according to students opinions.

It is deliver basis on students study they use mobile phone be positively and get more and more benefit this technology and remove the negative way of using cellular phone, avoiding use phones when not specific work to do and save the golden time.

REFERENCES

1. Aoki, K., Downers, E.J. (2004). An analysis of young people's use of and attitudes toward cell phones. *Telemetric and Informatics*, 20(4), 349-364.
2. Wajcman, J., Bittman, M., Jones, P., Johnstone, L. and Brown, J. (2007). *The impact of the mobile phone on work/life balance*. Canberra: Australian Mobile Telecommunications Association & Australian National University.
3. Campbell, R. (2006). Teenage girls and cellular phones: Discourses of independence, safety and 'rebellion'. *Journal of Youth Studies*, 9(2), 195-212.
4. Government of Pakistan (2013). Economic Survey of Pakistan, Ministry of Finance, Islamabad, Pakistan.
5. Head, M. and Ziolkowski, N. (2012). Understanding student attitudes of mobile phone features: Rethinking adoption through conjoint, cluster and SEM analyses. *Computers in Human Behavior*, 28(6), 2331-2339.
6. Ishii, K. (2011). Examining the adverse effects of mobile phone use among Japanese adolescents. *Keio Communication Review*, 33(33), 69-83.
7. Tessler, M., Jamal, A. and Robbins, M. (2012). New findings on Arabs and democracy. *Journal of Democracy*, 23(4), 89-103.
8. Ling, McGu, R., (2003). *The mobile connection*, (3rd Edition) New York: Morgan Kaufmann Publishing Co.
9. McGuigan, J. (2005). Towards a Sociology of the Mobile Phone. *Human Technology*, 1(1), 45-57.
10. Ravichandran, S.V. (2009). *Mobile Phones and Teenagers: Impact, Consequences and Concerns*. Parents (Doctoral dissertation, Unitec New Zealand).
11. Townsend, A.M. (2002). Mobile communications in the twenty-first century city. In *Wireless world* (pp. 62-77). Springer, London.
12. Yang, H.J., and Lay, Y.L. (2012). Factors affecting Student's Mobile Phone Dependence and Anxiety, Proceedings of the *World Congress on Engineering and Computer Science*, Vol. 2, San Francisco, USA.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 75-78

A CLUSTER RANDOMIZED CONTROLLED TRAIL FOR THE PREVENTION OF TRANSMISSION OF HIV IN KHAIRPUR CITY

Ahmed Hussain Shar[1], **Zaibun-Nisa Memon**[2] and **Faisal Afzal Siddiqui**[3]
Department of Statistics, Shah Abdul Latif University, Khairpur, Pakistan
Email: [1]shar.ahmedhussain123@gmail.com
[2]zaib.nisa@salu.edu.pk
[3]brc.khi@gmail.com

ABSTRACT

We have study on the impact of prevention transmission of HIV in the Khairpur city that we have find out the following results. This research most result a cluster randomized controlled trail support Civil Hospital Khairpur (CHK) of HIV knowledge see in the table parameter estimates in column coefficient the result most chances of HIV are age group 22-30 (1.095) and most chances of be HIV are them whose are children 5-7 (1.111) is coefficient Which is negatively result given table in coefficient column .

KEYWORDS

Birth, Health, Child, Vaccine, Family Planning.

INTRODUCTION

2.1 million children suffering from HIV (Human Immunodeficiency Virus) in the sub Saharan Africa (USAID, 2017). (1).

(Goga et al., 2014) the prevalence of MTCT of HIV in south Africa (SA) by 4—8 weeks postnatal in (2012-13) was 2.6% with Mpumalanga province having an MTCT rate of 1.5% (Goga et al., 2016) and surrounded by the same group ,collective MTCT was at 3 months (2.7) 6 months (3.5%) 9 months (3.7%) 12 months (3.9%) and 18 months (4,3%) (Goga et al., 2016). In the most recent public antenatal sentry HIV prevalence survey (2013) among women 15—49 years the national HIV prevalence was 29.7% and in Mpumalanga province 37.5% (National Department of Health South Africa, 2013). These high rates of HIV in rural areas a lot been attributed to provide and staff shortages as well as to limited access to care (6).

It is the virus that can lead to acquired immunodeficiency disease or ADIS, if not treated so one you get HIV you have it for life.

How many people suffering from HIV in Khairpur have constricted germ from HIV, (2) infected many causes of HIV they are road accident, Suring used another person, drinking the darrow, or used other person bled for shaving etc.

METHODS

This study get was a Civil Hospital Khairpur randomized controlled trail I used a questionnaire ask questions every male and situation of HIV(investigational) on the primary data in survey collected the data within 32 weeks with child, and 6 weeks ,6 months and 12 month postpartum. In count to assessment participants attend three group and one person analysis intercession (or time equal organize) sessions prior to birth, and two person counseling sessions postpartum (25).

RESULTS

Obligation and randomization Table 1 given blow summarizes civil hospital and patients identification male and female numbers the follow started from September 2018 to October 2018 and the trail ended in November this illustrated males 80% have HIV patients and females 70% resultant in 150 patients taken from Civil Hospital Khairpur.

Table 1
Baseline Demographic and Characteristics of Male and Female (N=150)
Total (N=150)

Age Group		3.(2.0)
	15-22	39(26.0)
	22-30	42(28.0)
	30-45	40.5(27.0)
	50-60	21(14.0)
	60	4.5(3.0)
Gender	Male	82.5(55.0)
	Female	67.5(45.0)
Education Level	No education	82.5(55.0)
	Middle	9(6.0)
	Metric	12(8.0)
	Inter	16.5(11.0)
	Graduation	15(10.0)
	Master	15(10.0)
Work	No Work	88.5(59.0)
	Job	61.5(41.0)
Mat- Status	Never Married	49.5(33.0)
	Married	100.5(67.0)
Month Exp	300-1000	78(52.0)
	1000-3000	54(36.0)
	3000-5000	12(8.0)
	5000-7000	6(4.0)
Month Income	10,000-30,000	76.5(51.0)
	30,000—50,000	51.(34.0)
	50,000—70,000	7.5(5.0)
	70,000—80,000	3(3.0)
No of Children	3-4	15(10.0)
	4-5	16.5(11.0)
	5-7	18(12.0)
	7-11	28.(19.0)
HIV Positive	Yes	100(100.0)
HIV Knowledge	Yes	13.5(9.0)
	No	136.5(91.0)
Other HIV Patient in Home	Yes	42(28.0)
	No	108(72.0)

Parameter Estimates

Gender		B	Std. Error	Wald	Df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
Male	Intercept	37.633	7204.101	.000	1	.996			
	[age_grp=]	19.614	.000	.	1	.	329803914.772	329803914.772	329803914.772
	[age_grp=15-22]	.306	1.559	.038	1	.845	1.357	.064	28.839
	[age_grp=22-30]	1.095	1.555	.495	1	.482	2.988	.142	62.964
	[age_grp=30-45]	.394	1.468	.072	1	.788	1.483	.084	26.332
	[age_grp=50-60]	.485	1.456	.111	1	.739	1.624	.094	28.160
	[age_grp=60+]	0[b]	.	.	0	.	.	.	.
	[month_exp_grp=1000=300]	-19.639	7204.100	.000	1	.998	2.956E-009	.000	.[c]
	[month_exp_grp=300=1000]	-21.005	7204.100	.000	1	.998	7.543E-010	.000	.[c]
	[month_exp_grp=3000=500]	-20.274	7204.100	.000	1	.998	1.567E-009	.000	.[c]
	[month_exp_grp=5000=700]	0[b]	.	.	0	.	.	.	.
	[month_income_grp=10000=30000]	-17.671	1.030	294.108	1	.000	2.116E-008	2.808E-009	1.594E-007
	[month_income_grp=30000=50000]	-19.544	1.231	251.874	1	.000	3.253E-009	2.912E-010	3.635E-008
	[month_income_grp=50000=70000]	-15.851	1.652	92.051	1	.000	1.306E-007	5.126E-009	3.329E-006
	[month_income_grp=6000=10,000]	-17.717	.000	.	1	.	2.022E-008	2.022E-008	2.022E-008
	[month_income_grp=70000=80000]	0[b]	.	.	0	.	.	.	.
	[No_children_grp=0=3]	.067	1.632	.002	1	.967	1.069	.044	26.192
	[No_children_grp=3=5]	.131	1.548	.007	1	.933	1.139	.055	23.699
	[No_children_grp=5=7]	1.111	1.563	.506	1	.477	3.038	.142	65.015
	[No_children_grp=7=9]	.544	1.566	.121	1	.728	1.724	.080	37.097
	[No_children_grp=9=11]	0[b]	.	.	0	.	.	.	.

a. The reference category is: Female.
b. This parameter is set to zero because it is redundant.
c. Floating-point overflow occurred while computing this statistic. Its value is therefore set to system missing.

DISCUSSION

This study assessed the crash of behavioral intervention to enhance HIV patients male and female given the effect of HIV in use from Civil Hospital Khairpur completed current an adding together effectual in HIV transmission several fundamentals many explanation for the HIV that the 2018 on put of alternative B+ programmed in district in which are male and female living at HIV virus obtainable lifelong. This research most result a cluster randomized controlled trail support Civil Hospital Khairpur (CHK) of HIV knowledge seen in the above table parameter estimates in column coefficient the result most chances of HIV are age group 22-30 (1.095) and most chances of be HIV are them whose are children 5-7 (1.111) is coefficient Which is negatively result given above table in coefficient column that male and female are no chances of HIV because that are some reasons.

CONCLUSION

The HIV (human immunodeficiency virus) is the most effective for are age group 22-30 (1.095) male and female and most effect are chance far married or maternal them have children are 5-7 (1.111) that resulted taken from (CHK).

REFERENCE

1. Goga, A., Jackson, D., Lombard, C., Ramokolo, V., Ngandu, N., Sherman, G., Puren, A., Doherty, T., Bhardwaj, S., Noveve, N. and Ramraj, T. (2016). South African PMTCT Evaluation Study Group. Highest risk of mother-to-child transmission of HIV or death in the first 6 months postpartum: results from 18 month follow-up of an HIV-exposed cohort, South Africa. 21st International AIDS Conference, Durban, abstract TUAE0106, 2016.
2. Goga, A.E., Dinh, T.H., Jackson, D.J., Lombard, C., Delaney, K.P., Puren, A., Sherman, G., Woldesenbet, S., Ramokolo, V., Crowley, S., Doherty, T., Chopra, M., Shaffer, N. and Pillay, Y. (2015). South Africa PMTCT Evaluation Team. First population-level effectiveness evaluation of a national programme to prevent HIV transmission from mother to child, South Africa. *J. Epidemiol Community Health*. 69(3), 240-8. https://doi.org/10.1136/jech-2014-204535.s
3. Goga, A.E., Dinh, T.H., Jackson, D.J., Lombard, C.J., Puren, A., Sherman, G., Ramokolo, V., Woldesenbet, S., Doherty, T., Noveve, N. and Magasana, V. (2016). Population–level effectiveness of PMTCT Option A on early mother–to–child (MTCT) transmission of HIV in South Africa: implications for eliminating MTCT. *Journal of Global Health*, 6(2), 020405. doi: 10.7189/jogh.6.020405
4. National Department of Health, South Africa (2015). *The 2013 National antenatal sentinel HIV prevalence survey in South Africa*. Pretoria: National Department of Health.
5. UNAIDS (2018). Fact Sheet – World Aids Day 2018, 2017 Global HIV Statistics. Joint United Nations Programme on HIV/AIDS (UNAIDS).
http://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 79-86

USE OF SOCIAL MEDIA AND ITS IMPACT ON STUDENT ACADEMIC PERFORMANCE: A STUDY OF THE STUDENTS OF SHAH ABDUL LATIF UNIVERSITY KHAIRPUR

Mohammad Shan[1], Mazhar Ali Noonari[2] and **Faisal Afzal Siddiqui[3]**
Department of Statistics, Shah Abdul Latif University, Khairpur, Pakistan
Email: [1]yaxun2255@gmail.com
[2]mazharalinoonari@yahoo.com
[3]brc.khi@gmail.com

ABSTRACT

Social media is one of the more largest network where millions of people links together in order to share data such as information, videos, chatting, feeling and ideas. This paper concentrates on the “use of social media and its impact on academic performance of Shah Abdul Latif University Khairpur”. Our data collected through by questionnaire paper and the study further confirmed that most of the students visits social media sites and spend thirty to three hours per day. The study also tells that the social media sites has effect the student’s academic performance and tells about there is a relationship between the time spent on social media sites and student’s academic performance. In this study primary data has been used and data collected through questionnaire method, and take 120 same of size. Conclusion was drawn with the help of tables generated through SPSS, these tables indicating the use of social media and its impact on the student academic performance of the student of Shah Abdul Latif university Khairpur.

KEYWORD

Social Media, internet use, academic performance, students, Khairpur.

INTRODUCTION

The word social media is combination of two words one is “social” and other is “media”. The term “social” refer to interacting with other people by sharing information with them and receiving information from them. The word “media” refer to an instrument of communication, like the internet (while tv, radio, and newspaper are example of more traditional forms of media).

From these two separate terms, we can pull a basic definition together: “Social media are web based communication tools that enable people to interact with each other by both sharing and consuming information.

Social media has exploded as a category of online discourse where people create content, share it, bookmark it and network at a prodigious rate. Because of its ease of use, speed and reach, social media is fast changing the public discourse in society and setting trends and agenda in topics that range from the environment and politics to technology and the entertainment industry (Asur and Huberman, 2010). In the last ten years, the online world has changed dramatically, thanks to the invention of social media, young men and women now exchange ideas, feelings, personal information, pictures and videos at a truly astonishing rate. Seventy-three percent of wired American teens now use social media websites (Oberst, 2010).

Martn, (2008) & Lusk, (2010) share the same concept of social media. To them social media is the use of Facebook, Blogs, Twitter, My Space and LinkedIn for the purpose of communication, sharing photos as well as videos. However, for the purpose of this study social media is captured within the use of internet through Facebook, WhatsApp, Twitter, Skype, Myspace as well as Yahoo Messenger for communication sharing of ideas, sharing of photos and videos by users. The increased use of Social Networking Websites has become an international phenomenon in the past several years. What started out as a hobby for some computer literate people has become a social norm and way of life for people from all over the world (Boyd. 2007). Teenagers and young adults have especially embraced these sites as a way to connect with their peers, share information, reinvent their personalities, and showcase their social lives (Boyd, 2007).

Many researchers such as Choney (2010), San Miguel (2009) and Enriquez (2010) studies on students' use of the social media sites revealed a negative effect of the use of social media sites on students' academic performance. Nielsen Media Research study conducted in June 2010 stated that almost 25% of students' time on the internet is spent on social networking sites (Jacobsen & Forste 2011).

Social networking websites provide tools by which people can communicate, share information, and create new relationships. With the popularity of social networking websites on the rise, our social interaction is affected in multiple ways as we adapt to our increasingly technological world. The way web users interact and talk to each other has changed and continues to change. These users now socialize through the internet and it takes away from the person socialization that has been around forever. Social networking websites have affected our social interaction by changing the way we interact face-to-face, how we receive information, and the dynamics of our social groups and friendships (Asur and Huberman, 2010).

The internet has created a platform for millions of computers at numerous sites in various countries, belonging to thousands of businesses, governments, research institution, educational institutions and other organizations to link up with one another. It provides very rich medium for information dissemination, exchange and collabo-rative interaction among individuals and computers without regards for geographical limitation of space (Ugwulebo and Okoro, 2016).

Social media change the world dramatically due to the invention of social media. The use of social media very common in the young generation where more than million people link together and share their ideas, feeling, information, videos, music, pictures and documents. The communication through internet and social media are little different as compared to person to person communication or face to face communication. Social media sites become more popular after the invention of Facebook, Myspace and WhatsApp. For example, Facebook has more than 500 million members and more than 80 percent approximately are Facebook users. Teenager and young adults are connected with each other through WhatsApp, Myspace, Facebook etc. and share ideas, information, and other important documents etc. YouTube are great resource to find picture and videos on a topic you are interested in. Many people are use the internet for their research.

OBJECTIVE OF THE STUDY

The objective of the study is to the use of social media and its impact on institution academic student performance at SALU.

The specific objective of this study is following.

- To examine the impact of the social media on the academic performance among the institution students.
- To identify the purpose obtained from using social media sites.
- To examine social media sites improve communication skills.
- To examine how many students are familiar with the social media networking.

Research Question

- Do you think social media effect student academic performance or work?
- For what purpose you used social networking sites?
- Do you think social media sites improves communication skill?
- What you say that most students are familiar with social media networking?

Hypothesis

H1: There is a relationship between time spent on social media and academic performance.

METHODOLOGY

The survey method was used to collect data and information from the students of Shah Abdul Latif University Khairpur during the year 2018, and questionnaire was designed to collect data and information.

Table 1
Favourite Social Media Site

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
WhatsApp	47	39.2	39.2	39.2
Facebook	43	35.8	35.8	75.0
Twitter	14	11.7	11.7	86.7
Gmail	2	1.7	1.7	88.3
Instagram	12	10.0	10.0	98.3
n0thing	1	.8	.8	99.2
YouTube	1	.8	.8	100.0
Total	120	100.0	100.0	

Field Data, 2018

From the table 1 respondent gave the above as their fovourite social media site. WhatsApp represents 39.2%, Facebook 43 representing 35.8%, twitter 14 representing 11.7%, Gmail 2 representing 1.7%, Instagram 12 representing 10%, nothing 1 representing .8% and YouTube 1 representing .8%. The analyzing show that the WhatsApp is the most favorite social media site among the students.

Table 2
Time Spend on Social Media per Day

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
30 mints-1 hour	29	24.2	24.2	24.2
2 hours	33	27.5	27.5	51.7
3 hours	20	16.7	16.7	68.3
4 hours	10	8.3	8.3	76.7
More than 4 hours	24	20.0	20.0	96.7
All Day	3	2.5	2.5	99.2
No Answer	1	.8	.8	100.0
Total	120	100.0	100.0	

Field Data, 2018

When the respondent asks the question how much they spent time on social media per day. 29 respondents representing 24.2% indicates that they spent thirty to one hour, 33 respondents representing 27.5% indicates that the spent two hours, 20 respondents representing 16.7% indicates that the spent three hours daily, 10 respondent representing 8.3% indicates that they spent four hours daily, 24 respondent representing 20% indicates that they spent five hours daily, 3 respondent representing 2.5% indicates that they spent all day and 1 respondent representing .8% indicates that they gave no answer. The analyzing show that 52% students use social media sites thirty to two hours daily.

Table 3
Effects of the Social Media on Student Academic Performance

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
Agree	66	55.0	55.0	55.0
Disagree	16	13.3	13.3	68.3
Not Sure/Certain	18	15.0	15.0	83.3
Partial	8	6.7	6.7	90.0
strongly Agree	12	10.0	10.0	100.0
Total	120	100.0	100.0	

Field Data, 2018

When the researcher asked the question of is the use of social media effect the student academic performance. 66 respondent representing 55% was agree, 16 respondent representing 13.3% were answered disagree, 18 respondent representing 15% were answered not sure, 8 of them (6.7%) were answered partial and also 12 of them (10%) were answered strongly agree. The table show that most of the respondents 55% agreed the use of social media effect student academic performance. This study confirms

Mehmood (2013) study that the use of social media effects the student academic performance.

Table 4
Social Media Improves Academic Performance

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
Agree	70	58.3	58.3	58.3
Strongly Agree	24	20.0	20.0	78.3
Disagree	11	9.2	9.2	87.5
Strongly Disagree	8	6.7	6.7	94.2
Not Sure	7	5.8	5.8	100.0
Total	120	100.0	100.0	

Field Data, 2018

When the researcher asked the question whether the social media improves student academic performance. 70 respondent representing 58.3% gave positive (agree) answer, 24 of them (20%) were strongly agree, 11 from them (9.2%) were gave negative (disagree) answer, 8 respondents representing 6.7% are strongly disagree and 7 of them (5.8%) were not sure about them. The above table show that 58.3% respondents are were agree that's show the use of social media improves the respondents academic performance.

Table 5
Most Students are Familiar with Social Media Networking

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
Agree	67	55.8	55.8	55.8
Strongly Agree	17	14.2	14.2	70.0
Disagree	13	10.8	10.8	80.8
Strongly Disagree	6	5.0	5.0	85.8
Not Sure	17	14.2	14.2	100.0
Total	120	100.0	100.0	

Field data, 2018

When the researcher asked the question whether the students are familiar with social media sites. 67 of them (55.7%) were answered agree, 17 of them (14.2%) were answered disagree, 13 of them (10.8%) were answered disagree, 6 of them (5%) were answered strongly disagree and 17 of them (14.2%) were answered not sure. The above analyzed table show that 55.8% most students are familiar with social media networking.

Table 6
Purpose of using Social Media Sites

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
Application/Music	9	7.5	7.5	7.5
Videos	14	11.7	11.7	19.2
Educational	30	25.0	25.0	44.2
Informational	52	43.3	43.3	87.5
For All	11	9.2	9.2	96.7
For Other Purpose	4	3.3	3.3	100.0
Total	120	100.0	100.0	

Field data, 2018

When the researcher asked the question whether purpose of using social media sites. 9 respondents representing 7.5% using social media sites for download applications, 14 respondents representing 11.7% using social media sites for download videos, 30 respondents representing 25% using social media sites for education works, 52 respondents representing 43.3% using social media sites for information, 11 respondents representing 9.2% using social media sites for all above options, 4 respondents representing 3.3% using social media sites for other purpose. The above table show that most of the students (43.3%) use social media sites for study purpose.

Table 7
Social Media Improves Communication Skills

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
Agree	67	55.8	55.8	55.8
Strongly Agree	36	30.0	30.0	85.8
Disagree	5	4.2	4.2	90.0
strongly Disagree	7	5.8	5.8	95.8
Not Sure	5	4.2	4.2	100.0
Total	120	100.0	100.0	

Field data, 2018

When the researcher asked the question whether the use of social media improves communication skills. 67 from them (55.8%) gave positive (agree) answer, 36 from them (30%) were answered strongly agree, 5 from them (4.2%) gave negative (disagree), 7 of them (5.8%) were strongly disagree and 5 of them (4.2%) respondents not sure that the use of social media improves student’s communication skills. The above table shows that most of the respondent approximately 56% agree that the use of social media improves student’s communication skills.

Table 8
Social Media have more Friends as Compared to Real Life

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
Yes	58	48.3	48.3	48.3
No	38	31.7	31.7	80.0
Not Sure	24	20.0	20.0	100.0
Total	120	100.0	100.0	

Field data, 2018

When the respondent asked the question whether do you have more friends as compare to real life. 58 respondent representing (48.3%) were answered yes, 38 respondent representing (31.7%) were answered no and 24 respondent representing were answered not sure. The above table show that 48.3% most of the respondent were agree about this question that they have more friends on social media sites as compared to real life.

Hi: There is a relationship between the use of social media and student academic performance.

Table 9
Correlations

		Time Spend on Social Media	Academic Performance
Time Spend on Social Media	Pearson Correlation	1	-.021
	Sig. (2-tailed)		.822
	N	120	120
Social Media Improves Academic Performance	Pearson Correlation	-.021	1
	Sig. (2-tailed)	.822	
	N	120	120

Field data, 2018 Correlation is significant at the 0.01 level (2-tailed)**

From the Table 9: There is a relationship between the time spent on social media and academic performance because of the our Pearson correlation is equal to one that's show that there is a strong positive relationship between academic performance (dependent variable) and spent time (independent variable) at a significant of 0.01 level. Therefore, spending more time on social media contributes to low academic performance.

CONCLUSION

The purpose of this study is to check or examine the use of social media and its impact on the student academic performance of the students of Shah Abdul Latif University Khairpur. The study further confirmed that most of the students visit on social media sites and spent thirty to three hours daily they visit to use WhatsApp and Facebook. The study also confirmed that most of the students are familiar with social media sites and also tells that social media sites improves communication skills of the students. As the result the study confirmed that the time spend on social media sites effect the student's academic performance and there is a strong positive correlation between the spend on social media and academic performance. The hypothesis Hi tested on time spend on social media and academic performance tells the strong positive relationship between them at level 0.01.

REFERENCES

1. Asur, S. and Huberman, B.A. (2010). *Predicting the Future with Social Media. Social Computing Lab: HP Labs*, Palo Alto, California. 1-8.
2. Boyd, D.M. (2008). *Taken out of context: American teen sociality in networked publics*. University of California, Berkeley. Available at: www.danah.org/papers/TakenOutOfContext.pdf. Accessed July 16, 2010
3. Choney, S. (2010). Facebook Use Can Lower Grades by 20 Percent, Study Says. http://www.msnbc.com/id/39038581/ns.technology_and_sciencetech_and_gadgets/. Retrieved 14/08/13.
4. Enriquez, J.G. (2010). *Facebook and Other Online Social Networking Sites Can Lower Grades*, Study Says. http://seerpress.com/facebook-and-other-online-social-networking-sites-can-lower-grades-studysays/6935/. Retrieved 14/08/13.
5. Jacobsen, W.C. and Forste, R. (2011). The wired generation: Academic and social outcomes of electronic media use among university students. *Cyber Psychology*, Behavior, and Social Networking, 14(5), 275-280.
6. San Miguel, R. (2010). Study on Facebook and Grades Becomes Learning Experience for Researcher. Tech News World. http://www.technewsworld.com/rsstory/66805.html?wlc=1286985671&wlc=12871 5471 Retrieved 20/08/13.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 87-98

THE ADVANCEMENT OF TECHNOLOGIES IN INFORMATICS

Nadia Qasim[1], **Muhammad Qasim Rind**[2]
and **Muhammad Saleem Sheikh**[3]

[1] Agilenex Enterprise Solutions Leeds, London
Email: nadia.nady@gmail.com

[2] National College of Business Administration and Economics,
ECC, Lahore, Pakistan. Email: qasim.rind@yahoo.com

[3] Preston University, Islamabad, Pakistan
Email: drsmsaleem@yahoo.com

ABSTRACT

This paper gives an overview of technological advancements in the field of informatics. It will touch future technologies to be emerged in computer hardware and software. The newest inventions, more developed software/hardware are replacing older ones daily. These technologies improve not only very common areas of our daily life, but also areas of education, health, production industries, and so on. Thus, recent advancements in computer and software technologies are the base for the society of tomorrow. The first informatics revolution was the microprocessor, the collections of millions of tiny circuits that serve as the “brains” of personal computers and that are embedded in an ever-expanding number of products, from video games, to cars, to refrigerators. The rapid advancements in fiber optic technologies have also played important role to the informatics revolution. The fiber optics technology enables data, including voices captured in digital form, to be converted into tiny pulses of light and then transmitted at high speeds through glass fibers wrapped into large capacity telecommunication cables.

Today's computers operate using semiconductors, metals and electricity whereas it is imagined that future computers might use atoms, DNA or light. Moore's Law predicts doubling, but when computers go from quartz to quantum, the factor will be off the scale. The Giant networks of computers will be in your clothing, your house and your car. In future there will be Quantum Computers, and DNA Computers.

A contributing factor to the growing technology sector is human capital. The majority of technological firms worldwide have leveled the baseline production of new technology to the point where they seek new areas of improvement for their products. Thus, it is concluded that the informatics revolution drives the extraordinarily rapid decline in the cost and rapid increase in the processing power of digital technologies.

KEY WORDS

Quantum Computers, DNA Computers, Optical Computers and Telecommunications Networks.

1. INTRODUCTION

Computer Science is a study of science related to computer and it is a huge field to be developed every day. The newest inventions, more developed software/hardware are replacing older ones daily. It is updating only in a few days. The first computer invented was named as Difference Engine. It is passed through millions of inventions to nowadays. The latest breakthrough in Information Technology are 3-D Metal Printers, Artificial Embryos, Sensing City, AI for everyone, Dueling Neural Networks, Babel-Fish air buds, Perfect Online Privacy, Genetic Fortune Telling, Material’s Quantum Leap. These are all new inventions in computer science till now.

The emerging information technologies are imaginatively novel advances that decision makers are just beginning to notice them. The emerging products or processes may have been invented some time ago, but now they are applied in practical life. This paper will explore emerging technologies to use in near future. (Kendall K. E., 1997). Over the last few decades, advancements in computer and software technologies have reached an impressive level. These technologies improve not only very common areas of our daily life, but also areas of education, health, production industries, and so on. Thus, recent advancements in computer and software technologies are the base for the society of tomorrow (K. K. Mishra, 2014).

The transformation of the technology sector in the U.S. market resulted in need for software developers, computer and information systems managers, and computer systems analysts. New jobs such as these are commodities in the globalized world of technology, especially for companies recruiting individuals from technologically advanced countries. The growing market for tech jobs will continue to increase as technologies become even further integrated into society. More and more jobs will become available to individuals that obtained degrees in technology orientated fields.

The Information technology has been one of the most encouraging research areas throughout the globe over the past two decades. It is being applied in private and public sector organizations in the field of commerce, trade, healthcare, education, entertainment, environmental management etc. Information technology is indispensable and will continue to fuel further advances in all facets of human endeavors.

2. ADVANCEMENT IN PC'S SOFTWARE

Computer Software refers to the instructions and commands needed to run computer hardware and to perform expected services. There are many classifications of software products, but they usually fall in two main categories: System software and Application software. System software is usually the low level software required to manage computer resources and support the execution of application programs. Application software is software that performs specific functions needed directly by the end users.

System software products include operating systems, network operating systems, data base management software, programming languages software and other software development tools. Application development products include end-user applications, office automation software and various utilities. Some software products may fall into more than one of these categories.

Network operating systems are a subset of operating systems. Their role is to start-up, perform and monitor communications equipment and networks. Most of these systems are proprietary but perform services defined by standards such as TCP/IP (Transmission Control Protocol/Internet Protocol), which are a suite of communications protocols used to connect computers on the Internet. Database management software (DBMS) defines structures and tools to enable and control storage, modification, access and retrieval of data stored in a computer data base. Once exclusively proprietary, open source DBMS such as My SQL are increasingly used to create large-scale EMB applications

2.1 Entry of Window-8 for 64-Bit Processor

In 1986, Intel introduced first 32-bit CPU which was fully operated by 32-bit Windows Operating System, and Windows NT 3.1 operating system. After 1993, the 64-bit processor was introduced in the desktops and notebooks computer systems, but Microsoft didn’t produce to utilize fully 64-bit processor Windows. The 64-bit versions of Windows came later with Windows XP. In fact, virtually every computer sold today has a 64-bit processor under the hood. Later on Micro soft developed Window 7 to utilize full capacity of 64 –bit processor. Now Microsoft has switch over to 64-bit exclusively with Windows 8. Meanwhile, Mac OS X Leopard is already on 64-bit processor, and some hardware manufacturers are currently trying to transition customers to 64-bit versions of Windows. In future 128-bit processor computing machine will be introduced, it will be a big jump. Let's tackle one sea change at a time--and prepare for that move around 2025.

2.2 Entry of Google's Desktop Operating System

The gOS Linux is independent operating system which is built around Google Web apps. Now Google has concentrated to develop software from Web browsers to cell phones, time has come that are spending most of your time in the Google verse and never have to leave it. Google Checkout provides an alternative to PayPal. Street View is well on its way to taking a picture of every house on every street .Though Google seems to have covered everything, many observers believe that logically it will next attempt to attack one very big part of the software market: the operating system. The Chrome browser is the first toe Google has dipped into these waters.

2.3 Seamless Voice Recognition Software

Voice recognition software is interaction of humans and machines. A search for “voice recognition” on Google delivers more than 9.5mm results and most of these are recent results focused on companies who are diving into this category. They are raising money to solve problems where voice recognition is a path towards a solution. Voicea is one of these companies that focus on voice recognition to deliver a better experience for people in meetings by offering Eva, an in-meeting virtual assistant. Eva listens, records the conversation and provides a means to take notes, identify action items and organize follow up that can be delivered via email or pushed into tools like Sales force and Slack. Most of the population are using voice technology services such as Apple Siri, Google

Home, Alexa, etc. Thus voice recognition technology has always been viewed as a helping to assist the Management. But all that may change in future with the rise of sophisticated voice and speech recognition technologies.

2.4 Incoming Driverless Cars

The onset of driverless cars is becoming more of reality day by day. As safety increases in these autonomous vehicles, vehicle owners are taking more notice and considering the possibilities ahead. With manufacturers such as Tesla, Uber and Google leading the charge, it is only a matter of time before one drives into your neighborhood.

Offering an abundance of features, these driverless cars are sure to intrigue vehicle owners and turn heads on the road. Additionally, self-driving vehicles could reduce traffic fatalities by roughly 90%, saving hundreds of thousands of lives and more than $190 billion in healthcare costs every year, according to a report by consulting firm McKinsey & Company. Members of the Forbes Technology Council provide a glimpse of what consumers can expect from driverless cars that they aren’t able to take advantage of in their current vehicles. Driverless cars used to be confined to the realm of science fiction, but now they're set to come to a road near you, with the likes of Tesla, Mercedes-Benz, BMW, Google and Audi among the companies with systems ready to deploy. From reducing the number of accidents, to improving emissions compliance and easing congestion, driverless cars are claimed to cure a number of ills, and to assess how likely that is, In Ul driverless cars are under trial in Bristol, Greenwich and Milton Keynes cities. Some of the most amazing features of driver less car will be as under:

i. **Listening and Reacted to Specialized Sounds**
Driving is a comprehensive undertaking which involves many of driver’s senses working in tandem to allow the car to be safely driven on the road. One of the most complex and hardest to emulate senses among these is the sense to hear and understand sounds and driverless cars can do that equally well. Waymo, Google’s self-driving arm, just put forward a magnanimous display of this very capability in one of its recent tests. It was “teaching” its autonomous cars to learn to understand, react and give way to emergency and law enforcement vehicles, in an unprecedented display of the capabilities of the driverless cars.

ii. **360-Degree Camera:**
To be highly efficient in maneuverability and control, driverless cars need to have a full perspective on all the sides of the car. Driverless cars nowadays are coming with fully automated 360-degree camera systems that can manage to give excellent feedback regarding other surrounding cars and trucks, near the car to make it easier for the onboard system to navigate and drive the car forward in a safe way.

iii. **Recognition of Different Objects on Road**:
Not everything on the road looks similar. There can be humans walking by as pedestrians, dogs playing the street corner, unheralded stones lying on the road among others, and recognizing each one accurately and then taking necessary action, not just to protect the car, but also the living things in front of it is no easy feat. Machine learning has allowed these driverless cars to be exemplarily capable of doing exactly that, by amassing a massive data base of all

things that can be found on a normal or even an extreme road setting, and then learning to react accurately according to the situation. Extensive tests are being run by major companies developing this technology to hone this feature towards perfection.

iv. **Platooning**

It's common knowledge that reducing aerodynamic drag can help immensely in saving a whole lot more fuel, but even our aero dynamic designs have reached their threshold in the amount of aerodynamic drag they can neutralize. Enter driverless cars, and this fact can be soon translated into something tangible advantageous. Platooning is a concept in which cars can tail behind one another to reduce aerodynamic drag for the vehicles in the middle of the pack to a wide extent and save fuel. Driverless cars can easily tow behind one another and maintain a certain distance because of their supreme coordination and communication prowess that will allow different cars in a single platoon to act as one. This feature will be the most beneficial for Driverless trucks, who travel for extremely long distances, and could save up a hefty amount through fuel efficiency.

The features of driverless cars are, even now, highly impressive, but they need to be much more advanced if they are to make it safely and economically on to roads and into public use on a wider scale and there are many manufacturers, tech firms, and even some governments, who are pushing the envelope towards trailblazing this lucrative path and trying to be the first one to bring commercially viable autonomous cars on to roads within the next decade.

2.5 Edge Computing Software

Edge computing is a way to streamline the flow of traffic from IoT devices and provide real-time local data analysis. Edge computing allows data produced by internet of things (IoT) devices to be processed closer to where it is created instead of sending it across long routes to data centers or clouds. Edge computing is rapidly becoming a key part of the Industrial Internet of Things (IoT) equation to accelerate digital transformation. Edge computing is not a new concept, but several trends have come together to create an opportunity to help industrial organizations turn massive amounts of machine-based data into actionable intelligence closer to the source of the data.

Doing this computing closer to the edge of the network lets organizations analyze important data in near real-time which is a need of organizations across many industries, including manufacturing, health care, telecommunications and finance. In most scenarios, the presumption that everything will be in the cloud with a strong and stable fat pipe between the cloud and the edge device. The edge computing is the key factor in making IoT (Internet of Things) work. There has been an unprecedented rise in data generation which has led to an increase in demand for means of gathering, storing and utilizing all of that data in more compatible and centralized means.

2.6 Prescriptive Analytics Software

Predictive analytics uses data mining, machine learning and statistics techniques to extract information from data sets to determine patterns and trends and predict future outcomes. The future of business is never certain, but predictive analytics makes it clearer. Incorporating this software into your business is a sure way of taking a peek into what is likely to happen beyond the present and manipulating it to your advantage. It uses different methods of analyzing and interpreting various data to create a near accurate forecast of events that are likely to occur in some sectors of your business which impacts its overall performance. Predictive analytics can be used to gauge future customer or client reactions to specific measures that your company takes, and this is important for your customer relations management (CRM). It not only helps make decisions that maintain good customer relations but also captures their loyalty and identify important trends. The software can also be used for various purposes such as determining the impact of taking certain risks as well as manipulating the market to maintain a strong competitive advantage. The list of top most Predictive Analytics Software is given as under:

RapidMiner Studio, KNIME Analytics Platform, SAP Predictive Analytics, IBM Predictive Analytics, Dataiku DSS, Microsoft Azure Machine Learning, Microsoft R, SAS Predictive Analytics, DataRobot, Google Cloud Prediction API, Anaconda, Oracle Data Mining ODM, Angoss Predictive Analytics, Minitab, STATISTICA, H2O.ai, Alteryx Analytics, AdvancedMiner, Lavastorm Analytics Engine, GoodData, Analytic Solver, ABM, Portrait Predictive Analytics, MATLAB, Mathematica, TIMi Suite, Predixion Insight, DMWay, Grapheur, QIWare, CMSR Data Miner Suite, Skytree, Emcien, DataRPM, Feature Labs, RapidMiner Server, TIBCO Spotfire, GMDH Shell, HP Haven Predictive Analytics, GraphLab Create, Information Builders WebFOCUS Platform, Actian Analytics Platform, Alpine Chorus, Viscovery Software Suite, FICO Model Central, Rapid Insight Veera, Salford Systems SPM, Vanguard Business Analytics Suite are the top predictive analytics software. The large organizations and companies started working on Big Data and created their way to be better at predictive analytics. This was helping medium and large businesses tremendously for improving customer experience, processes and many other things.

3. ADVANCEMENT IN PC'S HARDWARE

Computer hardware refers to computing devices and the equipment attached to them. These may include personal computers (PCs), medium range servers, legacy mainframes (large computers that were predominant up to the 1990's), storage devices, visual presentation devices, Internet/communication equipment, imaging/printing devices, power supply management and personal identification equipment. Computer hardware evolves very quickly offering increasing capacity and lower acquisition cost. As a general trend, tasks earlier done by mainframes and medium range servers are gradually being done by smaller but powerful PCs.

3.1 Entry of Memristor Circuit

This simple memristor circuit could soon transform all electronic devices. There are three types of circuit components--resistors, inductors, and capacitors. The fourth type of component, one that would be able to measure the flow of electric current: the memristor. Now, Hewlett-Packard has built the memristor can "remember" how much current has passed through it. And by alternating the amount of current that passes through it, a memristor can also become a one-element circuit component with unique properties.

Memristors will theoretically be cheaper and far faster than flash memory, and allow far greater memory densities. They could also replace RAM chips as we know them, so that, after you turn off your computer, it will remember exactly what it was doing when you turn it back on, and return to work instantly. This lowering of cost and consolidating of components may lead to affordable, solid-state computers that fit in your pocket and run many times faster than today's PCs.

3.2 The 32-Core Central Processing Unit

A central processing unit (CPU) is the electronic circuitry within a computer that carries out the instructions of a computer program by performing the basic arithmetic, logic, controlling, and input/output (I/O) operations specified by the instructions With the gigahertz race largely abandoned, both AMD and Intel are trying to pack more cores onto a die in order to continue to improve processing power and aid with multitasking operations. Miniaturizing chips will fit these cores and other components into a limited space. A six-core CPU based on the Itanium design should be out imminently, when Intel then shifts focus to a brand-new architecture called Nehalem, to be marketed as Core i7. Core i7 will feature up to eight cores, with eight-core systems are available.

3.3 Nehalem and Swift Chips

When Advanced Micro Devices, Inc. , (AMD) purchased graphics card maker ATI Technologies Inc.(ATI) was a semiconductor technology corporation based in Markham, Ontario, Canada, that specialized in the development of graphics processing units and chipsets, most industry observers assumed that the combined company would start working on a CPU-GPU fusion. Chip makers expect the performance of such on-die. A graphics processing unit (GPU) is a specialized electronic circuit designed to rapidly manipulate and alter memory to accelerate the creation of images in a frame buffer intended for output to a display device. GPUs are used in embedded systems, mobile phones, personal computers, workstations, and game consoles. GPUs to fall somewhere between that of today's integrated graphics and stand-alone graphics boards--but eventually, experts believe, their performance could catch up and make discrete graphics obsolete. One potential idea is to devote, say, 4 cores in a 16-core CPU to graphics processing, which could make for blistering gaming experiences.

Intel's soon-to-come Nehalem chip includes graphics processing within the chip package, but off of the actual CPU die. AMD's Swift (aka the Shrike platform), the first product in its Fusion line, reportedly takes the same design approach, and is also currently on tap for 2009. Putting the GPU directly on the same die as the CPU presents challenges--heat being a major one--but that doesn't mean those issues won't be worked out. Intel's two Nehalem follow-ups, Auburndale and Havendale, both slated for late 2009, may be the first chips to put a GPU and a CPU on one die, a die is a specialized tool used in manufacturing industries to cut or shape material mostly using a press. Like molds, dies are generally customized to the item they are used to create.

3.4 Desktop External Hard Drives: USB 3.0

In order to enhance the transfer speed of USB the USB Implementers Forum, Inc., devised the new generation of USB: SuperSpeed USB 3.0. While interaction between the user and computer shouldn't change much with USB 3.0, data transfer speeds have skyrocketed from 480MB per second to 4.8GB per second. Just imagine filling your newly purchased desktop external hard drive with thousands of files in mere minutes or even seconds. SuperSpeed USB 3.0 has redefined speed to a whole new level.

Thus, USB connector has been one of the greatest success stories in the history of computing, with more than 2 billion USB-connected devices sold to date. But in an age of terabyte hard drives, the once-cool throughput of 480 megabits per second that a USB 2.0 device can realistically provide just doesn't cut it any longer. USB 3.0 (aka "SuperSpeed USB") promises to increase performance by a factor of 10, pushing the theoretical maximum throughput of the connector all the way up to 4.8 gigabits per second, or processing roughly the equivalent of an entire CD-R disc every second. USB 3.0 devices will use a slightly different connector, but USB 3.0 ports are expected to be backward-compatible with current USB plugs, and vice versa.

3.5 Wearable Computers

Margaret Rouse defined wearable computer as an electronic device capable of storing and processing data that is incorporated into a person's clothing or personal accessories. Some wearable computers are basically desktop or notebook computers that have been scaled down for body-wear. Others employ brand new technology. Both general and special purposes are envisioned. A number of wearables have been designed for the disabled. Among the challenges of wearable computers are: how to minimize their weight and bulkiness, how and where to locate the display, and what kind of data entry device to provide. Some of the applications envisioned for wearable computers include:

- Augmented memory, a concept originated by Thad Starner and being developed by Bradley Rhodes at the MIT Media Lab, in which as you enter a room, your wearable computer could sense the people present and remind you of their names or personal history, or a scheduler could whisper the time of an important meeting in your ear, or a "remembrance agent" could look for related documents by observing the words you were typing
- Immediate access to important data for anyone whose occupation requires mobility, such as real estate agents, rural doctors, fire and police professionals, lawyers in courtrooms, horse bettors, military personnel, stock brokers, and many others
- The ability to take notes immediately. For example, for reporters, geologists, botanists, vendor show representatives, field service repair personnel.

Thus, Wearable computer is a computer for carrying on the body, for example, on the wrist. It allows you to work, communicate, entertain right along ensuring mobility and hands-free and/or eyes-free access to the device. The latest trend in computing is wearable computers. Essentially, common computer applications (e-mail, database, multimedia, and calendar/scheduler) are integrated into watches, cell phones, visors and even clothing.

3.6 Quantum Computers

A quantum computer is a computer that makes direct use of distinctively quantum mechanical phenomena to perform operations on data. In a classical computer, the amount of data is measured by bits; in a quantum computer, the data is measured by qubits. The basic principle of quantum computation is that the quantum properties of particles can be used to represent and structure data and that quantum mechanisms can be devised and built to perform operations with these data. The Bloch sphere is a representation of a qubit, the fundamental building block of quantum computers. A quantum computer is a device for computation that makes direct use of distinctively quantum mechanical phenomena, such as superposition and entanglement, to perform operations on data. In a classical (or conventional) computer, information is stored as bits; in a quantum computer, it is stored as qubits (quantum binary digits).

A quantum computer is a device that performs quantum computing. Such a computer is completely different from binary digital electronic computers based on transistors and capacitors. Whereas common digital computing requires that the data be encoded into binary digits (bits), each of which is always in one of two definite states (0 or 1), quantum computation uses quantum bits or qubits, which can be in super positions of states. A quantum Turing machine is a theoretical model of such a computer and is also known as the universal quantum computer.

3.7 The DNA Computers

The concept of DNA computing was first introduced in 1994. It deals with the "biochips" made of DNA that are able to perform billions of calculations at once by multiplying themselves in number. In other words, a DNA Computer grows as it computes. In a recent development, the researchers from the University of Manchester have shown that the creation of this conceptual computer is possible in real life.

DNA computing was invented by the famous cryptographer Leonard_Adleman_who used DNA to solve the "traveling salesman" problem. The problem aimed at finding out the shortest route between a numbers of cities by going through each city only once. Adleman showed that billions of molecules in a drop of DNA had so much computational power that can simply overpower silicon and the powerful human-based computers. **DNA computing** is a branch of computing which uses DNA, biochemistry, and molecular biology hardware, instead of the traditional silicon-based computer technologies. Research and development in this area concerns theory, experiments, and applications of DNA computing.

3.8 Optical Computers

The computers we are using today consists of transistors and semiconductors to control electricity. Computers of the future may utilize crystals and metamaterials to control light. Optical computers make use of light particles called photons. Light travels at 186,000 miles per second. That's 982,080,000 feet per second -- or 11,784,960,000 inches. In a billionth of a second, one nanosecond, photons of light travel just a bit less than a foot, not considering resistance in air or of an optical fiber strand or thin film. Just right for doing things very quickly in microminiaturized computer chips. Newer advances have produced a variety of thin films and optical fibers that make optical interconnections and devices practical. Scientists are focusing on thin films made of organic molecules, which are more light sensitive than inorganics. Organics can perform functions such as switching, signal processing and frequency doubling using less power than inorganics. Inorganics such as silicon used with organic materials let us use both photons and electrons in current hybrid systems, which will eventually lead to all-optical computer systems.

4. ADVANCEMENT IN MOBILE PHONES

Mobile Computing is using a computer (of one kind or another) while on the move. With this definition, all three of the above examples will be covered by the term Mobile Computing. Mobile Computing is when a (work) process is moved from a normal fixed position to a more dynamic position. The 5G wireless is the fifth-generation of wireless cellular service. It will be faster than the 4G technology. The 4G is a big step up from 3G and is up to 10 times faster than 3G services. Now the 4G technology is reached to its peak, it is predicted that 5th generation communication technology will come in 2020.

4.1 Open Access Wireless Network

The reason most cell phones are so cheap is that wireless carriers subsidize them so you'll sign a long-term contract. Open access could change the economics of the mobile phone (and mobile data) business dramatically as the walls preventing certain devices from working on certain networks come down. We could also see a rapid proliferation of cell phone models, with smaller companies becoming better able to make headway into formerly closed phone markets.

4.2 Touch Screen Cell Phones

Last year Microsoft introduced Surface, a table with a built-in monitor and touch screen; many industry watchers have seen it as a bellwether for touch-sensitive computing embedded into every device imaginable. Surface is a neat trick, but the reality of touch devices may be driven by something entirely different and more accessible: the Apple iPhone.

4.3 Cell Phones Are the New Paper

Cell phone will work as plastic papers; you can drop paper boarding passes and event tickets and just flash your phone at the gate. Log in to your airline's Web site. Check in. Print out your boarding pass. Hope you don't lose it. Hand the crumpled pass to a TSA security agent and pray you don't get pulled aside for a pat-down search. When you're ready to fly home, wait in line at the airport because you lacked access to a printer in your hotel room. Applications to eliminate the need for a printout in nearly any situation are flooding the market. Cellfire offers mobile coupons you can pull up on your phone and show to a clerk; Tickets.com now makes digital concert passes available via cell phone through its Tickets@Phone service. The final frontier, though, remains the airline boarding pass, which has resisted this next paperless step since the advent of Web-based check-in.

4.5 Location Identifier Cell Phones

Right Now, only a handful of devices sport GPS service. In the near future, it will be the norm. Global Positioning System (GPS) is taking off, as phone makers, carriers, and service providers have realized that consumers generally have no idea where they are, ever. A location-based service (LBS) takes raw GPS data that pinpoints your location and enhances this information with additional services, from suggesting nearby restaurants to specifying the whereabouts of your friends.

5. SUMMARY AND CONCLUSION

The advancement in technology has been exceptionally fast in the 20th and 21st century. With electronic technology and machines being produced and improved all the time. Every day one company brings out something more advanced in an attempt to win the consumer war against another company. This consumerism is driving the rate of advancement faster and faster each year.

IT information has not only brought the world closer together, but it has allowed the world’s economy to become a single interdependent system. This means that we can not only share information quickly and efficiently, but we can also bring down barriers of linguistic and geographic boundaries. The help of information technology, communication has also become cheaper, quicker, and more efficient. We can now communicate with anyone around the globe by simply text messaging them or sending them an email for an almost instantaneous response. The internet has also opened up face to face direct communication from different parts of the world thanks to the helps of video conferencing.

Apple brought out a phone that could do everything and no other company in the mobile phone market could compare to it. We make phone calls, text, watch videos, take pictures and videos, listen to music, and connect to the internet, app’s for almost anything. We can download games, books, news, social networking sites in seconds. Of course this was a positive as people could use the phone to store data for work, education, entertainment. But the negative claims began as well, too expensive, quick battery power loss.

The rate of technological advancement is increasing with time, society is looking to create and develop easier ways to live and lengthen their lives. The internet is a massive source of information that millions of people use and depend on every day. It’s updated all the time from many places across the globe and is storing the history and heritage of the world that we live in.

Technology is good because it simplifies our work environment. Technology makes tasks easier, quicker, more efficient, and better. On the other hand, technology can make people lazier, products are made more cheaply which can be good for companies, but bad for consumers, and technology can have adverse effects on a person’s health and safety. Example of person health it might be eyes problems or headache because of spending time on computer without a break. Thus, Technology is all about what you do with it, its humanity’s choice on whether the advancement of technology will be used for positive or negative outcomes.

REFERENCES

1. Null, C. (2008). Hot New Technologies That Will Change Everything. *PC World*, OCT 28, 2008.
2. Parker, C. (2015). *The advancement of new technology*. *Positive or negative*? https://www.linkedin.com/pulse/advancement-new-technology-positive-negative-colette-parker.
3. Mishra, K.K., Misra, A.K., Mueller, P., Martinez Perez, G., Bhatia, S.K. and Wang, Y. (2014). Recent Advancements in Computer & Software Technology. *The Scientific World Journal*, Vol. 2014, Article ID 609512, 1 page, Hindawi Publishing Corporation, http://dx.doi.org/10.1155/2014/609512.
4. Kendall, K. E. (1997). The significance of information systems research on emerging technologies. *Decision Sciences*, 28(4), 775-792.
5. Perry, P. (2017). *Microsoft Plans to Have a DNA-Based Computer by 2020*. https://bigthink.com/philip-perry/microsoft-plans-to-have-a-dna-based-computer-by-2020,25 November, 2017.
6. *Future Technologies and Their Effect on Society*. https://www.futureforall.org/ computers/quantumcomputers.htm, 25th Nov. 2018.
7. ButtePatil, P.R. (2012). Technological Advancements and Its Impact on Humanity. *B. Tech-Agricultural Engineering*. (2012), 1-4.
8. Rosenberg, R.S. (2004). *The social impact of computers*, Elsevier Academic Press (2004, 3rd Ed.).
9. Yu F., Chang, C.C., Lu, Y., Shu, J., Gao, Y., Yue, G. and Chen, Z. (2014). Recent Advances in Information Technology. *The Scientific World Journal*, (6 pages), Article ID 746479, Vol. 2014. www.hindawi.com/journals/tswj/si/837060/cfp/
10. UK Essays (2018). Trends in Computer Hardware and Software Platforms Information Technology Essay. Retrieved on 01.01.2019 from www.ukessays.com /essays/information-technology/trends-in-computer-hardware-and-software-platforms-information-technology-essay.php?vref=1

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 99-112

COMPARATIVE ANALYSIS OF LOAD BALANCING ALGORITHMS IN CLOUD COMPUTING

Muhammad Qasim Rind[1], **S. Tauseef-ur-Rehman**[2] and **Nadia Qasim**[3]

[1] National College of Business Administration and Economics
ECC, Lahore, Pakistan. Email: qasim.rind@yahoo.com

[2] Ex-Professor, Preston University, Islamabad, Pakistan

[3] Agilenex Enterprise Solutions Leeds, London
Email: nadia.nady@gmail.com

ABSTRACT

Today cloud computing has become the cornerstone of information technology, the next stage in web development. It provides users with a great deal of computing and storage services over the Internet. The load balancing is necessary for important activities in virtual cloud environments. Due to extensive usage of cloud computing many customers around the world places a greater demand on the nature and types of services as well as the quality of services that are provided. To combat the needs load balancing is a key and hence has emerged as vital area of research. Thus various algorithms have been developed to cater for increasing demands from the clients in shape of availability of remote nodes. The effective workload balancing ensures an efficient use of resources for customers on demand and improves the overall performance of the cloud. This article provides a brief discussion of the current load balancing methods in the cloud computing process and compares them more based on a variety of parameters such as processing time, latency and response time, complexity and scalability.

KEY WORDS

Cloud computing, cloud services, workload balancing, balancing methods.

1. INTRODUCTION

Cloud computing is the most advanced technology in the IT industry and a new delivery method for cost-based services. According to various standards the definition of cloud computing can be briefly summarized as: "Cloud computing is an application, system and/or basic infrastructure which is as a service on the Internet that can be accessed from the web browser on pay-per-use basis.

Cloud computing has become one of the common technologies adopted by numerous industries, providing flexible and efficient file storage and retrieval [1]. The main problem is to create a schedule of requests included in a manner that is proportional to the minimal response and effective use of resources at a time and at the same time resources should not be used in full. The cloud computing system is heavily based on a virtualization where numerous logical servers are employed to offset load on a server. This approach improves the energy efficiency and allows virtual machines to access one physical server. Cloud computing is an Internet computing experience. The main theme is to provide all services dynamically when a user request where the range of services includes operating system, application development framework. network, or storage. These services are categorized as Infrastructure as a Service (IaaS), platform as a Service (PaaS), and Software as a Service (SaaS). Cloud computing is divided into three categories, the public private and hybrid cloud. The cloud computing window, as shown in Figure 1, shows that it contains three deployment forms and three service models, which are as follows.

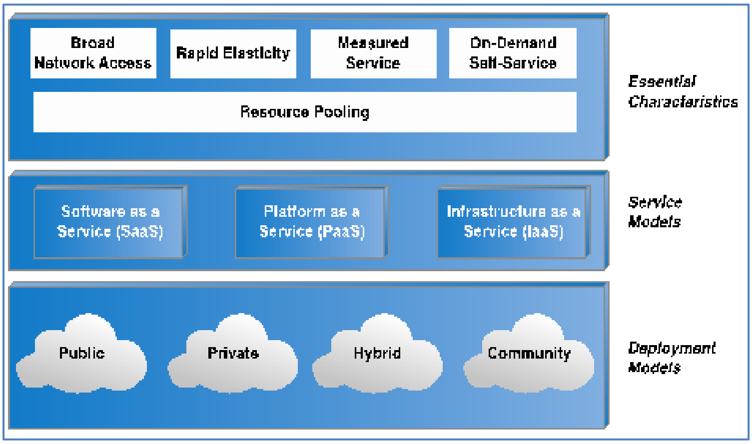


Fig. 1: Cloud Computing Framework [1]

1.1 Service Model

The service model refers to the domain of service provided by the cloud. It has three categories described below:

1.1.1 Infrastructure as a Service (IaaS)

The basic cloud service model is known as infrastructure as a service (IaaS). It offers hardware as a service organization, such as server space, network equipment, and session CPUs, virtual and actual hardware, and other resources.

1.1.2 Software as a Service (SaaS)

SaaS provides a variety of user services, such as applications through various cloud providers over the Internet. The client uses the program for each use. It is generally referred as "software on demand".

1.1.3 Platform as a Service (PaaS)

In Paas models cloud service providers provide platform computing such as operating system support and implementation of programming language and web server. Examples are GAE and Microsoft Azure [2].

1.2 Deployment Model

The deployment model is further categorized as:

1.2.1 Public Cloud

Available from anywhere in the world. An example of this cloud network on Google is available to anyone after the service level agreement between the service provider and the user. They are available after subscription.

1.2.2 Private Cloud

In this cloud, employees can access company data or hire a partner.

1.2.3 Hybrid Cloud

There is both a general and a special cloud combination.

2. LOAD BALANCE IN CLOUD COMPUTING

Adaptive Load Balancing is a technology focusing on the budgetary requirements. Here the term budgetary is used in context of time for execution or the resources utilized therein. Collectively the two terms have a combinatorial value termed as cost. Load balancing means provision of significant time resources and efficient use of resources through the adjustment of total loads between the various contracted clouds [3]It also encompass the ability to solve the problem of excessive use of virtual machines and use them incompletely. The term load balancing also embrace solving the problem of load balancing overload and aim at maximum productivity and better resource utilization and reduce response time. Load balancing is a prerequisite for maximizing cloud performance and efficient resource utilization.

The advantage of clouds has been enhanced by the resource allocation method and the ability to provide a deterrent task. From the resource allocation algorithm to the collateral adaptive system functions, the cloud is able to predict things, but this approach does not solve the problem of response and effective use of cost of time. Load balancing has two main functions, one is resource allocation or provision of other resources and second is planning in a distributed environment. Ensuring efficient resource and resource planning is in addition to the following tasks:

- Resources are readily available.
- Resources are used efficiently under low / high load conditions.
- Reduce resource usage costs.
- The load capacity increases the output speed to a maximum hour and a minimum response time.

Load balancing is an effective and critical cloud computing concept and helps optimize resource utilization, thus reducing resource consumption. So, one has to split the load on the structure of the contract based on the cloud, so that each resource is under the same amount of work at any moment. It is done with a load balancer. The load balancer determines that different requests are assigned to different servers. The load balancer uses another programming algorithm to determine the number and nature of the tasks to be received by the server.

3. CLASSIFICATION OF LOAD BALANCING ALGORITHM

With cloud computing, another algorithm was proposed for load balancing, such as load balancing techniques based on honey bees, taking random samples, assembling the asset, balancing and loads balancing, diluted, and WCAP, and JIQ and CLBVM. The main goal is to achieve high productivity and less response time. The load balancing algorithm usually has two types such as: static load balancing algorithm and dynamic load balancing algorithm.

3.1 Static Load Balancing Algorithm

The upload does not depend on the current state of the system; i.e.; the present status of q length etc., but requires knowledge of system deployment and resources. The load balancing algorithms are static in the sense that once adopted they do not evolve or change thereby dividing the load into a solid foundation for a set of rules, such as the amount of input work. There are four different types of static load balancing techniques: the Robin Round algorithm, the central manager algorithm, the threshold algorithm and the random algorithm.

3.2 Dynamic Load Balancing Algorithm

The dynamic algorithm is more flexible than a fixed algorithm is not dependent on prior knowledge, but based on the current state of the system. In a diffused system, there are two different methods for balancing dynamic load: distributed and not assigned. Distribute a unit to implement this algorithm with all the nodes in the system and a common task of loading between these servers balancing. Interactions between nodes in order to achieve a pregnancy balance can be of two types: cooperative and non-cooperative. In the first decade, the contract is working together to achieve the common goal of improving the overall response time, etc. In the second form, each node works independently of the local target.

4. TYPES OF LOAD BALANCING ALGORITHMS

In this paper, we have discussed main algorithms which are given below. Some of the algorithms have specific procedures and these are also described .

4.1 Round Robin VMLB Algorithms

It is a basic load balancing mechanism that maintains new fixes that reach virtual machines that can be accessed in a circular manner. The main benefit of this algorithm is its simplicity and ease of operation. The main problems are pre-defined knowledge of user functions and system resources. Moreover, the current state of the system is not compensated. [4]

4.2 Load-Alignment Virtual Throttle

In this dynamic approach, the user submits a request to the Data Center Console (DCC). Then the VM Load Balancer Data Center Controller asks you to identify the appropriate virtual machine that can be easily handled by the specific workload. VM Load Balancer maintains a list of virtual devices and its status (free / busy). If you see the amount of VM memory, cores, or access base, VM Load Balancer takes a cloud-shake request and assigned a cloud-based request to this virtual machine. Otherwise, the customer has to wait in line until the appropriate VM access is available. Among all, this is the best way to load balancing, because it supports the data center's current state of all VMs. But the main disadvantage is that it works correctly only if the entire data center's VM has the same hardware architecture. [4]

4.3 ESCE VM Load Balancing Algorithm

The ESCE implies the same implementation that is now comprehensive. It's also called the Active VM Load Balancing algorithm. This algorithm is based on spectrum distribution technology. Workload is also distributed among CCTV devices in the data center. Queue contains all Cloud lay requests to be performed by VM. ESCE VM Load Balancer (VMLB) also supports a list of virtual machines. VM Load Balancer continuously checks the task queue and VM list. If the VM is found free, the cloud request will be selected on the VM. At the same time, VMLB checks the congested VM. When a loaded virtual machine is found, then the VMLB puts a load in the standby mode or under the installed virtual machine to reduce the load from the overloaded VM. The main disadvantage is high arithmetic expenditure. [4]

4.4 HTV Balancing Algorithm with Increased Priority

This algorithm performs tasks on data center servers efficiently and reliably. This algorithm focuses on three parameters, such as server load, server statistics, and time constraints for tasks. The algorithm calculates the load factor and performance of each virtual machine, and has a significant access to many virtual machines depending on the completion time and the time it takes to increase performance and speed. [5]

4.5 Virtual Machine Algorithm and Physical Machine

In general, balancing, this would increase the use of material resources and reduce the energy consumption approach. Calculates the effectiveness of the policy compared to the current load balance access and value migration and energy consumption. The results of the experiment show that this method yields better results than current load balancing methods. [6]

4.6 Composite Procedure

Load balancing is necessary to provide the service provider with adequate resources. The load balancer is part of the cloud to spread the amount of work between many virtual machines on a network server to achieve optimal resource consumption, reduce processing time, low response time average, and prevent overload. This combined approach is effectively used to handle load-splitting methods well. This method benefits from the introduction of ESD and encoded algorithms. [7]

4.7 Dynamic Distributed Algorithm Based On Priority

It is used to efficiently balance load, improve system consistency, reduce response time and increase transmission speed. Prioritizing the allocation of virtual machine resources provides the best response time and processing time. Load balancing provides all instances of the network node to carry out the same workload at any time. Prioritize resources to improve the use of resources and minimize the response time of cloud services. [8]. The following are sub categories of Dynamic Distributed Algorithm Based On Priority.

4.7.1 Procedure: Burstness-Aware Load Balancing

This procedure takes two algorithms for load balancing. If RR is used for explosion, and Random is used not in consecutive position. Fuzzy logic-based validation is used to process a user request, and the request is submitted to a balanced VM. Experimental results show that the algorithm, in comparison with other algorithms, offers average response time and average processing time. [9]

4.7.2 Honeybee Behavior Load Balancing Method

High priority tasks are separated from the overloaded virtual machine and installed under the loaded virtual machine, taking into account the lowest similar priorities, cost effective virtual machine and the least possible completion time, which also balances the workload of Proactive. The minimum estimated cost, cost, and priority time during this task sends out to achieve a minimum completion time, a small task waiting time, and ultimately better access to resources. [10]

4.7.3 Procedure: Weighted Based Optimized Load Balancing

The loads are distributed among the servers that generate the VM by evaluating the weight as a parameter. The results are compared with Rubin and EIPR round robin algorithms. The simulation results confirmed that this algorithm equals the workload between virtual machines. [11]

4.7.4 New Enhanced Load Balancing Procedure

Firstly, the adaptive load balancing strategy was developed in accordance with the discoveries of genetics. Secondly, an improved load balancing strategy includes other constraints, such as suitability and selection of the initial set of resources. These limits significantly affect the algorithm's results. This advanced load balancing algorithm provides better results than the current genetic approach. [12]

4.8 Global Load Balancing Framework

An existing window representing the global server load balancing on sites in the cloud, using a two-step load balancing structure. This system represents a load balancing mechanism in various data centers managed by cloud service providers. [13] In addition, how a request is processed with or without priority in terms of workload, also as a client request and load balancer.

4.9 Procedure: Dynamic load balancing

Load balancing algorithms play an important role in balancing loads between data centers and using efficient computing resources. In this paper, the performance of a dynamic load balancing algorithm was assessed by dividing data centers in different regions. The proposed algorithm has been shown to improve the efficiency of data center computing and reduce user response time. [14]

5. RELATED WORK REGARDING LOAD BALANCING

Load balancing is a type or method, used to divide work across multiple computers over a band-based network connection to reduce the total return time. Load balancing eliminates performance deterrents such as resource congestion and traffic between servers and data, so data can be sent and received immediately. In order for the new broadcast to survive in the market for competitive information systems, it uses a cloud-based service to develop a widespread data center. Load balancing helps balance user requests in a distributed data center, so it's important to use load balancing on the logistics system's web servers. Meenakshi Sharma proposed a centralized distribution of load for virtual politics (CLBVM) to load balancing in a distributed environment, but fault tolerant systems were not considered resulting in skewed asymmetrical research.[17]. Bhupendra Verma analyzed load balancing virtual machine and their algorithm of load balancing virtual machine is restricted to frameworks having specific instance of agreement within the broader domain of IaaS. It introduces an active load balancing management weighted algorithm to achieve the best performance response time data. The results were collected using Cloud Sim processing time [18].

Dhaval Limbani Bhavesh Oza's Recommended policy and routing algorithm is based on requests from expanded user-driven users, for example, by choosing an efficient data center cost with the Cloud Analyst [19]. Neeraj Bhargava and Rita Bhargava et.al analyzed the round robin and discussed uses other data center algorithms. Their algorithm calculates the total response at best performance [20]. S. mohanum Priya and B. Subramani recommended using an active load balancing algorithm and resource planning algorithm familiar with high performance in cloud systems. The result of the proposed algorithm test is the effective virtual machine selection for the process and the minimum time for execution of the task. Increases performance and reduces response time and costs [21]. Harvinder Singh and Rakesh Chandra Jaanwar conducted a comparative study of pregnancy algorithms based on parameters such as budget, response, and fee reduced by Round Robin load balancing algorithms and asset monitoring time [22]. Mishra Sreenu Naik Bhukya's priority and algorithms and round-robin service is an expanded priority broker that distributes programs based on the classification of data centers and provide the best performance [23]. Deepak Capgit proposed and compared the new service brokerage algorithm (DC selection) with the existing service brokerage algorithm. The proposed algorithm reduced service response time and improved data center performance [24].

Slesha Nayak and Prangesh Patel presented a comparative study of existing throttle algorithms and the proposed shock algorithm for load balancing in the cloud computing process. Both have been tested and compared based on response time and data center costs and costs through Cloud Analyst [25]. Bhaon Mallik load and Garnetha based on criteria such as time data processing and response time, etc., using the Round Robin algorithm and throttle charts with the Cloud analyst [3]. Sarika Vasantrao Bodke offered a comparison between the various policy areas used in the compliance process with circular heartbeat and Robins and my first-time (FCFS) algorithms for carrying Equalizations. This comparison indicates that the response time is effectively reduced, but no cost is taken into account [26]. Mamta Kanchi and Sanya tyagi proposed a hybrid approach to balancing the virtual machine load. This algorithm simultaneously divides the workload between virtual machines available in the data center to reduce the total response time and processing time of the data center [27].

5.1 Problems with existing algorithms

In a widespread data center, the main offer of service brokers suggests leading the user's request for the best data center performance, since the best broker service policy is effective in choosing the best data center services given to many issues, such as response time to service and cost of time. To direct the user to the best data center, there are many load balancing algorithms such as network response time, service oriented routing, etc. Adjusted algorithm The Chromium Balancing Throttle and Load Balancing algorithm proposed a virtual machine that effectively reduces the response time but does not reduce costs [17] analyzed the Jasmin robin algorithm using other datacenters and user base

(UB). This result shows that the reaction time is minimal, but it should also reduce costs [20]. Priya has proposed a new load balancing algorithm in the virtual machine, which defines the process default machine and the minimum time to complete the task. Reduces response times and overall costs, but there are some issues to improve the efficiency of response time along with cost performance [21].

Harvinder analyzed three current load balancing algorithms: Robben's round, active monitoring, and throttle algorithm. The experimental result of these reduction algorithms appears in the response, but the increase in costs [22]. Rakesh The Robben ROW algorithm is a priority and extended priority that distributes requests based on the classification of data centers and provides better performance during the reaction but does not reduce the response time [23]. Deepak Capgit proposed and compared the new service brokerage algorithm (DC selection) with the existing service brokerage algorithm. The proposed algorithm reduced the response time of the service, but costs were not taken into account [24]. Slesha offered a throttle algorithm and compared it with the existing throttle algorithm. The proposed algorithm reduces the response time and costs for better performance, but as the data center increases, response time costs will increase [25].

Rina discussed various load balancing algorithms and compared them based on criteria such as time data processing and response time, etc. using a Round Robin algorithm and a throttle graph with the Cloud analyst. The result of both algorithms shows low response time and cost, but if data centers are increased, response time and costs will increase with the use of both individual and butterfly algorithms [2]. Podcas also introduced a roundabout algorithm. In this algorithm, time is divided into several segments, and each node is given a given time slicing. Several operations have multiple requests. Each slot action is given. If the user's request is completed over time, the user should not expect or the user will need to monitor the next slot. This may result in a slow process and overload [27].

Mamta's algorithm proposed a workload balancing virtual machine, which is a round robin combination, smothered (ESCE), and the uniform distribution of the current implementation hybrid algorithm. The result shows a decrease in the total response time, but overall performance is not significantly improved using the ESCE [2] algorithm. M has suggested. Moradi, M.A. Dezfuli, M.H. Safavi, load balancing algorithm with a new probability. It is proposed to create this algorithm to determine the load balancing and to reduce the response time and the main goal is to select resources based on the time at least the last and best of the finish [28].

In the Table 1 given below, you can see many different algorithms and defects. We have analyzed the results and responses and costs of different algorithms to determine Robin's robust circle and active monitoring algorithms and compressed time.

Table 1
Current Algorithm Simulation Results

Parameters	Algorithm						
	Round Robin	**ESCE**	**Throttled**	**HTV**	**Dynamic Priority Based**	**Composite**	**Global Frame-work**
Static/Dynamic	Static	Dynamic	Dynamic	Dynamic	Dynamic	Dynamic	Dynamic
Throughput	Low	Average	Average	Average	Good	High	High
Response time	Low	Average	Average	Average	Good	Good	Good
Scalability	Low	Average	High	High	High	High	High
Priority	Low	High	High	High	High	High	High
Fault tolerance	Low	Low	Average	Average	High	High	High
Overhead Cost	High	Average	Average	Average	Low	Low	Low
Virtual Machine Usage/hour	High	Average	Average	Average	Average	Average	Average
Power Consumption (Idle nodes off)	High	High	High	High	Average	Average	Average
Complexity	Low	Low	Low	Low	High	High	High
Fairness	Low	Average	Average	Average	Average	Average	Average
Performance	Low	Average	Average	Average	Good	Good	Good

The above tables show that seven algorithms such as: Robben's round, ESCE, Throttled, HTV, Dynamic Priority Based, Composite, and Global Framework w**ere** tested with 12 performance parameters. It is observed that Round Robin algorithm, is static whereas six other algorithms such as: ESCE, Throttled, HTV, Dynamic Priority Based, Composite, and Global Framework are dynamic in nature. The overall performance of the last three algorithms like Dynamic Priority Based, Composite, and Global Framework were treated as high. Thus these algorithms are better than the ESCE, Throttled, HTV.

6. PERFORMANCE ANALYSIS OF LOAD BALANCING ALGORITHMS

The next section discusses various load balancing algorithms and shows relative results. The investigated factors are discussed below: [29]

6.1 Throughput

It is the amount of work that can be done by all nodes over a given period of time.

Response time - the time elapsed between the inserted request and the start of the response after the task is completed. Simply put, the time it takes to answer is with a special load balancing algorithm. You need to reduce this parameter.

6.2 Scalability

Ability to apply a computer (hardware / software / service) to continue working effectively, even if it varies by size, topography, etc. Priority is the choice of tasks based on factors such as cost, time, size, etc. [30]

6.3 Fault Tolerance

The system is designed to withstand and continue to operate despite faults.

6.4 Overhead

Indicates the system processing time required for installation, operation or any transaction.

6.5 Cost

Costs associated with system configuration and processing of functions requested by users, better defined as hourly usage of the device.

6.6 Energy Consumption

Information technology consumes enormous energy and generates high energy costs. Efficient energy management is crucial for the success of the IT environment, such as cloud computing, network computing, and so on.

6.7 Complexity

It is making the whole system difficult. As the number of users associated with clouds and their characteristics increases, the system becomes more complex.

6.8 Fairness

Indicates that each user has the same response time and that all of his functions are completed at about the same time.

6.9 Performance

The speed and accuracy, after which the works are completed and measured according to predefined criteria. In simple words, it is the overall system efficiency. It can be improved by reducing the response time and waiting time of tasks, while maintaining a reasonable system cost.

Load balancing uses cloud computing and improves performance by dividing the load between processors. Transferring jobs from one node to another through a network that involves a delay (delays queue + processing delays), where it is to identify the target node through remote processing. [31]

Distribution system model employed considers n. users that consumes m number of computing resources. Nash equilibrium can be defined as a distribution system model as a "s" strategy for each user "u"

$$Su = ArgMinD_j(S_1S_2.........S_n)$$

Nash equilibrium is achieved if no user can reduce the average expected response time by unilaterally changing his strategy. The normal speed should send jobs immediately. When it receives the contract it will process. These processors maintain a specific queue for the job submitted. There is a finite amount of time that is consumed while transferring the job and the subsequent wait in the queue. Therefore

$$Total\ Response\ Time = Processing\ Time + Wait\ Time + Response\ Time$$

$$RelativeProcessingRate = \frac{Job\ Processing\ Rate}{Lower\ Processing\ Rateinthe\ Cluster}$$

$$Job\ Generation\ Rate = \frac{User's\ JobGeneration\ Rate}{Total\ Job\ Generation\ Rate\ of\ all\ Users}$$

The table below compares the various load balancing algorithms described above based on the qualitative data required for cloud computing.

Table 2
Comparison of Different Variations

Authors (Year)	Algorithms							Results	
	Round Robin	ESCE	Throttled	HTV	Dynamic Priority Based	Composite	Global Frame-work	Average Response Time (ms)	Total Cost ($)
Limabani & Oza (2012)	*	*	*					*	999.33
Jasmin James (2012)	√	√	√					*	722.03
Harvinder Singh (2013)	*	√	√					122.96	
S. Mohana Priya (2013)	*	*	*					187.90	250.07
Rakesh kr. (2014)	*	*	*					107.61	149.31
Kunal Kishore (2014)	*	*	*					527.80	*
Ali Naser (2014)	*	*	√					153.43	61.47
Sunny Nandwani (2015)	√	√	√					100.95	*
Slesha (2015)	*	*	√					54.91	1.37
Reena Panwar (2015)	√	*	√					111.37	51.94
Sarika Vasantrao Bodke (2016)	√	*	√					*	10.43
Mamta Khanchi (2016)	√	*	√					101.33	*
Simar Preet Singh (2016)	√	*	√					63.22	*
Er. Pooja (2016)	*	*	*					62.56	13.28
Pradeep Singh (2016)	√	*	*					530.77	*
Saurabh Shukla (2016)	√	*	√					165.36	*
This Paper	√	√	√	√	√	√	√	41.55	10.85

After the cloud analyst successfully performs various performance tests, we get the total response time for various load balancing algorithms, as shown in the table above. The analysis results for the total response time and load are based on various algorithms in a cloud-based environment.

Table 3
Comparison of Overall Response Time and Processing Time Summary (Part-I)

Algorithm	**Round Robin**			**ESCE**			**Throttled**		
Response Time	Avg (ms)	Min (ms)	Max (ms)	Avg (ms)	Min (ms)	Max (ms)	Avg (ms)	Min (ms)	Max (ms)
Over all RT	1306.92	16.04	6358.23	128.89	11.55	5427.28	204.05	16.08	6358.06
Processing Time	1056.47	0.00	39667.38	572.87	0.00	33842.12	1040.85	0.01	39664.90

Table 3
Comparison of Overall Response Time and Processing Time Summary (Part-II)

Algorithm	**HTV**			**Dynamic Priority Based**			**Composite**		
Response Time	Avg (ms)	Min (ms)	Max (ms)	Avg (ms)	Min (ms)	Max (ms)	Avg (ms)	Min (ms)	Max (ms)
Over all RT	255.06	20.10	7947.57	293.32	23.12	9139.70	216.80	17.09	6755.43
Processing Time	1301.07	0.01	49581.13	1496.23	0.02	57018.29	1105.91	0.01	42143.96

Table 3
Comparison of Overall Response Time and Processing Time Summary (Part-III)

Algorithm	**Global Framework**		
Response Time	Avg (ms)	Min (ms)	Max (ms)
Over all RT	191.30	15.08	5960.68
Processing Time	975.80	0.01	37185.84

7. CONCLUSION

This study deals with various load balancing algorithms. The existing algorithms are static, dynamic, composite, and prioritized. The ultimate purpose of those algorithms is to reduce the response time and maximize the resource utilization. The results of the previous algorithms are limited to give improved result. Still there is plenty of space to improve the results to extract best service from cloud service providers. This study also shows the comparative results of the existing load balancing algorithms based on the parameters such as makes pan and resource utilization. This is a conceptual piece of work and some more work is to be done leading to development of new algorithm, its execution and implementation. Amendments would be made as per the ongoing changes in the technology of cloud computing and the problems that arise accordingly.

REFERENCES

1. Bhargava, N., Bhargava, R., Mathuria, M. and Daima, R. (2013). Performance Analysis of Cloud Computing for Distributed Client. *International Journal of Computer Science and Mobile Computing, IJCSMC*, 2(6), 97-104.
2. Bodake, S.V. (2016). Cloud Computing Load Balancing Algorithms Comparison Based Survey. *International Journal of Advanced Engineering, Management and Science*, 2(5), 282-287.

3. Chandra, H., Semwal, P. and Chopra, S. (2016). Load balancing in cloud computing using a novel minimum make span algorithm. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 5(4), 1160-1164.
4. Chronopoulos, A.T. *Game Theory Based Load Balanced Job Allocation*, San Antonio. Online available at http://graal.ens-lyon.fr/~lmarchal/aussois/slides/chronopoulos.pdf
5. Garg, D. and Saxena, U. (2016). Dynamic Queue Based Enhanced HTV Dynamic Load Balancing Algorithm in Cloud Computing. *International Journal of Innovative Research in Science, Engineering and Technology*, 5(1), 141-153.
6. Ghumman, N.S. and Sachdeva, R. (2016). An Efficient Approach for Load Balancing in Cloud Computing using Composite Techniques. *International Journal of Research in Engineering and Applied Sciences*, 6(2), 145-149.
7. Ghumman, N.S. and Sachdeva, R. (2016). An efficient approach for load balancing in cloud computing using composite techniques. *International Journal of Research in Engineering and Applied Sciences*, 6(2), 145-149.
8. Gouda, K.C., Radhika, T.V. and Akshatha, M. (2013). Priority based resource allocation model for cloud computing. *International Journal of Science, Engineering and Technology Research (IJSETR)*, 2(1), 215-219.
9. Iranpour, E. and Sharifian, S. (2018). A distributed load balancing and admission control algorithm based on Fuzzy type-2 and Game theory for large-scale SaaS cloud architectures. *Future Generation Comp. Syst.*, 86, 81-98.
10. Issawi, S.F., Al-Halees, A. and Radi, M. (2015). An Efficient Adaptive Load Balancing Algorithm for Cloud Computing Under Bursty Workloads. *Engineering, Technology & Applied Science Research*, 5(3), 795-800.
11. James, J. and Verma, B. (2012). Efficient VM load balancing algorithm for a cloud computing environment. *International Journal on Computer Science and Engineering*, 4(9), 1658-1663.
12. Kapgate, D. (2014). Efficient service broker algorithm for data center selection in cloud computing. *International Journal of Computer Science and Mobile Computing*, 3(1), 355-365.
13. Khanchi, M., and Tyagi, S. (2016). An Efficient Algorithm for Load Balancing in Cloud Computing. *International Journal of Engineering Sciences & Research Technology*, 5(6), 468-475.
14. Liang, P.H. and Yang, J.M. (2015). Evaluation of Two-Level Global Load Balancing Framework in Cloud Environment. *International Journal of Computer Science & Information Technology*, 7(2), 1-11.
15. Limbani, D. and Oza, B. (2012). A proposed service broker strategy in cloud analyst for cost-effective data center selection. *International Journal of Engineering Research and Applications, India*, 2(1), 793-797.
16. Mangla, E.R. and Singh, E.H. (2015). Recovery and user priority based load balancing in cloud computing. *International Journal of Engineering and Science and Research*, 4(2), 407-412.
17. Mishra, R.K. and Bhukya, S.N. (2014). Service broker algorithm for cloud-analyst. *International Journal of Computer Science and Information Technologies*, 5(3), 3957-3962.

18. Moharana, S.S., Ramesh, R.D. and Powar, D. (2013). Analysis of load balancers in cloud computing. *International Journal of Computer Science and Engineering (IJCSE)*, 2(2), 101-108.
19. Moradi, M., Dezfuli, M.A. and Safavi, M.H. (2010). A new time optimizing probabilistic load balancing algorithm in grid computing. In *2010 IEEE 2nd International Conference on Computer Engineering and Technology*, 1, 232-237.
20. Nandhalakshmi, B.N. and Mahalingam (2015). Efficient Load Balancing in Cloud Computing Using Weighted Throttled Algorithm. *International Journal of Innovative Research in Computer and Communication Engineering*, 3(6), 5409-5415.
21. Nayak, S. and Patel, P. (2015). Analytical Study for Throttled and propsed Throttled algorithm for load Balancing in Cloud Computing using Cloud Analyst. *International Journal of Science Technology & Engineering*, 1(12), 90-100.
22. Panwar, R. and Mallick, B. (2015). A comparative study of load balancing algorithms in cloud computing. *International Journal of Computer Applications*, 117(24), 33-37.
23. Panwar, R. and Mallick, B. (2015). Load balancing in cloud computing using dynamic load management algorithm. In *2015 IEEE International Conference on Green Computing and Internet of Things (ICGCIoT)*, 773-778.
24. Pooja, Er. and Thapar, E.V. (2016). An Enhanced Virtual Machine Load Balancing Algorithm for Cloud Environment. *International Journal of Emerging Research in Management &Technology*, 5(5), 166-170.
25. Priya, S.M. and Subramani, B. (2013). A New Approach for Load Balancing in Cloud Computing. *International Journal of Engineering and Computer Science*, 2(5), 1636-1640.
26. Sharma, M., Sharma, P. and Sharma, S. (2012). Efficient Load Balancing Algorithm in VM Cloud Environment. *International Journal of Computer Science*, 3(1), 439-441.
27. Singh, H. and Gangwar, R.C. (2014). Efficient Approach for Load Balancing in Virtual Cloud Computing Environment. *Compusoft (An International Journal of Advanced Computer Technology)*, 3(10), 1204-1210.
28. Soni, S. and Yadav, P. (2016). A Load Balancing Approach to Minimize the Resource Wastage in Cloud Computing. *International Advanced Research Journal in Science*, Engineering and Technology, 3(3), 94-98.
29. Suryadevi, G., Vijayakumar, D., SabariMuthuKumar, R. and Srinivasagan, D.K. (2014). An Efficient Distributed Dynamic Load Balancing Algorithm for Private Cloud Environment. *International Journal of Innovative Research in Science,* Engineering and Technology, 3(3), 2181-2186.
30. Velte, A.T., Velte, T.J., Elsenpeter, R.C. and Elsenpeter, R.C. (2010). *Cloud Computing: A Practical Approach* (p. 44). New York: McGraw-Hill.
31. Zalavadiya, K., & Vaghela, D. (2016). Honey bee behavior load balancing of tasks in cloud computing. *International Journal of Computer Applications*, 139(1), 16-19.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 113-120

SOME NEW MEMBERS OF THE T-X FAMILY OF DISTRIBUTIONS

Farrukh Jamal[1] and **Arslan Nasir**[2]
The Islamia University of Bahawalpur, Bahawalpur, Pakistan.
Email: [1]drfarrukh1982@gmail.com
[2]arslannasir147@gmail.com

ABSTRACT

Alzaatreh et al., 2013 introduced the transformed-transformer (T-X) for developing the new family of probability of distributions. In this paper we define some new members of T-X family which also be viewed as new families of probability distributions. We check the existence of the proposed distributions through their graphs of probability distribution functions (pdf) and cumulative density functions (cdf).

1. INTRODUCTION

The recent literature has suggested several ways of extending well-known distributions. A common way consists in defining new classes of univariate continuous distributions by introducing additional shapes parameter(s) to a baseline distribution. The role of this additional parameter(s) is useful in exploring tail properties and also for improving the goodness-of-fit of the generator family. The well-known families are: the beta-G (Eugene et al., 2002; Jones, 2004), Kumaraswamy-G (Kw-G; Cordeiro and de Castro, 2011), McDonald-G (Mc-G; Alexander et al., 2012), gamma-X (Alzaatreh et al., 2014), gamma-G (type 1) (Zografos and Balakrishanan, 2009), gamma-G (type 2) (Ristic and Balakrishanan, 2012), gamma-G (type 3) (Torabi and Montazari, 2012), log-gamma-G (Amini et al., 2012), logistic-G (Torabi and Montazari, 2014), transformed-transformer (T-X; Alzaatreh et al., 2013), Exponentiated T-X (Alzaghal et al., 2013), Logistic-X Family (Tahir et al., 2016) and New Weibull-X Family (Zubair et al. 2018).

Let $v(t)$ be the probability density function (pdf) of a random variable, say $\boldsymbol{T}$, where $\boldsymbol{T} \in [m, n]$ for $-\infty \leq m < n < \infty$ and let $W[F(x;\xi)]$ be a function of cumulative distribution function (cdf) of a random variable, say $\boldsymbol{X}$, depending on the vector parameter ξ satisfying the conditions given below:

$$W[F(x;\xi)] \in [m,n]$$

1. $W[F(x;\xi)]$ is differentiable and monotonically increasing,
2. $W[F(x;\xi)] \to m$ as $x \to -\infty$ and $W[F(x;\xi)] \to n$ as $x \to \infty$.

Alzaatreh et al. (2013), defined the cdf of the T-X family of distributions by

$$G(x;\xi) = \int_m^{W[F(x;\xi)]} v(t)dt \quad \boldsymbol{x} \in \boldsymbol{R}, \tag{1}$$

where $W[F(x;\xi)]$ satisfies the conditions stated above. The pdf corresponding to (1) is given by $g(x;\xi) = \left\{\frac{\partial}{\partial x} W[F(x;\xi)]\right\} v\{W[F(x;\xi)]\}$ $\quad \boldsymbol{x} \in \boldsymbol{R}$.

Using the T-X idea, several new classes of distributions have been introduced in the literature. Table 1 provides some W[F(x; ξ)] functions for some members of the T-X family.

Table 1
Different $\boldsymbol{W[F(x;\xi)]}$ Functions for Special Members of the T-X Family

S#	$\boldsymbol{W[F(x;\xi)]}$	Range of X	Members of T-X Family
1	$F(x;\xi)$	[0,1]	Beta-G (Eugene et al., 2002), Mc-G (Alexander et al., 2012)
2	$-\log[1-F(x;\xi)]$	$[0,\infty]$	Gamma-G Type-1 (Zografos and Balakrishnan, 2009)
3	$\frac{F(x;\xi)}{1-F(x;\xi)}$	$[0,\infty]$	Gamma-G Type-3 (Torabi and Montazeri, 2012)
4	$-\log[1-F^{\alpha}(x;\xi)]$	$[0,\infty]$	Exponentiated T-X (Alzaghal et al., 2013)
5	$\log\left[\frac{F(x;\xi)^{\alpha}}{1-F(x;\xi)^{\alpha}}\right]$	$[-\infty,\infty]$	A new method for generating families of continuous distributions (Alzaatreh et al., 2013)
6	$\log\left[\frac{F(x;\xi)}{1-F(x;\xi)}\right]$	$[-\infty,\infty]$	Logistic-G (Torabi and Montazeri, 2014)
7	$\log[-\log[1-F^{\alpha}(x;\xi)]]$	$[-\infty,\infty]$	The Logistic-X Family (Tahir et al., 2016)
8	$\frac{F(x;\xi)^{\alpha}}{1-F(x;\xi)^{\alpha}}$	0 to ∞	The generalized odd log-logistic family (Cordeiro et al., 2017)
9	$\frac{-\log[1-F(x;\xi)]}{1-F(x;\xi)}$	$[0,\infty]$	New Weibull-X Family (Zubair et al., 2018)
10	$\frac{-\log[1-\mathrm{F}^{\alpha}(\mathrm{X};\varepsilon)]}{1-\mathrm{F}^{\alpha}(\mathrm{X};\varepsilon)}$	$[0,\infty]$	Jamal Weibull-X Family **(Proposed)**
11	$\frac{\log[-\log\{1-F^{\alpha}(X;\varepsilon)\}]}{1-F^{\alpha}(X;\varepsilon)}$	$[-\infty,\infty]$	Nasir Logistic-X Family **(Proposed)**

Table 2
Different $\boldsymbol{W[1-F(x;\xi)]}$ Functions for Members of the Survival Family

S#	$\boldsymbol{W[1-F(x;\xi)]}$	Range of X	Members of Survival Family
1	$-\log[F(x;\xi)]$	$[0,\infty]$	Gamma-G Type-2 (Ristić and Balakrishnan, 2012)
2	$\frac{-\log[\mathrm{F}(\mathrm{X};\varepsilon)]}{\mathrm{F}(\mathrm{X};\varepsilon)}$	[0 to ∞]	Nasir Weibull-Generalized Family **(Proposed)**

From Table 2, we can see that they are not belong to the T-X family. However, one can say that it is derived to the T-X family. It is characterized by the cdf $G(x;\,\xi) = 1 - \int_m^{W[1-F(x,\xi)]} v(t)dt$ or the survival function (sf) as, $S(x;\,\xi) = \int_m^{W[1-F(x,\xi)]} v(t)dt$) with the same $W(u)$, i.e. $W(u)$ is increasing, $W(u) \to m$ when $u \to 0$ and $W(u) \to n$ when $u \to 1$.

2. SOME GENERALIZED FAMILIES BASED ON NEW MEMBERS OF THE T-X FAMILY

2.1 Jamal Weibull-X (JW-X) Family

If X is a Weibull random variable with parameter $\boldsymbol{\alpha}, \boldsymbol{\beta} > 0$ then from Table 1, S.No. 10 and form Eq. 1, we define the cdf and pdf of the JW-X family by

$$\boldsymbol{G(x;\alpha,\beta,\xi) = \int_0^{\frac{-\log(1-F^{a}(x;\xi))}{1-F^{a}(x;\xi)}} \alpha\beta\, e^{-\alpha t^{\beta}}\, dt = 1 - \exp\left[-\alpha\left\{\frac{-\log(1-F^{a}(x;\xi))}{1-F^{a}(x;\xi)}\right\}^{\beta}\right]}, \quad (2)$$

and

$$g(x;\alpha,\beta,\xi) = \alpha\beta a\,\frac{f(x;\xi)F^{a-1}(x;\xi)\{\log[1-F^{a}(x;\xi)]\}^{\beta}}{[F^{a}(x;\xi)-1]^{\beta+1}}\exp\left[-\alpha\left\{\frac{-\log(1-F^{a}(x;\xi))}{1-F^{a}(x;\xi)}\right\}^{\beta}\right]. \quad (3)$$

2.2 Nasir Logistic-X (NL-X) Family.

If X is a Logistic random variable with parameter $\boldsymbol{\lambda} > 0$ then from Table 1, S.No 11 and form Eq. 1, we define the cdf and pdf of the NL-X family by

$$\begin{aligned} G(x;\lambda,\xi) &= \int_{-\infty}^{\frac{\log[-\log[1-F^{a}(x;\xi)]]}{1-F^{a}(x;\xi)}} \lambda\, e^{-\lambda\, t}\left\{1+e^{-\lambda\, t}\right\}^{-2} dt \\ &= \left[1+\exp\left[\frac{-\lambda\,\log[-\log[1-F^{a}(x;\xi)]]}{1-F^{a}(x;\xi)}\right]\right]^{-1} \end{aligned} \quad (4)$$

and

$$g(x;\lambda,\xi) = \frac{\lambda\alpha f(x;\xi)F^{\alpha-1}(x;\xi)}{[1-F^{\alpha}(x;\xi)]^{2}}\left[\frac{1}{\log[1-F^{\alpha}(x;\xi)]} + \log[-\log[1-F^{\alpha}(x;\xi)]]\right] \frac{\exp\left[\frac{-\lambda\,\log[-\log[1-F^{\alpha}(x;\xi)]]}{1-F^{\alpha}(x;\xi)}\right]}{\left[1+\exp\left[\frac{-\lambda\,\log[-\log[1-F^{\alpha}(x;\xi)]]}{1-F^{\alpha}(x;\xi)}\right]\right]^{2}}. \quad (5)$$

2.3 Jamal Logistic-X (JL-X) Family

If X is a Logistic random variable with parameter $\boldsymbol{\lambda} > 0$ then from Table 1, S.No 5 and form Eq. 1, we define the cdf and pdf of the JL-X family by

$$\begin{aligned} G(x;\lambda,\xi) &= \int_{-\infty}^{\log\left[\frac{F(x;\xi)^{a}}{1-F(x;\xi)^{a}}\right]} \lambda\, e^{-\lambda\, t}\left\{1+e^{-\lambda\, t}\right\}^{-2} dt \\ &= \left\{1+\exp\left[-\lambda\,\log\left(\frac{F(x;\xi)^{a}}{1-F(x;\xi)^{a}}\right)\right]\right\}^{-1}, \end{aligned} \quad (6)$$

and

$$(x;\lambda,\xi) = \frac{a\, f(x)}{F(x)[1-a(x)]}\exp\left[-\lambda\,\log\left(\frac{F(x;\xi)^{a}}{1-F(x;\xi)^{a}}\right)\right]\left\{1+\exp\left[-\lambda\,\log\left(\frac{F(x;\xi)^{a}}{1-F(x;\xi)^{a}}\right)\right]\right\}^{-2}. \quad (7)$$

2.4 New Family of Distributions with Survival Function

2.4.1 Nasir Weibull-Generalized (NW-G) Family

If X is a Weibull random variable with parameters $\boldsymbol{\alpha}, \boldsymbol{\beta} > 0$ then from Table 2, S.No. 2, we define the sf and pdf of the NW-G family by

$$S(x;\alpha,\beta,\xi) = \int_0^{\frac{-\log F(x;\xi)}{F(x;\xi)}} \alpha\beta\, e^{-\alpha t^{\beta}} dt = 1 - \exp\left[-\alpha\left\{\frac{-\log F(x;\xi)}{F(x;\xi)}\right\}^{\beta}\right], \tag{8}$$

and

$$g(x;\alpha,\beta,\xi) = \alpha\beta\,\frac{f(x;\xi)\{-\log F(x;\xi)\}^{\beta-1}}{F(x;\xi)^{\beta+1}} \exp\left[-\alpha\left\{\frac{-\log F(x;\xi)}{F(x;\xi)}\right\}^{\beta}\right][1-\log F(x;\xi)]. \tag{9}$$

2.5 Some Special Models

2.5.1 Jamal Weibull-Lomax (JW-Lx) Distribution

Let **X** follows Lomax distribution with pdf and cdf respectively, $f(x;\theta,\sigma) = \frac{\sigma}{\theta}\left(1+\frac{x}{\theta}\right)^{-\sigma-1}$ and $F(x;\theta,\sigma) = 1-\left(1+\frac{x}{\theta}\right)^{-\sigma}$,where $\theta \geq 0$ and $\sigma \geq 0,$ then from Eq 2 and 3 the cdf and pdf of JW-Lx distribution as follow

$$G(x;\theta,\sigma,\alpha,\beta,a) = 1 - \exp\left[-\alpha\left\{\frac{-\log\left[1-\left\{1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right\}^{a}\right]}{1-\left\{1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right\}^{a}}\right\}^{\beta}\right],$$

and

$$g(x;\theta,\sigma,\alpha,\beta,a)\,\frac{\alpha\beta\sigma}{\theta}\frac{\left\{1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right\}^{a-1}}{\left(1+\frac{x}{\theta}\right)^{\sigma+1}}\left\{\frac{-\log\left[1-\left\{1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right\}^{a}\right]}{1-\left\{1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right\}^{a}}\right\}^{\beta-1}$$
$$\times\frac{\left[1-\log\left[1-\left\{1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right\}^{a}\right]\right]}{\left[1-\left\{1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right\}^{a}\right]^{2}}\times\exp\left[-\alpha\left\{\frac{-\log\left[1-\left\{1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right\}^{a}\right]}{1-\left\{1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right\}^{a}}\right\}^{\beta}\right].$$

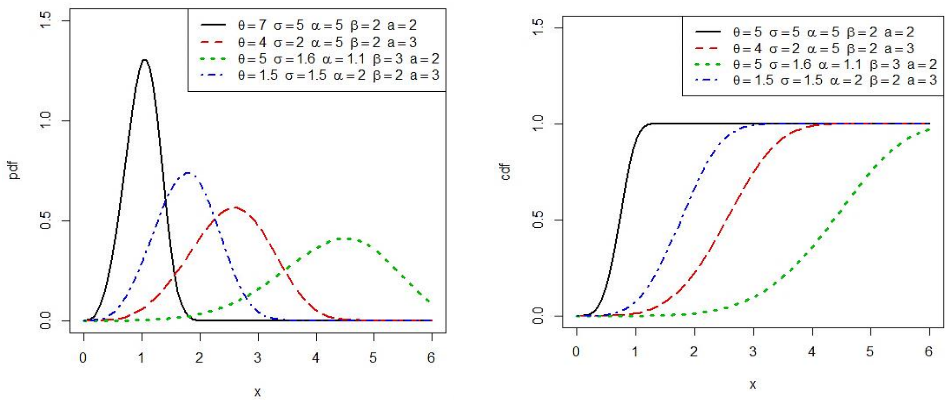


Figure 1: Plots of pdf and cdf for JW-Lx Distribution for Different Values of the Parameters

2.5.2 Nasir Logistic-Lomax (NL-Lx) Distribution

Let **X** follows Lomax distribution with pdf and cdf respectively, $f(x;\theta,\sigma)=\frac{\sigma}{\theta}\left(1+\frac{x}{\theta}\right)^{-\sigma-1}$ and $F(x;\theta,\sigma)=1-\left(1+\frac{x}{\theta}\right)^{-\sigma}$, where $\theta \geq 0$ and $\sigma \geq 0$, then from Eq. 4 and 5 the cdf and pdf of NL-Lx distribution given as

$$G(x;\theta,\sigma,a,\lambda)=\left[1+\exp\left[\frac{-\lambda\,\log\left[-\log\left[1-\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{a}\right]\right]}{1-\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{a}}\right]\right]^{-1}$$

and

$$g(x;\theta,\sigma,a,\lambda)=\frac{\lambda a\frac{\sigma}{\theta}\left(1+\frac{x}{\theta}\right)^{-\sigma-1}\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{a-1}}{\left[1-\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{a}\right]^{2}}$$
$$\left[\frac{1}{\log\left[1-\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{a}\right]}+\log\left[-\log\left[1-\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{a}\right]\right]\right]$$
$$\times\frac{\exp\left[\frac{-\lambda\,\log\left[-\log\left[1-\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{a}\right]\right]}{1-\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{a}}\right]}{\left[1+\exp\left[\frac{-\lambda\,\log\left[-\log\left[1-\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{a}\right]\right]}{1-\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{a}}\right]\right]^{2}}.$$

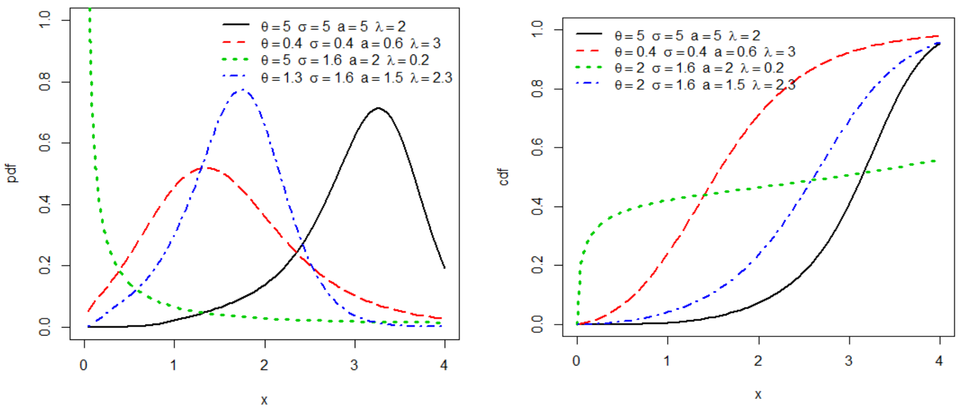


Figure 2: Plots of pdf and cdf for NL-Lx Distribution for Different Values of the Parameters

2.5.3 Jamal Logistic-Lomax (JL-Lx) Distribution

Let **X** follows Lomax distribution with pdf and cdf respectively, $f(x;\theta,\sigma) = \frac{\sigma}{\theta}\left(1+\frac{x}{\theta}\right)^{-\sigma-1}$ and $F(x;\theta,\sigma) = 1-\left(1+\frac{x}{\theta}\right)^{-\sigma}$, where $\theta \geq 0$ and $\sigma \geq 0,$ then from Eq. 6 and 7 the cdf and pdf of JL-Lx distribution given as

$$G(x;\theta,\sigma,a,\lambda) = \left\{1+\exp\left[-\lambda\ \log\left(\frac{\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{a}}{1-\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{a}}\right)\right]\right\}^{-1},$$

and

$$g(x;\theta,\sigma,a,\lambda) = \frac{a\frac{\sigma}{\theta}\left(1+\frac{x}{\theta}\right)^{-\sigma-1}}{\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]\left[1-\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{a}\right]}$$

$$\exp\left[-\lambda\ \log\left(\frac{\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{a}}{1-\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{a}}\right)\right]\left\{1+\exp\left[-\lambda\ \log\left(\frac{\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{a}}{1-\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{a}}\right)\right]\right\}^{-2}.$$

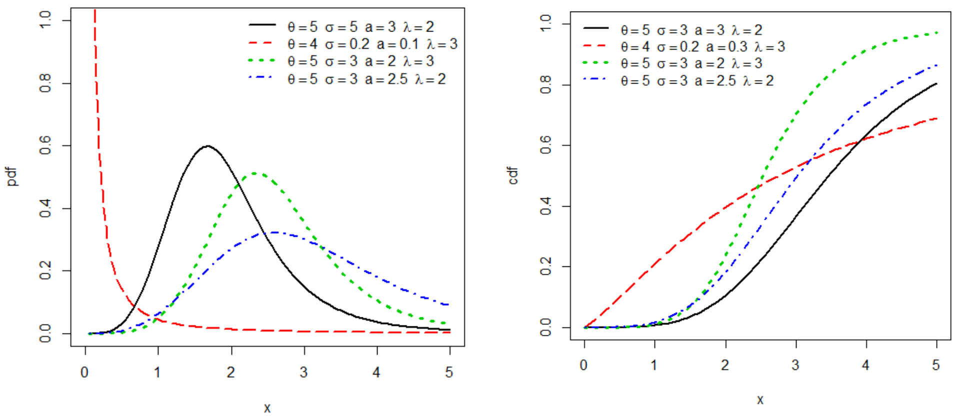


Figure 3: Plots of pdf and cdf for JL-Lx Distribution for Different Values of the Parameters

2.5.4 Nasir Weibull-Lomax (NW-Lx) Distribution

Let **X** follows Lomax distribution with pdf and cdf respectively, $f(x;\theta,\sigma) = \frac{\sigma}{\theta}\left(1+\frac{x}{\theta}\right)^{-\sigma-1}$ and $F(x;\theta,\sigma) = 1-\left(1+\frac{x}{\theta}\right)^{-\sigma}$, where $\theta \geq 0$ *and* $\sigma \geq 0$, then from Eq. 8 and 9 the cdf and pdf of NW-Lx distribution given as

$$S(x;\theta,\sigma,\alpha,\beta) = 1-\exp\left[-\alpha\left\{\frac{-\log\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]}{1-\left(1+\frac{x}{\theta}\right)^{-\sigma}}\right\}^{\beta}\right],$$

and

$$g(x;\theta,\sigma,\alpha,\beta) = \alpha\beta\frac{\frac{\sigma}{\theta}\left(1+\frac{x}{\theta}\right)^{-\sigma-1}\left\{-\log\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]\right\}^{\beta-1}}{\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]^{\beta+!}}$$

$$\exp\left[-\alpha\left\{\frac{-\log\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]}{\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]}\right\}^{\beta}\right]\left[1-\log\left[1-\left(1+\frac{x}{\theta}\right)^{-\sigma}\right]\right].$$

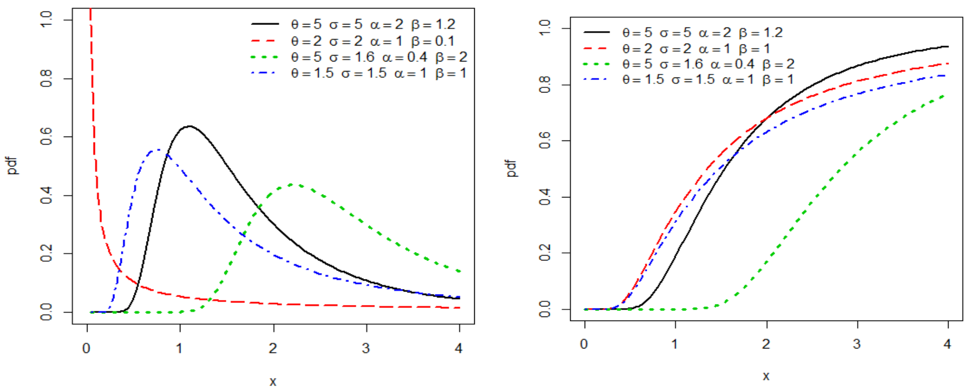


Figure 4: Plots of PDF and cdf for NW-Lx Distribution for Different Values of the Parameters.

REFERENCES

1. Ahmad, Z., Elgarhy, M. and. Hamedani, G.G. (2018). A new Weibull-X family of distributions: properties, characterizations and applications. *Journal of Statistical Distributions and Applications*, 5, 1-18.
2. Alexander, C., Cordeiro, G.M., Ortega, E.M.M. and Sarabia, J.M. (2012). Generalized beta-generated distributions. *Computational Statistics and Data Analysis*, 56, 1880-1897.
3. Alzaatreh, A., Famoye, F. and Lee, C. (2013). A new method for generating families of continuous distributions. *Metron*, 71, 63-79.
4. Alzaatreh, A., Famoye, F. and Lee, C. (2014). The gamma-normal distribution: Properties and applications. *Computational Statistics and Data Analysis*, 69, 67-80.
5. Alzaghal, A., Lee, C. and Famoye, F. (2013). Exponentiated T-X family of distributions with some applications. *International Journal of Probability and Statistics*, 2, 31-49.
6. Cordeiro, G.M., Alizadeh, M., Ozel, G., Hosseini, B., Ortega, E.M.M. and Altun, E. (2017). The generalized odd log-logistic family of distributions: properties, regression models and applications. *Journal of Statistical Computation and Simulation*, 87(5), 908-932.
7. Cordeiro, G.M., de Castro, M. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, 81, 883-893.
8. Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics–Theory and Methods*, 31, 497-512.
9. Jones, M.C. (2004). Families of distributions arising from the distributions of order statistics. *Test*, 13, 1-43.
10. Ristic, M.M. and Balakrishnan, N. (2012). The gamma-exponentiated exponential distribution. *Journal of Statistical Computation and Simulation*, 82, 1191-1206.
11. Ristić, M.M. and Balakrishnan, N. (2012). The gamma-exponentiated exponential distribution. *Journal of Statistical Computation and Simulation*, 82(8), 1191-1206.
12. Tahir, M.H., Cordeiro, G.M., Alzaatreh, A., Mansoor, M. and Zubair, M. (2016). The Logistic-X family of distributions and its applications. *Communications in Statistics-Theory and Methods*, 45(24), 7326-7349.
13. Torabi, H. and Montazari, N.H. (2012). The gamma-uniform distribution and its application. *Kybernetika*, 48, 16-30.
14. Torabi, H. and Montazari, N.H. (2014). The logistic-uniform distribution and its application. *Communications in Statistics–Simulation and Computation*, 43, 2551-2569.
15. Zografos, K. and Balakrishnan, N. (2009). On families of beta- and generalized gamma-generated distributions and associated inference. *Statistical Methodology*, 6, 344-362.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 121-125

SURVEY OF PAKISTAN HEALTH SECTOR

Imran Anwar Ujan[1], **Arifa Bhutto**[1], **Imdad Ali Ismaili**[1]
and **Asadullah Shah**[2]

[1] Institute of Information & Communication Technology,
University of Sindh, Jamshoro, Pakistan

[2] Kulliyyah of Information & Communication Technology,
International Islamic University, Malaysia

ABSTRACT

The health sector is important for human development for healthy life. In Pakistan the basic healthcare facility and living healthy life is right of every people.

In this research the researchers are study the Pakistan economic survey 2016-17 in healthcare sector. Using the data of above survey to study the "Prime Minister's National Health Program", Polio Eradication Initiatives, Healthcare Facilities, and Healthcare Expenditure, Vertical Programmes and other elements will be discussed.

INTRODUCTION

Public welfare & development is an elementary right inherent in each personal as stipulated in the "constitution" of the "Islamic Republic of Pakistan" and "United Nations" charter. worldwide level and inside nation states, there exists a debate for welfare or development must be the yardstick of indications of economic development. As such, mentioned sub-sectors have always received the first priority in the various models of economic development. In "Pakistan", the consecutive governments, including the compulsory government undertake poverty improvement as part of different programs.

The health sector is delegated to the provinces of Pakistan under the 18th Constitutional Amendment and provinces. It has since created health policies. At the provinces, "KPK" recently enacted the "Medical Teaching Institutions", Injured "Persons Medical Aid Act", and "Child Nutrition Act". The governments of "Sindh", "Punjab" created the "Health Commissions" and created a health plan outline that delineates the programs that need the highest priority. The Federal program also implemented the "National Maternal and Child Health Program", "Malaria", "Tuberculosis', and "HIV/AIDS Control Program". Above programs have helped improve the overall health of the people. The coverage of immunization was also increased. Polio cases decreased. Deaths caused by Tuberculosis also decreased, whereas child and maternal mortality rates show an improvement from 2007.

"Prime Minister's National Health Program"

To achieve a worldwide health treatment, the "National Health Program" was inaugurated through the leadership of "Prime Minister" by the "Ministry of National

Health Services", "Regulations and Coordination (MoNHSRC)". Aims of program is to give assistance to the millions of poor citizens and give access to quality of healthcare.

"Polio Eradication Initiatives"

Incidence of Polio has been eradicating in other parts of the world. Improved and routine immunization helped eradicate this disease in Pakistan. According to the World Health Organization (WHO), there are only two polio cases reported for the present year until April 25, 2017 as compared to the nine cases reported on the previous year. Country is, indeed, close to eradicating Polio. This improvement indicates that the immunization program is achieving its desired objectives. Under the "Global Polio Eradication Initiatives (GPEIK)", the federal governments, "World Health Organization (WHO)", "United Nations Children's Fund (UNICEF)", and "US Centres for Disease Control (CDC)" were in cooperation to help eradicate polio around a globe. During the World Economic Forum in Davos, philanthropist, Bill Gates expressed commendation on the impressive success of polio eradication in Pakistan. In the presence of the Prime Minister of Pakistan, Bill Gates also mentioned the that year 2017 is the year of Polio eradication in Pakistan.

Table 11.1: Provinces Wise Polio Cases

Province	2010	2011	2012	2013	2014	2015	2016	2017
Punjab	7	9	2	7	5	2	0	1
Sindh	27	33	4	10	30	12	8	0
KPK	24	23	27	11	68	17	8	0
FATA	74	59	20	65	179	16	2	0
Balochistan	12	73	4	0	25	7	2	0
Gilgit-baltistan	0	1	1	0	0	0	0	1
Azad jammu & kashmir	0	0	0	0	0	0	0	0
TOTAL	**144**	**198**	**58**	**93**	**306**	**54**	**20**	**2**

Source: End Polio Pakistan (www.endpolio.com.pk)

"Health Facilities"

Public health events have consistently improved the number of physical organization and workforce in Pakistan. The doctors, dentists, hospital nurses and "LHVs" have also increased, including the availability of doctor, dentist, nurse in every hospital bed vis-a-vis the population. At present, population ratio of Doctors is at 1: 997; dentist is at 1 : 10658; and hospital bed is at 1 : 1584. The national health infrastructure makes up 1201 hospitals, 5518 "BHUs", 683 "RHCs", 5802 "Dispensaries", 731 "Maternity and Child Health Centers", and 347 "TB centers". The total beds available in the mentioned health services is estimated at 123394.

Table 11.2: Healthcare Facilities						(Numbers)
Health Manpower	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17
Registered Doctors	152368	160880	167759	175223	184711	195896
Registered Dentists	11649	12692	13716	15106	16652	18333
Registered Nurses	77683	82119	86183	90276	94766	99228
Population per Doctor	1162	1123	1099	1073	1038	997
Population per Dentist	15203	14238	13441	12447	11513	10658
Population per Bed	1647	1616	1557	1591	1604	1584
Source: Pakistan Bureau of Statistics						

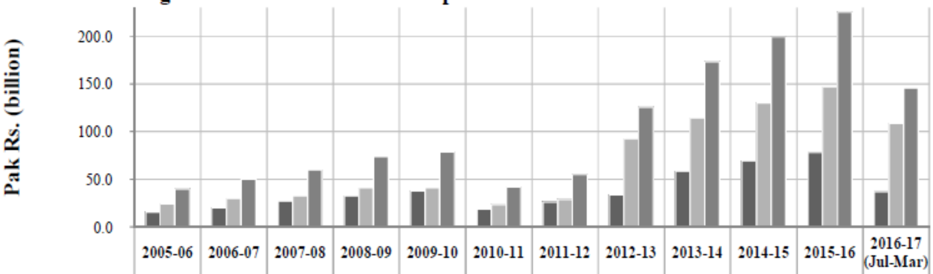
"Health Expenditure"

Health spending is steadily rising at a slow pace. The GDP of 0.5 to 0.8 percent spent on health for previous ten years. However, this percentage is below the "WHO" benchmark of having minimum 6 percent GDPO in order to deliver basic and lifesaving facilities to the people. In the fiscal year 2015 - 2016, the total expenditure rose to 13 percent over the fiscal year 2014 - 2015. Present fiscal year (July - March) 2016 - 2017, the expenditure of 145.97 billion, which shows an increase of 9 percent as compared to the previous year. Conferring to new report of "World Bank", "Pakistan's" per capital health expenditure is US $36.2, which is lesser than "WHO's "low income countries' benchmark of US $86.

Table 11.3: Health & Nutrition Expenditures (2000-01 to 2016-17)					(Rs. billion)
Fiscal Years	Public Sector Expenditure (Federal and Provincial)			Percentage Change	Health Expenditure as % of GDP
	Total Health Expenditures	Development Expenditure	Current Expenditure		
2000-01	24.28	5.94	18.34	9.98	0.58
2001-02	25.41	6.69	18.72	4.63	0.57
2002-03	28.81	6.61	22.21	13.42	0.59
2003-04	32.81	8.50	24.31	13.85	0.58
2004-05	38.00	11.00	27.00	15.84	0.58
2005-06	40.00	16.00	24.00	5.26	0.49
2006-07	50.00	20.00	30.00	25.00	0.54
2007-08	59.90	27.23	32.67	19.80	0.56
2008-09	73.80	32.70	41.10	23.21	0.56
2009-10	78.86	37.86	41.00	6.86	0.53
2010-11	42.09	18.71	23.38	-46.63	0.23

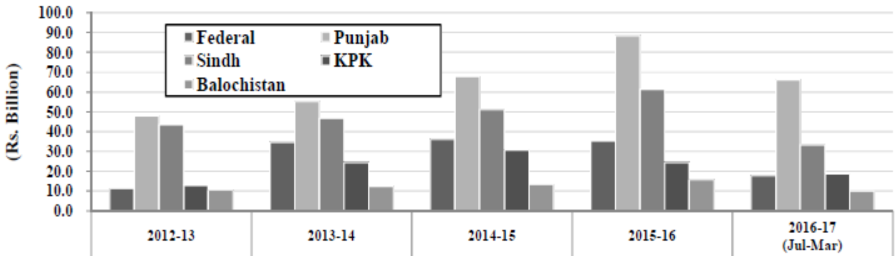
Table 11.3: Health & Nutrition Expenditures (2000-01 to 2016-17)					(Rs. billion)
Fiscal Years	Public Sector Expenditure (Federal and Provincial)			Percentage Change	Health Expenditure as % of GDP
	Total Health Expenditures	Development Expenditure	Current Expenditure		
2011-12	55.12	26.25	28.87	30.96	0.27
2012-13	125.96	33.47	92.49	128.51	0.56
2013-14	**173.42**	**58.74**	**114.68**	**37.68**	**0.69**
2014-15	**199.32**	**69.13**	**130.19**	**14.94**	**0.72**
2015-16	**225.33**	**78.07**	**147.26**	**13.05**	**0.76**
July-March					
2015-16	133.93	39.94	94.00	-	-
2016-17	145.97	37.47	108.50	8.99	0.46
Source: Finance Division (PF Wing)					

Fig-11.1: Total Public Sector Expenditure on Health



	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17 (Jul-Mar)
Development	16.0	20.0	27.2	32.7	37.9	18.7	26.3	33.5	58.7	69.1	78.1	37.5
Current	24.0	30.0	32.7	41.1	41.0	23.4	28.9	92.5	114.7	130.2	147.3	108.5
Total	40.0	50.0	59.9	73.8	78.9	42.1	55.1	126.0	173.4	199.3	225.3	146.0

Fig-11.2: Share in Total Public Sector Health Expenditure



	2012-13	2013-14	2014-15	2015-16	2016-17 (Jul-Mar)
Federal	11.4	34.7	36.2	35.3	17.8
Punjab	47.8	55.1	67.7	88.5	66.2
Sindh	43.5	46.6	51.2	61.3	33.4
KPK	12.8	24.5	30.8	24.4	18.7
Balochistan	10.4	12.5	13.5	15.8	10.0

Source: Finance Division (PF Wing)

Various programs and projects are given funds by means of the PSDP and provincial and governments. This includes the “Civil Registration and vital statistics”, “Family Planning and Primary Health Care (FP&PHC)”, “Expanded Program on Immunization”, “Malaria Control Program”, “Tuberculosis (TB) Control Program”, “HIV/AIDS Control Program”, “Maternal Neonatal & Child Health (MNCH) Program”, “Prime Minister’s Program for Prevention and Control of Hepatitis, and Cancer Treatment Program”.

CONCLUSION

High population growth characterizes the health profile of Pakistan. Despite the given significant investments in the health sector ended the year, the progress degree is still lower than furthermost countries. This accounts for the domination of malnutrition and poor dietary practices of the country. “SDGs” serve as a national commitment to help alleviate health outcomes. The “Federal and provincial governments” are in cooperation to help bring developments by means of implementing and monitoring the health sector strategic framework in order to attain the health-related “SDGs”.

DATA SOURCE

Pakistan Economic Survey 2016-17, Economic Adviser’s Wing Economic Adviser’s Wing, Finance Division, Government of Pakistan.

REFERENCES

1. Government of Pakistan (2017). *Pakistan Economic Survey 2016-17*, Economic Adviser’s Wing, Finance Division, Government of Pakistan, Islamabad.
2. Mashhadi, S.F., Hamid, S., Roshan, R. and Fawad, A. (2016). Healthcare in Pakistan–a systems perspective. *Pakistan Armed Forces Medical Journal*, 66(1), 136-42.
3. Mahdavian, M., Nazarian, H., Mahdavian, M. and Wattanapongsakorn, N. (2014). An investigation of the success of hospital information systems implementation: A case study. In *2014 IEEE International Computer Science and Engineering Conference (ICSEC)*, 329-333.
4. Rabbani, F., Jafri, S.M.W., Abbas, F., Shah, M., Azam, S.I., Shaikh, B.T., Brommels, M. and Tomson, G. (2010). Designing a balanced scorecard for a tertiary care hospital in Pakistan: a modified Delphi group exercise. *The International Journal of Health Planning and Management*, 25(1), 74-90.
5. Malik, M.A. and Khan, H.R. (2009). Understanding the implementation of an electronic hospital information system in a developing country: a case study from Pakistan. In *Proceedings of the Third Australasian Workshop on Health Informatics and Knowledge Management*, Volume 97 (pp. 31-36). Australian Computer Society, Inc.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 127-146

IDENTIFYING THE PROBLEM SOLVING STRATEGIES OF MATHEMATICS STUDENTS OF SECONDARY LEVEL

Atiqa Yasmeen
University of Education, Lahore, Pakistan
Email: atiqay@gmail.com

ABSTRACT

This study sought to "identify the problem solving strategies of mathematics students of secondary level". It is a descriptive research. Survey method was used for this study. The secondary school students studying mathematics constituted the population of this study. All the secondary school science students of Lahore were considered as the population for the study. The sample comprised of 250 secondary school science students. The researchers selected the sample of 130 girls and 120 boys by using convenient sampling technique. Statistical Package for Social Sciences (SPSS), version 15 was used for entering the data. Means scores and standard Deviation were found and t-test was used to analyze data. Applying t-tests the results shows that all the secondary students whether they belongs to the public or a private schools having any gender does not perform well in strategy "logical reasoning" and they performed very good in strategies "Guess and check" and "Solve a simpler problem. The suggestion was that such type of knowledge should be provided by the teachers to the pupils which is helpful to them to get awareness about problem solving strategies.

INTRODUCTION

Kantowski (1980, p. 195) stated that "a problem is a situation for which the individual who faces it has no procedure that will guarantee a solution". That person's relevant knowledge must be put together in a new way to solve the problem.

According to Mayer and Wittrock (2006, p. 287).problem solving is "cognitive processing directed at achieving a goal when no solution method is obvious to the problem solver"

There are some skills and strategies of problem solving. With the help of those strategies one can solve daily life problems. Our main concern is with the problem solving strategies. Some of those strategies are as following:

- Guess and check is a problem-solving strategy that students can use to solve mathematical problems by guessing the answer and then checking that the guess fits the conditions of the problem.
- Look for a pattern means some problems can be solved by recognizing a pattern. Making a table can help the students recognize the pattern.
- Work backward is the strategy of undoing key elements in the problem in order to find a solution.
- Make an organized list is a problem-solving strategy which enables students to organize the data and visually consider their options when answering a problem. "In the attempt to produce an organized list, students will encounter frequent and repeated patterns." (Muckerheide et al., 1999).

- Eliminate possibilities is a strategy where students use a process of elimination until they find the correct answer. This is a problem-solving strategy that can be used in basic math problems or to help solve logic problems
- Act it out help the students visualize what the problem is and what is involved. They go through the actions of what the problem says and at the end come to a conclusion.
- Use logical reasoning is a problem-solving strategy that involves the use of Venn diagrams or charts to help students use logic to solve a problem.
- Solve a simpler problem is useful when solving a complex problem because it allows you to reduce large numbers to small numbers, or reducing the number of items given in a problem.
- Draw a diagram is a strategy that incorporates the use of drawing pictures to represent the problem. This strategy is frequently used to help solve some problems.
- Reasonable answer refers to the results of a calculation or problem-solving operation reflecting what is reasonable within the context of the given factors or values.
  (Florida Department of Education, Division of Public Schools and Community Education, Bureau of Exceptional Education and Student Services 2010)

OBJECTIVES

This study was designed with following objectives in view:

- To identify the problem solving strategies of science students at secondary level.
- To find the effects of demographics variables (school type, grade, gender) on science students problem solving strategies.

NULL HYPOTHESIS

H_1: There is no significant mean difference between public and private school students' "Guess and check" problem solving strategy at secondary level.

H_2: There is no significant mean difference between public and private school students' "look for a pattern" problem solving strategy at secondary level.

H_3: There is no significant mean difference between public and private school students' "work backward" problem solving strategy at secondary level.

H_4: There is no significant mean difference between public and private school students' "Make an organized list" problem solving strategy at secondary level.

H_5: There is no significant mean difference between public and private school students' "Eliminate possibilities" problem solving strategy at secondary level.

H_6: There is no significant mean difference between public and private school students' "act it out" problem solving strategy at secondary level.

H_7: There is no significance mean difference between public and private school students' "use logical reasoning" problem solving strategy at secondary level.

H_8: There is no significant mean difference between public and private school students' "Solve a simpler problem" problem solving strategy at secondary level.

H_9: There is no significant mean difference between public and private school students' "Draw a diagram" problem solving strategy at secondary level.

H_{10}: There is no significant mean difference between public and private school students' "Reasonable answer" problem solving strategy at secondary level.

H_{11}: There is no significant mean difference between 9^{th} and 10^{th} class students in "guess and check" problem solving strategy at secondary level.

H_{12}: There is no significant mean difference between 9^{th} and 10^{th} class students in "look for a pattern" problem solving strategy at secondary level.

H_{13}: There is no significant mean difference between 9^{th} and 10^{th} class students in "work backward" problem solving strategy at secondary level.

H_{14}: There is no significant mean difference between 9^{th} and 10^{th} class students in "make an organized list" problem solving strategy at secondary level.

H_{15}: There is no significant mean difference between 9^{th} and 10^{th} class students in "eliminate possibilities" problem solving strategy at secondary level.

H_{16}: There is no significant mean difference between 9^{th} and 10^{th} class students in "act it out" problem solving strategy at secondary level.

H_{17}: There is no significant mean difference between 9^{th} and 10^{th} class students in "use logical reasoning" at secondary level.

H_{18}: There is no significant mean difference between 9^{th} and 10^{th} class students in "solve a simpler problem" in problem solving strategy at secondary level.

H_{19}: There is no significant mean difference between 9^{th} and 10^{th} class students in "draw a diagram" problem solving strategy at secondary level.

H_{20}: There is no significant difference between 9^{th} and 10^{th} class students in "reasonable answers" problem solving strategy at secondary level.

H_{21}: There is no significant mean difference between male and female students in "guess and check" problem solving strategy at secondary level.

H_{22}: There is no significant difference between male and female students in "look for a pattern" problem solving strategy at secondary level.

H_{23}: There is no significant mean difference between male and female students in "work backward" problem solving strategy at secondary level.

H_{24}: There is no significant mean difference between male and female students in "make an organized list" problem solving strategy at secondary level.

H_{25}: There is no significant mean difference between male and female in "eliminate possibilities" problem solving strategy at secondary level.

H_{26}: There is no significant mean difference between male and female class students in "act it out" problem solving strategy at secondary level

H_{27}: There is no significant difference between male and female student in use "logical reasoning" problem solving strategy at secondary level.

H_{28}: There is no significant mean difference between male and female students in "solve a simpler problem" in problem solving strategy at secondary level

H_{29}: There is no significant difference between male and female students in "draw a diagram" problem solving strategy at secondary level.

H_{30}: There is no significant difference between male and female students in "reasonable answers" problem solving strategy at secondary level.

RESEARCH METHODOLOGY

Research Design: It is a descriptive research. Survey method was used for this study. In survey research, the researcher selects a sample of respondents from a population and administers a standardized questionnaire or a test to them.

Population: All the secondary school science students of Lahore were considered as the population for the present study.

Sampling: The sample comprised of 250 secondary school science students. The researchers selected the sample of 130 girls and 120 boys by using convenient sampling technique.

Instrument of the Study: For this study, the researchers developed a test to identify the mathematics student's problem solving strategies as a tool of study. It was consisted on 10 strategies and 30 items. Each strategy having 3 questions. The detail of the strategies and its items are given in the following table.

ANALYSIS

This chapter outlines the analysis and interpretation of data. The data collected for the research was analyzed and represented in the form of tables which were followed by their interpretations. The objectives of the study are to "identify the problem solving strategies of science students at secondary level".

After the collection of test, data were entered on SPSS then researchers find out frequencies and percentages of each statement.

After applying all the tests the researchers formulated tables. Then researcher's interpreted the data.

Data Analysis

Table 1: Analysis and Interpretation 4.1

S#	Questions	Correct answers		Incorrect answers	
		f	%	f	%
1	What value of x is a solution of x+14.9=30.2?	210	84	40	16
2	Asia has an area of 44579000 square kilometer write 44579000 in scientific notation?	227	90.8	23	9.2
3	Every morning Morgan swims 0.5 km. how many meters does she swim?	143	57.2	107	42.8

Above table shows that

- 84% of students found correct value of x while 16% students found incorrect value of question 1.
- 90.8% students found correct scientific notation while 9.2% students found incorrect scientific notation of question 2.
- 57.2% students found correct numbers of meters while 42.8% students found incorrect numbers of meters in question 3.

Table 2: Analysis and Interpretation 4.2

S#	Questions	Correct answers		Incorrect answers	
		f	**%**	**f**	**%**
4	What is equivalent to $5^{4?}$	225	90	25	10
5	How many millimeters are in 13 centimeters?	138	55.2	112	44.8
6	Use the formula A=*bh* to find the height of a parallelogram with a base of 34 millimeters and an area of 612 square kilometers?	129	51.6	121	48.4

Above table shows that

- 90% students originate correct equivalent value of 54 while 10% students found incorrect value of question 4.
- 55.2% students found correct amount of millimeters in centimeters while 44.8% students found incorrect amount of millimeters of question 5.
- 51.6% students originate the correct height of parallelogram while 48.4% students found incorrect height of parallelogram of question 6.

Table 3: Analysis and Interpretation 4.3

S#	Questions	Correct answers		Incorrect answers	
		f	**%**	**f**	**%**
7	A taxi driver charges $2.00 plus $0.80 for each mile traveled. Which expression could be used to find the cost of a taxi ride if “m” miles are traveled?	60	24	190	76
8	Kim’s time for 5k race was four minutes less than Tanya’s time. If Tanya’s time is t, which expression represents the Kim’s time?	103	41.2	147	58.8
9	The temperature at 6:00 A.M. was -5ºF. What was the temperature at 8:00 A.M. if it had risen 7 degrees?	121	48.4	129	51.6

Above table shows that

- 24% students identify correct cost of a taxi ride while 76% students found incorrect cost of question 7.
- 41.2% students identify correct time of Kim’s race while 58.8% students found incorrect time of question 8.
- 48.4% students identify correct value of temperature while 51.6% students found incorrect value of temperature of question 9.

Table 4: Analysis and Interpretation 4.4

S#	Questions	Correct answers		Incorrect answers	
		f	%	f	%
10	Find the ratio that represents 20%?	73	29.2	177	70.8
11	Find the LCM of 27 and 30?	138	55.2	112	44.8
12	Find the number that is not rational.	127	50.8	123	49.2

Above table shows that

- 29.2% students explore the correct ratio while 70.8 students found incorrect ratio of question 10.
- 55.2% students explore correct LCM while 44.8% students found incorrect LCM of question 11.
- 50.8% students explore the number that is not rational while 49.2% students found incorrect number.

Table 5: Analysis and Interpretation 4.5

S#	Questions	Correct answers		Incorrect answers	
		f	%	f	%
13	An elephant in a zoo eats 58 cabbages in a week. About how many cabbages does an elephant eat in one year?	161	64.4	89	35.6
14	What value of x makes 2/3x=-9 a true sentence?	74	29.6	176	70.4
15	What is the reciprocal of 2. 1/5?	84	33.6	166	66.4

Above table shows that

- 64.4% students construct the correct amount of cabbages while 35.6% students found incorrect value of cabbages in question 13.
- 29.6% students construct the correct value of “x” while 70.4% students found incorrect value in question 14.
- 33.6% students construct the correct reciprocal of 2. 1/5 while 66.4% students found incorrect answers of question 15.

Table 6: Analysis and Interpretation 4.6

S#	Questions	Correct answers		Incorrect answers	
		f	%	f	%
16	What is the theoretical probability of choosing a vowel from the word MATHEMATICS?	144	57.6	106	42.4
17	Which of the following is the greatest common factor of 84 and 49?	189	75.6	61	24.4
18	The ornithologist said there is a 0.6754 chance of the bird returning next year to its nest in a tree in a Glowinski family’s backyard. What is the value of the 5 in this number?	107	42.8	143	57.2

Above table shows that

- 57.6% of the students found correct probability while 42.4% of the students found incorrect probability of question 16.
- 75.6% of students found correct common factor while 24.4% of the students found incorrect common factor of question 17.
- 42.8% of students found correct value of 5 while 57.2% of students found incorrect value of 5 of question 18.

Table 7: Analysis and Interpretation 4.7

S#	Questions	Correct answers		Incorrect answers	
		f	%	f	%
19	Which of the following flowchart symbols is not a polygon?	107	42.8	143	57.2
20	Which of the following regular shapes cannot be used by itself to make a tessellation?	63	25.2	187	74.8
21	Δ ABC with vertices A (-2, -5), B (4, 1), and C (3, -2) is reflected over the y-axis. Find the coordinates of the new figure.	95	38	155	62

Above table shows that

- 42.8% of the students establish correct symbol of polygon while 57.2% of the students found incorrect symbol of polygon of question 19.
- 25.2% of the students establish correct regular shape to make a tessellation while 74.8% of the students found incorrect regular shape to make a tessellation of question 20.
- 38% of the students establish correct coordinates while 62 % of the students found incorrect coordinates of question 21.

Table 8: Analysis and Interpretation 4.8

S#	Questions	Correct answers		Incorrect answers	
		f	%	f	%
22	Find 1/7× 2/9	161	64.4	89	35.6
23	Which of the square root of 441?	205	82	45	18
24	Which of the reasonable estimate for the square root of 66?	192	76.8	58	23.2

Above table shows that

- 64.4% of the students found correct value while 35.6% of the student's incorrect value of question 22.
- 82% of the students found correct square root while 18% of the students found incorrect square root of question 24.
- 76.8% of the students found correct reasonable estimate while 23.2% of the students found incorrect reasonable estimate of question 24.

Table 9: Analysis and Interpretation 4.9

S#	Questions	Correct answers		Incorrect answers	
		f	%	f	%
25	Choose the best estimate for 100.5%	121	48.4	129	51.6
26	A football player has made 80% of the field goals, he has attempted in his career. If he attempts 5 field goals in a game, how many would he be expected to make?	132	52.8	118	47.2
27	Which of the following is the value of p in p/7=21/49?	168	67.2	82	32.8

Above table shows that

- 48.4% of the students originate correct estimate while 51.6 % of the students found incorrect estimate of question 25.
- 52.8% of the students originate correct answer while 47.2% of the students found incorrect answer of question 26.
- 67.2% of the students originate correct value while 32.8 % of the students found incorrect value of question 27.

Table 10: Analysis and Interpretation 4.10

S#	Questions	Correct answers		Incorrect answers	
		f	%	f	%
28	23 is what percent of 64? Round to the nearest tenth	129	51.6	121	48.4
29	A cookie company received 1,600 e-mails in one week. Of those e-mails, 12.5% were people requesting catalogues. How many people requested catalogues that week?	129	51.6	121	48.4
30	Alan wants to buy a new snowboard that has a regular price of $169. This week the snowboard is on sale at 35% discount. What is the sale price of the snowboard?	128	51.2	122	48.8

Above table shows that

- 51.6% of the students construct correct percent while 48.4 % of the students found incorrect percent of question 28.
- 51.6% of the students construct correct catalogues while 48.4% of the students found incorrect catalogues of question 29.
- 51.2% of the students construct correct sale price while 48.8% of the students found incorrect catalogues of question 30.

Descriptive Statistics

Table 11: Descriptive Statistics

S#	Strategies	Mean	Standard Deviation
1	Guess and check	2.3200	.79759
2	Look for a pattern	1.9680	.84963
3	Work backward	1.1360	.84401
4	Make an organized list	1.3520	.89851
5	Eliminate possibilities	1.2760	.87343
6	Act it out	1.7600	.95606
7	Logical reasoning	1.0600	.89195
8	Solve a simpler problem	2.2320	.93255
9	Draw a diagram	1.6840	.88276
10	Reasonable answers	1.5440	1.04135

Above table shows that

- The strategy “logical reasoning” has the smallest mean from all the strategies.
- “Guess and check” and “Solve a simpler problem” have the greatest mean from all the strategies.

Hypothesis

Hypothesis details

Hypothesis 1

There is no significant mean difference between public and private school students’ “Guess and check” problem solving strategy at secondary level.

Table 12: Hypothesis 1 Table

Variables	N	Mean	df	t	Sig
Public	144	2.3611	248	.950	.343
Private	106	2.2642			

Table shows that t-value (.95) is not significant at $P \leq 0.05$ level of significance. So, our null hypothesis that” There is no significant mean difference between public and private school students’ “Guess and check” problem solving strategy at secondary level” is accepted.

Hypothesis 2

There is no significant mean difference between public and private school students’ “look for a pattern” problem solving strategy at secondary level.

Table 13: Hypothesis 2 Table

Variables	N	Mean	df	t	Sig
Public	144	1.8472	248	-2.651	.009
Private	106	2.1321			

This table shows that that t-value (-2.651) is significant at P≤0.05 level of significance. So, our null hypothesis that" There is no significant mean difference between public and private school students' "look for a pattern" problem solving strategy at secondary level" is rejected.

So, it is concluded that there is significant mean difference between public and private school students' "look for a pattern" problem solving strategy at secondary level.

Hypothesis 3

There is no significant mean difference between public and private school students' "work backward" problem solving strategy at secondary level.

Table 14: Hypothesis 3 Table

Variables	N	Mean	df	t	Sig
Public	144	1.1111	248	-.543	.588
Private	106	1.1698			

Above table shows that that t-value (-.543) is significant at P≤0.05 level of significance. So, our null hypothesis that" There is no significant mean difference between public and private school students' "work backward" problem solving strategy at secondary level" is rejected.

So, it is concluded that there is significant mean difference between public and private school students' "work backward" problem solving strategy at secondary level.

Hypothesis 4

There is no significant mean difference between public and private school students' "Make an organized list" problem solving strategy at secondary level.

Table 15: Hypothesis 4 Table

Variables	N	Mean	df	t	Sig
Public	144	1.3264	248	-.525	.6
Private	106	1.3868			

This table shows that that t-value (-.525) is significant at P≤0.05 level of significance. So, our null hypothesis that" There is no significant mean difference between public and private school students' "make an organized list" problem solving strategy at secondary level" is rejected.

So, it is concluded that there is significant mean difference between public and private school students' "make an organized list" problem solving strategy at secondary level.

Hypothesis 5

There is no significant mean difference between public and private school students' "Eliminate possibilities" problem solving strategy at secondary level.

Table 16: Hypothesis 5 Table

Variables	N	Mean	df	t	Sig
Public	144	1.3125	248	.769	.442
Private	106	1.2264			

This shows that t-value (.769) is not significant at P≤0.05 level of significance. So, our null hypothesis that" There is no significant mean difference between public and private school students' "eliminate possibilities" problem solving strategy at secondary level" is accepted.

Hypothesis 6

There is no significant mean difference between public and private school students' "act it out" problem solving strategy at secondary level.

Table 17: Hypothesis 6 Table

Variables	N	Mean	df	t	Sig
Public	144	1.9167	248	3.071	.002
Private	106	1.5472			

Above table shows that t-value (3.071) is not significant at P≤0.05 level of significance. So, our null hypothesis that" There is no significant mean difference between public and private school students' "act it out" problem solving strategy at secondary level" is accepted.

Hypothesis 7

There is no significant mean difference between public and private school students' "Use logical reasoning" problem solving strategy at secondary level.

Table 18: Hypothesis 7 Table

Variables	N	Mean	df	t	Sig
Public	144	1.0278	248	-.665	.507
Private	106	1.1038			

This table shows that that t-value (-.665) is significant at P≤0.05 level of significance. So, our null hypothesis that" There is no significant mean difference between public and private school students' "use logical reasoning" problem solving strategy at secondary level" is rejected.

So, it is concluded that there is significant mean difference between public and private school students' "use logical reasoning" problem solving strategy at secondary level.

Hypothesis 8

There is no significant mean difference between public and private school students' "Solve a simpler problem" problem solving strategy at secondary level.

Table 19: Hypothesis 8 Table

Variables	N	Mean	df	t	Sig
Public	144	2.1944	248	-.741	.459
Private	106	2.2830			

Above table shows that that t-value (-.741) is significant at P≤0.05 level of significance. So, our null hypothesis that" There is no significant mean difference between public and private school students' "Solve a simpler problem" problem solving strategy at secondary level" is rejected.

So, it is concluded that there is significant mean difference between public and private school students' "Solve a simpler problem" problem solving strategy at secondary level.

Hypothesis 9

There is no significant mean difference between public and private school students' "Draw a diagram" problem solving strategy at secondary level.

Table 20: Hypothesis 9 Table

Variables	N	Mean	df	t	Sig
Public	144	1.7361	248	1.088	.278
Private	106	1.6132			

Above table shows that t-value (1.7361) is not significant at P≤0.05 level of significance. So, our null hypothesis that" There is no significant mean difference between public and private school students' "draw a diagram" problem solving strategy at secondary level" is accepted.

Hypothesis 10

There is no significant mean difference between public and private school students' "Reasonable answer" problem solving strategy at secondary level.

Table 21: Hypothesis 10 Table

Variables	N	Mean	df	t	Sig
Public	144	1.6944	248	2.696	.008
Private	106	1.3396			

This shows that t-value (2.696) is not significant at P≤0.05 level of significance. So, our null hypothesis that" There is no significant mean difference between public and private school students' "reasonable answer" problem solving strategy at secondary level" is accepted.

Hypothesis 11

There is no significant mean difference between 9th and 10th class students in "guess and check" problem solving strategy at secondary level.

Table 22: Hypothesis 11 Table

Variables	N	Mean	df	t	Sig
9th class	154	2.2792	248	-1.024	.307
10th class	96	2.3854			

Above table shows that t value (-1.024) is not significant at p ≤0.05 level of significance so our null hypothesis that there is no significant mean difference between 9th and 10th students in guess and check problem solving strategy at secondary level is rejected.

So, it is concluded that there is significant mean difference between 9th and 10th students in guess and check problem solving strategy at secondary level.

Hypothesis 12

There is no significant mean difference between 9th and 10th class students in "look for a pattern" problem solving strategy at secondary level.

Table 23: Hypothesis 12 Table

Variables	N	Mean	df	t	Sig
9th class	154	1.9156	248	-1.237	.217
10th class	96	2.0521			

This table shows that t value (-1.237) is not significant at $p \leq 0.05$ level of significance so our null hypothesis that there is no significant mean difference between 9th and 10th class students in look for a pattern problem solving strategy at secondary level is rejected.

So, it is concluded that there is significant mean difference between 9th and 10th class students in look for a pattern problem solving strategy at secondary level.

Hypothesis 13

There is no significant mean difference between 9th and 10th class students in "work backward" problem solving strategy at secondary level.

Table 24: Hypothesis 13 Table

Variables	N	Mean	df	t	Sig
9th class	154	1.0390	248	-2.323	0.21
10th class	96	1.2917			

Above table shows that t value (-2.323) is not significant at $p \leq 0.05$ level of significance so our null hypothesis that there is no significant mean difference between 9th and 10th class students in work backward problem solving strategy at secondary level is rejected.

So, it is concluded that there is significant mean difference between 9th and 10th class students in work backward problem solving strategy at secondary level.

Hypothesis 14

There is no significant mean difference between 9th and 10th class students in "make an organized list" problem solving strategy at secondary level.

Table 25: Hypothesis 14 Table

Variables	N	Mean	df	t	Sig
9th class	154	1.2403	248	-2.517	0.12
10th class	96	1.5313			

This shows that t value (-2.517) Is not significant at $p \leq 0.05$ level of significance so our null hypothesis that there is no significant mean difference between 9th and 10th class students in make an organized list problem solving strategy at secondary level is rejected.

So, it is concluded that there is significant mean difference between 9th and 10th class students in organized list problem solving strategy at secondary level.

Hypothesis 15

There is no significant mean difference between 9th and 10th class students in "eliminate possibilities" problem solving strategy at secondary level.

Table 26: Hypothesis 15 Table

Variables	N	Mean	df	t	Sig
9th class	154	1.2403	248	-8.19	.414
10th class	96	1.3333			

This table shows that t value (-8.19) is not significant at $p \leq 0.05$ level of significance so our null hypothesis that there is no significant mean difference between 9th and 10th class students in eliminate possibilities problem solving strategy at secondary level is rejected.

So, it is concluded that there is significant mean difference between 9th and 10th class students in eliminate possibilities problem solving strategy at secondary level.

Hypothesis 16

There is no significant mean difference between 9th and 10th class students in "act it out" problem solving strategy at secondary level.

Table 27: Hypothesis 16 Table

Variables	N	Mean	df	t	Sig
9th class	154	1.8377	248	1.632	.104
10th class	96	1.6354			

This indicates that value (1.632) is not significant at $p \leq 0.05$ level of significance so our null hypothesis that there is no significant mean difference between 9th and 10th class students in act it out problem solving strategy at secondary level is accepted.

So, it is concluded that there is significant mean difference between 9th and 10th class students in act it out problem solving strategy at secondary level.

Hypothesis 17

There is no significant mean difference between 9th and 10th class students in "use logical reasoning" at secondary level.

Table 28: Hypothesis 17 Table

Variables	N	Mean	df	t	Sig
9th class	154	1.0779	248	.402	.688
10th class	96	1.0313			

This shows that t value (.402) is not significant at $p \leq 0.05$ level of significance so our null hypothesis that there is no significant mean difference between 9th and 10th class students in use of logical reasoning at secondary level is accepted.

So, it is concluded that there is no significant difference between 9th and 10th class students in use logical reasoning at secondary level.

Hypothesis 18

There is no significant mean difference between 9th and 10th class students in “solve a simpler problem” in problem solving strategy at secondary level.

Table 29: Hypothesis 18 Table

Variables	N	Mean	df	t	Sig
9th class	154	2.1688	248	-1.359	.175
10th class	96	2.3333			

This table shows that t value (-1.359) is not significant at $p \leq 0.05$ level of significance so our null hypothesis that there is no significant mean difference between 9th and 10th class students in solve a simpler problem in problem solving strategy at secondary level is rejected.

So, it is concluded that there is significant mean difference between 9th and 10th class students in solve a simpler problem in problem solving strategy at secondary level.

Hypothesis 19

There is no significant mean difference between 9th and 10th class students in “draw a diagram” problem solving strategy at secondary level.

Table 30: Hypothesis 19 Table

Variables	N	Mean	df	t	Sig
9th class	154	1.7078	248	.539	.590
10th class	96	1.6458			

Above table shows that t value (.539) is not significant at $p \leq 0.05$ level of significance so our null hypothesis that there is no significant mean difference between 9th and 10th class students in draw a diagram problem solving strategy at secondary level is accepted.

So, it is concluded that there is no significant mean difference between 9th and 10th class students in draw a diagram problem solving strategy at secondary level.

Hypothesis 20

There is no significant difference between 9th and 10th class students in “reasonable answers” problem solving strategy at secondary level.

Table 31: Hypothesis 20 Table

Variables	N	Mean	df	t	Sig
9th class	154	1.6623	248	2.295	.023
10th class	96	1.3542			

Above table shows that t value (2.295) is not significant at $p \leq 0.05$ level of significance so our null hypothesis that there is no significant difference between 9th and 10th class students in reasonable answers problem solving strategy at secondary level is accepted.

So, it is concluded that there is no significant difference between 9th and 10th class students in reasonable answers problem solving strategy at secondary level.

Hypothesis 21

There is no significant mean difference between male and female students in “guess and check” problem solving strategy at secondary level.

Table 32: Hypothesis 21 Table

Variables	N	Mean	df	t	Sig
Male	130	2.2538	248	-1.367	.173
Female	120	2.3917			

Above table shows that t value (-1.367) is not significant at $p \leq 0.05$ level of significance so our null hypothesis that There is no significant mean difference between male and female students in guess and check problem solving strategy at secondary level is rejected.

So, it is concluded that there is significant mean difference between male and female students in guess and check problem solving strategy at secondary level.

Hypothesis 22

There is no significant difference between male and female students in “look for a pattern” problem solving strategy at secondary level.

Table 33: Hypothesis 22 Table

Variables	N	Mean	df	t	Sig
Male	130	2.0000	248	.619	.536
Female	120	1.9333			

Above table shows that t value (.619) is not significant at $p \leq 0.05$ level of significance so our null hypothesis that there is no significant difference between male and female students in “look for a pattern” problem solving strategy at secondary level is accepted.

So, it is concluded that there is no significant difference between male and female students in look for a pattern problem solving strategy at secondary level.

Hypothesis 23

There is no significant mean difference between male and female students in “work backward” problem solving strategy at secondary level.

Table 34: Hypothesis 23 Table

Variables	N	Mean	df	t	Sig
Male	130	1.0846	248	-1.002	.317
Female	120	1.1917			

This shows that t value (-1.002) is not significant at $p \leq 0.05$ level of significance so our null hypothesis that there is no significant mean difference between male and female students in work backward problem solving strategy at secondary level is rejected.

So, it is concluded that there is significant mean difference between male and female students in work backward problem solving strategy at secondary level.

Hypothesis 24

There is no significant mean difference between male and female students in "make an organized list" problem solving strategy at secondary level.

Table 35: Hypothesis 24 Table

Variables	N	Mean	df	t	Sig
Male	130	1.3923	248	.738	.461
Female	120	1.3083			

Above table shows that t value (.738) is not significant at $p \leq 0.05$ level of significance so our null hypothesis that there is no significant mean difference between male and female students in make an organized list problem solving strategy at secondary level is accepted.

So, it is concluded that there is no significant mean difference between male and female students in organized list problem solving strategy at secondary level.

Hypothesis 25

There is no significant mean difference between male and female in "eliminate possibilities" problem solving strategy at secondary level.

Table 36: Hypothesis 25 Table

Variables	N	Mean	df	t	Sig
Male	130	1.1846	248	-1.729	.085
Female	120	1.3750			

Above table shows that t value (-1.729) is not significant at $p \leq 0.05$ level of significance so our null hypothesis that there is no significant mean difference between male and female in eliminate possibilities problem solving strategy at secondary level is rejected.

So, it is concluded that there is significant mean difference between male and female in organized list problem solving strategy at secondary level.

Hypothesis 26

There is no significant mean difference between male and female class students in "act it out" problem solving strategy at secondary level.

Table 37: Hypothesis 26 Table

Variables	N	Mean	df	t	Sig
Male	130	1.5154	248	-4.360	.000
Female	96	2.0250			

This table shows that t value (-4.360) is not significant at $p \leq 0.05$ level of significance so our null hypothesis that There is no significant mean difference between male and female class students in act it out problem solving strategy at secondary level is rejected.

So, it is concluded that there is significant mean difference between male and female class students in act it out problem solving strategy at secondary level.

Hypothesis 27

There is no significant difference between male and female student in use "logical reasoning" problem solving strategy at secondary level.

Table 38: Hypothesis 27 Table

Variables	N	Mean	df	t	Sig
Male	130	1.0538	248	-.113	.910
Female	120	1.0667			

This table shows that t value (-.113) is not significant at $p \le 0.05$ level of significance so our null hypothesis that there is no significant difference between male and female student in use logical reasoning problem solving strategy at secondary level.

So, it is concluded that there is significant difference between male and female student in use logical reasoning problem solving strategy at secondary level.

Hypothesis 28

There is no significant mean difference between male and female students in "solve a simpler problem" in problem solving strategy at secondary level

Table 39: Hypothesis 28 Table

Variables	N	Mean	df	t	Sig
Male	130	2.1923	248	-.700	.485
Female	120	2.2750			

This shows that t value (-.700) is not significant at $p \le 0.05$ level of significance so our null hypothesis that there is no significant mean difference between male and female students in solve a simpler problem solving strategy at secondary level.

So, it is concluded that there is significant mean difference between male and female students in solve a simpler problem solving strategy at secondary level.

Hypothesis 29

There is no significant difference between male and female students in "draw a diagram" problem solving strategy at secondary level.

Table 40: Hypothesis 29 Table

Variables	N	Mean	df	t	Sig
Male	130	1.5692	248	-2.155	0.32
Female	120	1.8083			

This shows that t value (-2.155) is not significant at $p \le 0.05$ at level of significance so our null hypothesis that there is no significant difference between male and female students in draw a diagram problem solving strategy at secondary level is rejected.

So, it is concluded that there is significant difference between male and female students in draw a diagram problem solving strategy at secondary level.

Hypothesis 30

There is no significant difference between male and female students in "reasonable answers" problem solving strategy at secondary level.

Table 41: Hypothesis 30 Table

Variables	N	Mean	df	t	Sig
Male	130	1.2385	248	-5.062	.000
Female	120	1.8750			

Above table shows that t value (-5.062) is not significant at $p \leq 0.05$ level of significance so our null hypothesis that There is no significant difference between male and female students in reasonable answers problem solving strategy at secondary level is rejected.

So, it is concluded that there is significant difference between male and female students in reasonable answers problem solving strategy at secondary level.

CONCLUSION

If we talk about the conclusion of the study briefly then the strategy "logical reasoning" has the smallest mean, and "Guess and check" and "Solve a simpler problem" have the greatest mean from all the strategies. It means that all the secondary students whether they belongs to the public or a private schools having any gender does not perform well in strategy "logical reasoning" and they performed very good in strategies "Guess and check" and "Solve a simpler problem.

And now we will discuss separately about all the variables and all the strategies. Results about school type shows that there is no significant mean difference in responses between public and private school students in "guess and check", "eliminate possibilities", "act it out", "draw a diagram" and "reasonable answer" strategies. While private school students perform proficient than public school students in remaining strategies.

If we talk w.r.t grade then there is no significant mean difference in responses between 9th and 10th class students in "reasonable answers", "draw a diagram", "use logical reasoning" and "act it out" strategies. While 10th class students perform competent than 9th class students in remaining all the strategies. Conclusions drawn w.r.t gender shows that there is no significant mean difference in answers between male and female students in "look for a pattern" and "make an organized list" strategies. While in all the other strategies female students shows the effective answers.

REFERENCES

1. Klingler, K.L. (2012). *Mathematic Strategies for Teaching Problem Solving: The Influence of Teaching Mathematical Problem Solving Strategies on Students' Attitudes in Middle School*. Doctoral dissertation, University of Central Florida Orlando, Florida.
2. Laterell, C.M. (2013). What is problem solving ability. *LATM Journal*, 1(1).

3. Mayer, R. and Wittrock, M. (2009). *Problem solving*. Retrieved from education.com/reference/article/problem-solving.
4. Mayer, R.E. and Wittrock, M.C. (2006). Problem solving. *Handbook of Educational Psychology*, 2, 287-303.
5. McDougal, T. and Takahashi, A. (2014). *Teaching mathematics through problem solving*. Online available at https://www.nais.org/magazine/independent-teacher/fall-2014/teaching-mathematics-through-problem-solving/
6. Powell, S.R. (2011). Solving word problems using schemas: A review of the literature. *Learning Disabilities Research & Practice*, 26(2), 94-108.
7. Taplin, M. (2006). *Mathematics through problem solving*. Retrieved from https://staff.tarleton.edu/brawner/coursefiles/507/Article%20Mathematics%20through%20problem%20Solving%20.pdf
8. Tripathi, P.N. (2009). Problem solving in mathematics: A tool for cognitive development. In *International Conference to Review Research in Science, Technology and Mathematics Education*. Subramaniam, K. and Mazumdar, A. (Ed.),. Mumbai. India.
9. Wilson J.W, Fernandez M.L., and Hadaway, N. (2011). *Mathematical Problem Solving*. Retrieved from http://jwilson.coe.uga.edu/emt725/PSsyn/Pssyn.html.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 147-154

FACTORS AFFECTING THE UTILIZATION OF MATERNAL HEALTH CARE SERVICES AMONG WOMEN IN PUNJAB, PAKISTAN: INSIGHT FROM MICS 2014

Rubama Javed[1], **Anum Fatima**[2] and **Asifa Kamal**[3]
Lahore College for Women University, Lahore, Pakistan
Email: [1]rubamajaved14@gmail.com
[2]anumfatimam@gmail.com
[3]asifa.k53@gmail.com

ABSTRACT

Improvement in maternal health, improves the cycle of health of mothers and children. The disturbance in maternal health may lead to a very serious condition i.e., either the death of mother or child, or in some cases both. This research focuses on the Millennium Development Goal (MDG 5) related to factors affecting the utilization of maternal health care services among women in Punjab. This study focuses on visits of Antenatal Care (ANC), Professionally Assisted Delivery (PAD) and Postnatal Care Health Checks (PNC). Secondary data is used in this research taken from Punjab Bureau of Statistics, Lahore, MICS Data, 2014. Effect of Area of residence, Women's age, Highest level of Women's Education, Children Ever Born (CBE) and Wealth Index Quintiles is studied on three maternal health care related indicators i.e., Visits for Antenatal Care (ANC), Professionally Assisted Delivery (PAD) and Postnatal Care health checks (PNC) using Binary Logistic and Multinomial Logistic regression models. Binary logistic regression model showed that only two independent variables i.e. women's education and wealth index quintile have significant affect upon ANC visits. Multinomial regression model showed that women's education, Area, Wealth Index Quintile have significant contribution towards skilled professionally assisted delivery while the factors women's education (primary) and wealth index quintile are said to be significant for skilled postnatal care health checking. Finally it can be concluded that education plays a vital role to motivate Punjabi women to avail for health care facilities. Strong association with wealth quintiles and utilization of health care facilities recommends that Government should provide free and accessible health care services to Pakistani women.

KEY WORDS

Maternal Health, Visits for ANC, Professionally Assisted Delivery, Postnatal Care Health Checks, Binary Logistic Regression, Multinomial Logistic Regression.

1. INTRODUCTION

According to the World Health Organization (WHO) maternal health is defined as "health of women during pregnancy, childbirth and postpartum period". Decisions related to maternal health during pregnancy have implication in the later life of both mother and child. Decline in maternal mortality is one of the main worldwide agendas (Dar, 2013).

Antenatal Care Visits (ANC), Professionally Assisted Deliveries (PAD) and Post Natal Health Care (PNC) are indicators of utilization of maternal health care services. Antenatal care (ANC) and delivery care services are the indicators of safe maternity. ANC services facilitate pregnant woman and her child to interrelate with the prescribed health care system (Joshi et al., 2014). It helps to detect and treat the pregnancy complications, provision of delivery services and postnatal care services (Pervin et al., 2012; Trujillo et al., 2014). ANC and delivery care had strong and positive associations with health outcomes for both mothers and children (Houweling et al., 2007).

Women should have to visits for antenatal health care checkups for at least four times during her complete tenure of pregnancy (World Health Organization, 2018). A delivery conducted either in a hospital or medical clinic should be assisted by a skilled professional assistant i.e., doctor, nurse, Lady Health Visitor (LHV), Auxiliary Nurse Midwife (ANM) or other qualified assistant. Majority of maternal mortality occurred due to unskilled home deliveries (Yasir et al., 2009). A woman should also have to receive postnatal care after delivery. It is recommended that she should have a postnatal check-up within 42 days after delivery (World Health Organization, 2015).

Pakistan is one of the six countries furnishing to more than half of all maternal deaths globally (Hogan et al., 2010). In spite of development towards achieving the Millennium Development Goals (MDG), Pakistan’s maternal health indicators are some of the worst in the world (National Institute of Population Studies 2013). Hence, it becomes necessary to explore factors which are significantly affecting the utilization of maternal health care services among women of Punjab, Pakistan.

Utilization of reproductive health care services are linked with education, urbanization, religion and employment status (Maqsood, 2009; Midhet and Becker, 2010; Khan et al., 2009). According to Muchie (2017) educational level of women, age, region, child’s birth order and wealth indicator are considerable determinants of ANC visits. Pulok et al. (2016) showed that education of women and their husbands are linked with extensive use of maternal health care services. In general, women with low status are less likely to use modern facilities, whereas women with higher status take the initiative in seeking care for themselves and their children (Owoseni, Jegede, & Ibikunle, (2014). Rutaremwa et al. (2015) investigated determinants of maternal health services utilization and found that richer women were expected to use the ideal package of maternal health services.

Association exists between age and the use of skilled attendants at child birth (Celik and Hotchkiss, 2000). Gap in the utilization of health care services is wider between rural and urban areas (Mezmur, 2017). Education is linked with awareness of health care services which establish demand and use of these services (Mekonnen and Mekonnen, 2003). Maternal education and reading ability was also observed to be correlated with utilization of maternity care services (Gabrysch and Campbell, 2009). Women with higher number of children showed negative and statistically significant impact on health seeking behavior. Higher number of living children represents the experience of women with respect to childbirth; hence it is associated with low utilization of maternal care utilization (Prusty, Gouda and Pradhan, 2015).

OBJECTIVES

To identify the significant demographic factors affecting the utilization of antenatal health care visits, professionally assisted deliveries and postnatal health care facilities.

2. DATA COLLECTION AND METHODOLOGY

Data for current study is taken from Multiple Indicator Cluster Survey, Punjab (MICS, 2014). Respondents for current study are ever married women aged (15-49). Variable, antenatal health care visits (ANC) is dichotomized i.e., (number of ANC visits<4; number of ANC visits ≥ 4). Other two response variables i.e. PAD and PNC are categorized as skilled (if assisted by Doctor), semi-skilled (if assisted by skilled assistant, nurse/midwife, lady health visitor) and un-skilled (if assisted by traditional birth attendant, relative/friend). Age of women, place of residence, wealth quintiles, education of women and children ever born are taken as potential determinants of utilization of health care services.

Bivariate analysis is applied to scrutinize the strength of association between response variables and demographic factors. Chi square test of association is used for this purpose. Binary logistic and Multinomial regression model is fitted to identify significant factors affecting ANC visits, PAD and PNC.

3. STATISTICAL ANALYSIS

Initially, descriptive analysis is carried out, which describes the basic and significant features of the data related to three maternal health care services. Binary logistic regression model and multinomial logistic regression models are used to build models of three dependent variables.

4.1 Descriptive and Bivariate Analysis

The results of various factors across Visits of ANC, Professionally Assisted Delivery and Postnatal Care Health Checks for the data set of 2014, delivered by descriptive analysis, are displayed in Table 1.

Table 1
Percentage Distribution of Health Care Services vs Factors

	Antenatal Care Visits (ANC)			Professionally Assisted Deliveries (PAD)				Postnatal Health Care Checks (PNC)			
	≤4	>4	Chi Square	Skilled	Semi-skilled	Un-skilled	Chi Square	Skilled	Semi-skilled	Un-skilled	Chi Square
Women Education											
Primary	15.3	17.7	349.203 (0.000)	7.3	12.5	15.1	375.731 (0.000)	16.2	5.9	4.4	75.242 (0.000)
Middle	6.7	11.6		5.5	7.5	5.3		11.2	4.9	1.9	
Matric or above	9.8	38.9		18.6	19.3	8.8		41.9	10.5	3.1	
Area											
Urban	16.1	21.8	0.106 (0.745)	7.2	13.9	17.3	32.080 (0.000)	22.6	9.1	5.9	1.342 (0.511)
Rural	26.6	35.4		14.7	19.3	27.6		38.9	13.9	9.6	
Wealth Index Quintiles											
Poor	23.5	12.7	1264.55 (0.000)	7.0	9.8	25.5	1205.144 (0.000)	13.2	5.6	9.1	339.067 (0.000)
Middle	10.0	11.9		5.4	7.2	8.4		11.8	5.5	2.9	
Upper	9.4	32.5		12.9	15.4	8.4		38.2	10.6	3.1	
Children Ever Born											
≤ 2	15.4	20.7	8.369 (0.015)	8.7	11.7	15.6	5.572 (0.233)	22.2	8.6	5.6	4.798 (0.309)
3-4	11.7	17.4		7.8	9.1	11.6		19.4	5.6	3.9	
> 4	15.8	19.0		8.8	11.4	15.3		22.0	7.3	5.2	
Age of Women											
15-19	2.0	1.6	50.776 (0.000)	.8	1.1	1.6	52.993 (0.000)	1.7	0.8	0.4	24.827 (0.016)
20-24	9.2	12.0		5.4	6.7	8.5		13.0	4.8	4.0	
25-29	14.5	21.3		9.1	11.5	14.2		23.3	8.2	4.9	
30-34	10.3	14.9		6.2	8.7	10.2		16.8	5.4	3.1	
35-39	5.1	5.8		2.9	3.3	5.5		6.5	1.9	2.0	
40-44	1.4	1.3		0.7	1.0	1.7		1.7	0.4	0.5	
45-49	0.4	0.3		0.1	0.2	0.5		0.2	0.2	0.1	

From Table 1 it can be observed that women with matric and above education level have higher frequency of 4 or more ANC visits i.e. 38.9%. Visits for ANC are higher for rural areas than for urban areas. Majority of (35.4%) of rural women said that they visited more than 4 times ANC during pregnancy. Also women from upper class having 4 or more ANC visits have highest frequency observed (32.5%). In Punjab, women having 2 or less children that visited 4 or more times for ANC are 20.7%. Moreover middle age mothers (25-29 years) showed highest frequency of getting 4 or more ANC visits rather than other age groups.

Table 1 also shows that 19.3% of women with education matric or above go to semi-skilled professional assistants for delivery. Majority of rural women consulted unskilled professional assistant for delivery i.e., 27.6%. Poor class women showed the highest percentage for unskilled professionally assisted delivery that is 25.5%. Women having less than or equal to 2 children go towards the un-skilled professionally assistant delivery with the percentage of 15.6%. The women with age group 25-29 used maternal health services from unskilled assistants during delivery with percentage 14.2%.

For the postnatal health care checks the women with education matric and above 41.9% attended the skilled postnatal health care professionals. Most of rural women visited the skilled professionals of PNC (38.9%). Regarding the wealth status, 38.2% women from upper class visited skilled PNC professionals. Maximum percentage of women with 2 or less children visited skilled PNC professionals (22.2%) and 23.3% women aged 25-29 years visited the skilled professionals for postnatal health care checks.

It can also be observed in bivariate analysis that all the independent variables except area are significantly associated with number of ANC visits. For professionally assisted delivery, children ever born have not shown significant association with that response variable. Women's education, wealth index quintile and age of women have significant associations with postnatal health care checks.

4.2 Multivariate Analysis

To model visits for ANC Binary Logistic regression is used. While for Professionally Assisted Delivery and Postnatal Care Health Checks Multinomial Logistic regression model is employed. The results are documented in the Table 2:

Table 2
Odds Ratios for Models of Utilization of Maternal Health Care in Punjab, Pakistan

	Antenatal Care Visits (ANC)	**Professionally Assisted Deliveries (PAD)**		**Postnatal Health Care Checks (PNC)**	
		Skilled	**Semi-Skilled**	**Skilled**	**Semi-Skilled**
	Exp(B)	**Exp(B)**	**Exp(B)**	**Exp(B)**	**Exp(B)**
Women Education					
Primary	.416*	.384*	.575*	.316*	1.292
Middle	.560*	.561*	.734*	.519	.820
Area(urban)	1.027	.780*	1.179	1.059	1.190
Wealth index quintile	-				
Poor	.324*	.449*	.391*	.260*	.257*
Middle	.457*	.579*	.614*	.368*	.462*
Children ever born	-				
≤ 2	1.102	.788	.917	1.190	.633
3-4	1.115	1.017	.998	1.004	.773
Age of women	1.014	1.013	1.066	.987	.949
Constant	4.355*	-	-	-	-

*significant at 5%

For Antenatal Care Visits, in case of women education, by keeping matric or above as reference category the odds of women education being primary and middle are declined by 58.4%((1-0.46)*100) and 44% respectively. This means that women with low level of education have fewer odds of 4 or more visits for ANC as compared to higher education. In wealth index quintile, keeping upper class as reference category the odds of having the wealth status as poor and middle is declined by 67.6% and 54.3% respectively hence women with poor and middle wealth status have low odds of ANC visits as compared to rich women.

For Professionally Assisted Deliveries (PAD), Women education is one of the significant factors. Taking matric and above as reference category the odds of having primary education and middle education declined by 61.6% and 43.9% respectively. Hence women with higher level of education have more odds of skilled professionally assisted delivery as compare to unskilled assistant. Area is also a significant factor taking rural as reference category the odds of being from urban area declined by 22% i.e. women from rural areas have more odds for skilled professionally assisted delivery than an unskilled assistant. For wealth index quintile women with rich wealth status have more odds for skilled professionally assisted delivery than an unskilled assistant since the odds of having the wealth status as poor and middle, declined by 55.1% and 42.1% respectively.

For Postnatal Health Care Checks (PNC) the significant factors are Women's education (primary) and Wealth index quintile. Women with higher level of education have more odds of visiting skilled postnatal health care professionals than unskilled professionals. The odds of having primary and middle education have declined by 68.4% and 48.1% respectively. Also the odds of being poor and middle class declined by 74% and 63.2% respectively while we kept upper class as reference category. So we can conclude that women with rich wealth status have more odds for skilled postnatal health care checks than unskilled professionals.

4. COMMENTS AND CONCLUSION

Purpose of this study was to investigate factors responsible for the utilization of maternal health care services among child bearing women in Punjab, Pakistan. Bivariate analysis showed that Women's education, Wealth index quintile, Children ever born and Women's age are significantly associated with visits for Antenatal Care visits. Professionally Assisted Deliveries have significance associations with Women's education, Area, Wealth index quintile and Women's age. Women's education, Wealth index quintile and Women's age are significantly associated factors for Postnatal Health care checks.

Significance of some factors is changed in multivariate analysis. Women's Education and Wealth Index Quintile appeared to be significant factors for ANC and PNC models. Model for PAD have three significant factors namely Women Education, Area and Wealth Index Quintile. Hence it can be concluded that utilization of maternal health care services among women in Punjab, Pakistan is significantly affected by a woman's education level and her wealth status. Women with higher level of education and upper class wealth status are more likely to get 4 or more antenatal care visits and visit skilled professionals for deliveries and postnatal health care checks.

Finally it can be concluded that education plays a vital role to motivate Punjabi women to avail health care facilities. Strong association with wealth quintiles and utilization of health care facilities recommends that Government should provide free and accessible health care services to Pakistani women.

REFERENCES

1. Celik, Y. and Hotchkiss, D.R. (2000). The socio-economic determinants of maternal health care utilization in Turkey. *Social Science & Medicine*, 50(12), 1797-1806.
2. Fauzia, M. (2009). *Effect Of Socio-Cultural Exclusion and Community Level Factors on Reproductive Level; Comparison Between Urban and Rural Pakistani Women* (Doctoral dissertation, University of Punjab).
3. Gabrysch, S. and Campbell, O.M. (2009). Still too far to walk: literature review of the determinants of delivery service use. *BMC Pregnancy and Childbirth*, 9(1), 34. doi: 10.1186/1471-2393-9-34
4. Hogan, M.C., Foreman, K.J., Naghavi, M., Ahn, S.Y., Wang, M., Makela, S.M., Lopez, A.D., Lozano, R. and Murray, C.J. (2010). Maternal mortality for 181 countries, 1980-2008: A systematic analysis of progress towards Millennium Development Goal 5. *The lancet*, 375(9726), 1609-1623.
5. Houweling, T.A., Ronsmans, C., Campbell, O.M. and Kunst, A.E. (2007). Huge poor-rich inequalities in maternity care: An international comparative study of maternity and child care in developing countries. *Bulletin of the World Health Organization*, 85, 745-754.
6. Joshi, C., Torvaldsen, S., Hodgson, R. and Hayen, A. (2014). Factors associated with the use and quality of antenatal care in Nepal: a population-based study using the demographic and health survey data. *BMC Pregnancy and Childbirth*, 14(1), 94. http://www.biomedcentral.com/1471-2393/14/94
7. Khan, Y.P., Bhutta, S.Z., Munim, S. and Bhutta, Z.A. (2009). Maternal health and survival in Pakistan: issues and options. *Journal of Obstetrics and Gynecology Canada*, 31(10), 920-929.
8. Mekonnen, Y. and Mekonnen, A. (2003). Factors influencing the use of maternal healthcare services in Ethiopia. *Journal of Health, Population and Nutrition*, 21(4), 374-382.
9. Mezmur, M., Navaneetham, K., Letamo, G. and Bariagaber, H. (2017). Socioeconomic inequalities in the uptake of maternal healthcare services in Ethiopia. *BMC Health Services Research*, 17(1), 367.
10. Midhet, F. and Becker, S. (2010). Impact of community-based interventions on maternal and neonatal health indicators: Results from a community randomized trial in rural Balochistan, Pakistan. *Reproductive Health*, 7(1), 30. http://www.reproductive-health-journal.com/content/7/1/30
11. Muchie, K.F. (2017). Quality of antenatal care services and completion of four or more antenatal care visits in Ethiopia: a finding based on a demographic and health survey. *BMC Pregnancy and Childbirth*, 17(1), 300. https://doi.org/10.1186/s12884-017-1488-0

12. National Institute of Population Studies (2013). *Pakistan Demographic and Health Survey Report*. National Institute of Population Studies, Islamabad, Pakistan and Macro International Inc. Calverton: Maryland, USA.
13. Owoseni, J.S., Jegede, L.I. and Ibikunle, A.M. (2014). Socio-economic status and utilization of healthcare facilities in rural Ekiti, Nigeria. *Standard Research Journal of Medicine and Medical Sciences*, 2(1), 001-043.
14. Pervin, J., Moran, A., Rahman, M., Razzaque, A., Sibley, L., Streatfield, P.K., Reichenbach, L.J., Koblinsky, M., Hruschka, D. and Rahman, A. (2012). Association of antenatal care with facility delivery and perinatal survival–a population-based study in Bangladesh. *BMC Pregnancy and Childbirth*, 12(1), 111. https://doi.org/10.1186/1471-2393-12-111.
15. Prusty, R.K., Gouda, J. and Pradhan, M.R. (2015). Inequality in the utilization of maternal healthcare services in Odisha, India. *International Journal of Population Research*, Article ID 531485, 10 pages, http://dx.doi.org/10.1155/2015/531485.
16. Pulok, M.H., Sabah, M.N.U., Uddin, J. and Enemark, U. (2016). Progress in the utilization of antenatal and delivery care services in Bangladesh: where does the equity gap lie? *BMC Pregnancy and Childbirth*, 16(1), 200. https://doi.org/10.1186/s12884-016-0970-4
17. Rutaremwa, G., Wandera, S.O., Jhamba, T., Akiror, E. and Kiconco, A. (2015). Determinants of maternal health services utilization in Uganda. *BMC Health Services Research*, 15(1), 271. https://doi.org/10.1186/s12913-015-0943-8
18. Trujillo, J.C., Carrillo, B. and Iglesias, W.J. (2013). Relationship between professional antenatal care and facility delivery: an assessment of Colombia. *Health Policy and Planning*, 29(4), 443-449. https://doi.org/10.1093/heapol/czt033.
19. World Health Organization (2015). Postnatal care for mothers and newborns. Highlights from the *World Health Organization 2013 Guidelines*. *RHR/15.05, Geneva*.
20. World Health Organization (2018). Health situation and trend assessment. Antenatal care coverage. http://www.searo.who.int/entity/health_situation_trends/data/mnch_anc-with-atleast-4-visits/en/.
21. World Health Organization and UNICEF (2015). *Pregnancy, childbirth, postpartum and newborn care: a guide for essential practice*. 3rd edition. Geneva, Switzerland: WHO.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 155-159

PERCEPTIONS OF TEACHERS REGARDING IMPACT OF CO-CURRICULAR ACTIVITIES ON SOCIAL SKILLS OF STUDENTS

Syeda Mishal Akhtar
University of Education, Lahore, Pakistan
Email: syedamishalzehra@gmail.com

ABSTRACT

The purpose of this paper is to know the perceptions of teachers regarding impact of co-curricular activities on social skills of students. This descriptive exploratory research drew a random sampling of male and female who had given interviews about how co-curricular activities effects social skills of students. Teachers belong to Government sector School. Face to face interviews was taken. The transcripts of interviews (i.e. recorded audio) were made in order to ensure authenticity of data.

There is good effect of co-curricular activities on social skills of student. Students become confident, energetic due to co-curricular activities. Their communication skills enhance. They become social. They develop leadership qualities. Data revealed that there is positive impact of co-curricular activities on social skills of students.

1. INTRODUCTION

Extracurricular is a combination of the Latin prefix extra- meaning "on the outside" and the word curriculum, meaning "a running, course, career."

Extracurricular are those activities that held outer the realm of the set of courses of school or university education, performed by students. Extracurricular activities are present intended for all pupils. Students and Staff organize these activities under faculty support. The role of extracurricular activities in a student's growth has been given much importance. A student's social life is as significant as their academic life and for their development. There is so much pressure on students, especially in secondary schools, to perform well in their exams and attain high grades.

First time when Children come across to community is started in Infancy stage and it becomes more complex when they learn language. (Hetherington et al., 2006).

The reason behind a learners' failure is that he/she has lack of communication skill. He is unable to create good interaction with others. (Abdul Rahman, 2010). Developing pupils connects with their age fellows at 4 years age. When children progress through elementary and middle school, they desire to be accepted by their same age peers and to share activities with them (Rubin, Bukowski, & Parker, 2006).

Communal (outer of school) activities or extracurricular activities are such activities that engaged pupils with others (Larson & Verma, 1999; Gilman, 2001; King et al., 2003).

2. RESEARCH QUESTION

1. What is impact of extra-curricular activities on student social skill?
   A. How many students are participating in activities?
   B. In what activities are the students choosing?
   C. How can schools minimize the barriers to participation in activities to be more inclusive of all students?
2. How extracurricular activities effects student social skill?

3. RESEARCH METHODOLOGY

This study is stand on Qualitative approach. “Qualitative research is designed to reveal a target audience’s range of behavior and the perceptions that drive it with reference to specific topics or issues. It uses in-depth studies of small groups of people to guide and support the construction of hypotheses”.

Researcher used case study strategy to meet up my research “Robert K. Yin defines the case study research method as an empirical inquiry that investigates a contemporary phenomenon within its real-life context”.

Researcher used interview to meet my research question. Interviews are particularly useful for getting the story behind a participant’s experiences (McNamara, 1999).

Data was Collected through Government school located at Johar Town Lahore. Both Male and females were included in this study. Sample was selected randomly.

4. DATA ANALYSIS

	Communication Skill	**Confident**	**Leadership Qualities**	**Enthusiastic**	**Team Wok**
T1					1
T2				2	2
T3			3	3	
T4	4	4			
T5		5		5	
T6			6	6	
T7	7				
T8	8				
T9		9	9		
T10	10	10			
T11			11		11
T12	12	12			
T13		13		13	13
T14	14	14			14
Total	**06**	**07**	**04**	**05**	**05**

Communication Skill:

Communication skill of students enhance due to extracurricular activities as **Teacher 1** said their communication skills improve. They learn how to convey our message to others**. Teacher '4'** thought that "they learn how to talk with other as a result their communication skill improves." **Teacher 7** said that "due to extracurricular activities student become known to everyone. People will talk to him, so his communication skill enhances." **Teacher 10** supposed that "their communication skills improve. They learn how to convey our message to others." **Teacher 12** tells me "student go and meet different public sector students, his communication skill enhances." **Teacher 14** believed that Extracurricular activities are very good way to enhance communication skill as well as growth and learning of students.

Confident:

Students gain confident when they participate in extracurricular activities. As **teacher 4** inform me "we take students outside the school, they go for playing cricket. They become confident." **Teacher 5** notifies me "students become confident and energetic because of extracurricular activities." **Teacher 9** enlighten me "class wise arrangement can motivate the students to participate". Student gets more confident." **Teacher 10** acquaint with me that "if we offer them different activities like running, debates, drawing and different competition, they'll become confident." **Teacher 12** whispered that "they went to different schools their confidence increased." **Teacher 13** tells me "students become social and confident because of it." **Teacher 14** assumed that "because of these activities student become more social in their circle, in their school. Student can enhance their confidence level."

Leadership Qualities:

Learners develop Leadership Qualities due to co-curricular activities. As **Teacher 3** thought student become able to lead other people due to co-curricular activities. **Teacher 6** believed that co-curricular activities enhance student's leadership qualities. **Teacher 9** whispered it develops leadership qualities in students. **Teacher 11** assumed that they gain leadership qualities.

Enthusiastic:

Teacher 2 keenly study after performing these activities**. Teacher 3** believed Student become eager for studies**. Teacher 5** said after co-curricular activities students shows excitement towards studies. **Teacher 6** alleged Students become enthusiastic because of extracurricular activities. **Teacher 13** told Student will become active because of it.

Team Work:

Teacher 1 understood they work in teams and they learn how to work as a team. **Teacher 2** assumed that students learn how to solve their peers' issues. If there is some problem, he actively tries to solve it. Due to these activities' students become fit, positive and socially active. **Teacher 11** told they learn win and lose is a part of life. It is not important that every time we win, we have to accept our defeat. **Teacher 13** assumed Students become social and confident because of it. They help others. They understand how to share and cooperate with others. **Teacher 14** they will be able to differentiate themselves they're able to socialize themselves in school and after school**.**

5. DISCUSSION

Researcher found that all my participants said that there is good impact of extracurricular activities on student social skills. When literature is reviewed it is known from previous studies that activities. a benefit is a positive association between participating in various forms of physical activities and positive self-perception (Daley & Leahy, 2003).

Students become confident, energetic because of these activities. Their communication skills enhance. They become social. They develop leadership qualities. When literature is observed, consequences are similar to my research students have interactions with competent adult figures, it may lead to achievement of goals, development and improvement of skills, and enhanced social and leadership opportunities (Gilman, 2004).

6. CONCLUSION

To conclude my result that at there is good impact of extracurricular activities on student’s social skill.

Teachers said that students like hockey, cricket, football, badminton, running, rope pulling, and sand bowling pins. Few teachers said that students like Drawing and role play and matches are the favorite. One teacher said that students like Naat e Rasool and Recite of Holy Quran.

Most of the teachers believed that believed that 90-100% students participate. Few teachers believed that below 60% students participate in extracurricular activities.

Most of the teachers said that school does not have many facilities; we contribute and provide facilities to students. Few teachers said that we ask donor to provide facilities in our school. Some teachers said that school can manage through existed facilities by set timetable of school.

They learn how to work in team, their communication skill enhances, and they become confident, enthusiastic, active and participative in social activities. They develop leadership qualities and learn how to work in team.

REFERENCES

1. Brooks, B.A., Floyd, F., Robins, D.L. and Chan, W.Y. (2015). Extracurricular activities and the development of social skills in children with intellectual and specific learning disabilities. *Journal of Intellectual Disability Research*, 59(7), 678-687.
2. Moriana, J.A., Alós, F., Alcalá, R., Pino, M. J., Herruzo, J. and Ruiz, R. (2006). Extra-curricular activities and academic performance in secondary students. *Electronic Journal of Research in Educational Psychology*, 4(1), 35-46.
3. Wilson, N. (2009). *Impact of Extracurricular Activities on Students.* Research Paper for MS Degree, University of Wisconsin-Stout.
4. Hetherington, E.M., Parke, R.D., Gauvain, M. and Locke, V.O. (2006). Expanding the social world : peers and friends. *Child psychology*: *A Contemporary Viewpoint* (pp. 506-545). Boston: McGraw-Hill.

5. Gilman, R. (2004). Structured extracurricular activities among adolescent findings and implications for school psychologists. *Psychology in the Schools*, 41(1), 31-41.
6. Daley, A. and Leahy, J. (2003). Self-perceptions and participation in extracurricular physical activities. *The Physical Educator*, 60(2), 13-19.
7. McNamara, C. (1999). *General Guidelines for Conducting Interviews, Authenticity Consulting*, LLC, Retrieved from: http://www.managementhelp.org/evaluatn/intrview.htm
8. King, G., Law, M., King, S., Rosenbaum, P., Kertoy, M. and Young, N. (2003). A conceptual model of the factors affecting the recreation and leisure participation of children with disabilities. *Physical & Occupational Therapy in Pediatrics*, 23(1), 63-90.
9. Rubin, K., Bukowski, W. and Parker, J. (2006). Peer interaction and social competence. In W. Damon & R.M. Lerner (Series Eds.) & N. Eisenberg (Vol. Ed.), *Handbook of Child Psychology*: Vol. 3 (6th ed.). New York: Wiley
10. Rahman, R.A. (2010). *Kesan Latihan Kemahiran Berkomunikasi Terhadap Konsep Kendiri Pelajar Di Sebuah sekolah Menengah* (In English: Impact of Communication Skills Training on Individual Self-Concepts in A Secondary School). *Jurnal Pendidikan Malaysia*.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 161-168

DETERMINATION OF RISK FACTORS THAT CAUSE ANTISOCIAL BEHAVIOUR AMONG THE STUDENTS OF PUBLIC SECTOR UNIVERSITIES IN LAHORE

Anam Rafi[1], **Mahnaz Makhdum**[1] and **Mubbasher Munir**[2]
[1] Lahore College for Women University, Lahore, Pakistan.
Email: anamrafi17@gmail.com; minz_mak@hotmail.com
[2] University of Management and Technology Lahore Pakistan.
Email: muasher.munir@umt.edu.pk

ABSTRACT

The study of antisocial behavior problem was conducted to determine the prevalence and severity of antisocial behavior problem and identifying factors influence the risk of antisocial behavior problem. A sample of 500 students selected randomly from three public sector Universities of Lahore, Pakistan. The data was collected by using well designed questionnaire. The age ranged from 16 to 25 years. The STAB scores were computed.

The anti-social behavior was found to be prevalent in one fifth of the population. It was more prevalent in boys as compare to girls. The test of association was performed to test the significance of different risk factors and antisocial behavior problem. To determine the effect of different factors on antisocial behavior, the binary logistic regression model was applied on all data. The factors low supervision of parents, complex with friends, poor medium of instruction and carelessness about studies are risk factors which were found to be significant contributors towards anti-social behavior among youth.

1. INTRODUCTION

The anti-social behavior (ASB) was defined as "acting in a manner that caused or was likely to cause harassment, alarm or distress to one or more persons not of the same household as the perpetrator" by crime and disorder act (1998). The anti-social behavior is a serious problem among children, students and adults.

This is becoming a big problem with passage of time because it has negative impact on lives. The anti-social behavior (ASB) may include many things such as; criminal activities, drug dealing, physical assail, threats of violence, hurting someone, damaging property, abusing etc.

There are many factors of anti-social behavior that are affecting the lives of students. The four main factors in this research study were family factors, personal factors, peer or friend factors and school or education factors. Each factor has sub factors which define the impact of antisocial behavior on students as listed in Table 1.

Table 1
Cross Tabulation of Demographic Factors with Antisocial Behavior

S#	Variables	Label	Coding	Anti-Social Behavior		Total
				No(396)	Yes(104)	(500)
1	GPA of Respondents	GPA	<=2.5	16.91%	32.69%	20.2%
			>2.5 & <=3.5	62.12%	61.54%	62%
			>3.5	20.96%	5.77%	17.8%
2	Current Age Group	Age	<=18	10.35%	10.57%	10.4%
			>18&23<	68.69%	66.35%	68.2%
			>=23	20.96%	23.07%	21.4%
3	Gender of the Respondents	Gender	Male	50.06%	75%	60%
			Female	43.94%	25%	40%
4	Medium of Schooling	School	English Medium	65.91%	53.85%	63.4%
			Urdu Medium	34.09%	46.15%	36.6%
5	Parents Marital Relationship	Parents Relation	Living together	93.68%	84.62%	91.8%
			Divorce	2.27%	3.84%	2.6%
			Living separate	4.04%	11.54%	5.6%

The demographic factors GPA, age, gender, medium of school and parents' marriage relationship has been taken to study the effect of anti-social behavior University students.

OBJECTIVES

- To determine prevalence of antisocial behavior among students of public sector universities in Lahore.
- To determine risk factors which are associated to the antisocial behavior problems.
- Find out the strength of association between risk factors and occurrence of antisocial behavior problems.
- Use an appropriate model to determine the significant factors of antisocial behavior for prediction purposes.

2. RESEARCH METHODOLOGY

Population under Study and Sampling Scheme

The population under study is based on students of public sector universities in Lahore. The population size is number of all bachelor students in three public sector universities. Three universities were randomly selected from nine public sector universities. Simple random sampling technique was used for selection of sample.

Study Instrument

In this study the information collected from students by using a well-designed questionnaire. The questionnaire was developing by taking help of research literature (Parenting style, Peer pressure and the formation of Antisocial Behavior by Erinisha L. Johnson, 2012). The questionnaire includes demographic factors, STAB (Sub Type Antisocial Behavior) questions and risk factors (parental, personal, peer/friends and schooling). The factors were divided into sub factors. The STAB developed by taking help of (Development and Validation of the Subtypes of Antisocial Behavior Questionnaire by Burt S.A et al 2009).

STAB Scores and Score of Risk Factors

The 5 point likert scale “1 No way” to “5 generally” is use for STAB questions. The 13 questions were used to detect antisocial behavior problems in respondents. The average scores of 13 questions were calculated for each respondent. The students have average scores less than equals to 3 were categorizing as do not have antisocial behavior problem and greater than 3 were categorizing as have antisocial behavior problem.

Scores for each risk factor were computed by averaging responses of all questions asked on five-point likert scale for that specific risk factor. Respondents with score less than or equal to three are categorized as not possessing the risk factor and those having score greater than three are categorized as possessing the factor.

3. RESULTS

Descriptive Analysis

Demographic Risk Factors

The 61.54% students with GPA between 2.5 & 3.5 suffered from antisocial behavior problem. The 66.3% students of age between 18 and 23 years suffered with antisocial behavior problem. The 75% male students and 25% female students suffered with anti-social behavior problem. The males were more prevalent as compare to females that were suffering with antisocial behavior problem.

Cross Tabulation of Family Factors with Antisocial behavior were performed given in Table 2.

Table 2
Cross Tabulation of Family Factors with Antisocial Behavior

S#	Variables	Label	Coding	Antisocial Behavior		Total (500)
				No (396)	Yes (104)	
1	Low Income Level of Parents	Income level (F1)	No	71.96%	52.88%	68%
			Yes	28.03%	47.11%	32%
2	Ineffective Upbringing of Children	Children’s Upbringing (F2)	No	75%	52.88%	70.4%
			Yes	25%	47.11%	29.6%
3	Parent’s Low Supervision	Low Supervision (F3)	No	78.79%	52.88%	73.4%
			Yes	21.21%	47.11%	26.6%
4	High Conflict of parents and Home Environment	Parent’s Disputes (F4)	No	85.10%	65.38%	81%
			Yes	14.89%	34.62%	19%
5	Low Encouragement	Low Encouragement (F5)	No	77.02%	49.04%	71.2%
			Yes	22.98%	50.96%	28.8%
6	Anti-social Parents	Antisocial Parents (F6)	No	75.50%	54.81%	71.2%
			Yes	24.49%	45.19%	28.8%
7	Imposing Restrictions on Children	Imposing Restrictions (F7)	No	31.06%	16.35%	28%
			Yes	68.94%	83.65%	72%

Family Factors

In family factors 52.9% students suffered with antisocial behavior problem whose parents did not have low supervision on them and 47.1% whose parents have low supervision on them they feel shy, hesitation and did not feel comfortable. The 49% students suffered with antisocial behavior problem whose parents encourage them and 51% suffered whose parents low encourage them which shows that students feel low encouragement and they do not take interest in studies and other things. The 16.3% students suffered with antisocial behavior problem whose parents did not impose restrictions on them and 83.7% suffered whose parents impose restrictions on them which shows that they feel uncomfortable they cannot chose subjects of their choice which effect their studies.

Cross Tabulation of Different Personal Factors with Antisocial behavior are given in Table 3.

Table 3
Cross Tabulation of Personal Factors with Antisocial Behavior

S#	Variables	Labels	Coding	Antisocial Behavior		Total
				No(396)	Yes(104)	(500)
1	Lack of Confidence	Confidence (P1)	No	62.37%	51.92%	60.2%
			Yes	37.37%	48.07%	39.6%
2	Lack of Skills	Skills (P2)	No	72.22%	51.92%	68%
			Yes	27.78%	48.07%	32%
3	Afraid of Rejection Experience	Rejection Experience (P3)	No	79.29%	60.57%	75.4%
			Yes	20.71%	39.42%	24.6%
4	Abuse Others	Abuse (P4)	No	77.02%	49.04%	71.2%
			Yes	22.98%	50.96%	28.8%
5	Feel Discrimination	Discrimination (P5)	No	74.49%	52.88%	70%
			Yes	25.51%	47.11%	30%

Personal Factors

The 62.5% students suffer with antisocial behavior problem that did not have lack of confidence and 48.1% suffer who had lack of confidence. The 72.2% students did not suffer with antisocial behavior problem that did not have lack of skills and 48.1% students suffer that had lack of skills. The77% students that did not abuse other not suffer with antisocial behavior and 51% suffer that abuse on others. The 74.5% students did not feel discrimination not suffer with antisocial behavior and 47.1% feel discrimination and suffer with antisocial behavior

Cross Tabulation of Different Peer/Friends Factors with Antisocial are given in Table 4.

Table 4
Cross Tabulation of Peer/Friends Factor with Antisocial Behavior

S#	Variables	Labels	Coding	Antisocial Behavior No (104)	Antisocial Behavior Yes (396)	Total (500)
1	Peer Rejection	Rejection (FR1)	No	59.84%	42.31%	56.2%
			Yes	40.15%	57.69%	43.8%
2	Relationship with friends	Friendship (FR2)	No	44.94%	34.62%	42.8%
			Yes	54.79%	65.38%	57%
3	Peer pressure	Pressure (FR3)	No	75%	50.96%	70%
			Yes	25%	49.04%	30%
4	Complex with friends	Complex (FR4)	No	21.97%	53.85%	28.6%
			Yes	78.03%	12.12%	71.4%

Peer/ Friends Factors

The 57.7% students have peer rejection, suffer with antisocial behavior problem. The 75% students did not have pressure of friends, not suffer with antisocial behavior problem. The 86.6% students were not complex with friends, not suffer with antisocial behavior problem.

Cross Tabulation of Different School/ Education Factors with antisocial behavior are given in Table 5.

Table 5
Cross Tabulation of Schooling/Education Factor with Antisocial Behavior

S#	Variables	Labels	Coding	Antisocial Behavior No (104)	Antisocial Behavior Yes (396)	Total (500)
1	Ineffective response of teachers	Ineffective response (S1)	No	72.22%	60.57%	69.8%
			Yes	27.78%	39.42%	30.2%
2	Uncomfortabl e atmosphere of institute	Uncomfortable atmosphere (S2)	No	72.98%	47.11%	67.6%
			Yes	27.02%	52.88%	32.4%
3	Medium of Instruction	Medium (S3)	No	74.75%	43.26%	68.2%
			Yes	25.25%	56.73%	31.8%
4	Gender Conscious	Gender Conscious (S4)	No	69.44%	50%	65.4%
			Yes	30.55%	50%	34.6%

School/ Education Factors

The 56.7% students suffer with antisocial behavior problem that were have poor medium of instruction they feel hesitation in communication with others. The 50% students suffered with antisocial behavior that was gender conscious.

Bivariate Analysis

The association of all risk factors was calculated with antisocial behavior and to determine the strength of association between risk factors and occurrence of antisocial behavior, phi and contingency coefficient were computed for nominal variables and Kendall's tau-b was calculated for ordinal variables in Table 6.

Table 6
Bivariate Analysis

Predictor	Chi Square	P-Value	Kendall's Tau-b
GPA	-	0.000	-0.175
Gender	12.311	0.000	-
Age	-	0.878	-0.006
Medium of school	5.165	0.023	-
Parent's marital status	9.792	0.007	-
Low income level of parents	13.788	0.000	-
Ineffective upbringing of children	19.332	0.000	-
Parent's low supervision	28.307	0.000	-
High conflict of parents and Home environment	20.805	0.000	-
Low Encouragement	31.451	0.000	-
Anti-social parents	17.207	0.000	-
Imposing Restrictions on children	8.846	0.003	-
Lack of confidence	3.871	0.049	-
Lack of skills	15.598	0.000	-
Afraid of rejection experience	15.555	0.000	-
Abuse others	31.451	0.000	-
Feel Discrimination	18.317	0.000	-
Peer rejection	10.295	0.001	-
Relationship with friends	3.669	0.055	-
Peer pressure	22.665	0.000	-
Complex with friends	40.986	0.000	-
Ineffective response of teachers	5.299	0.021	-
Uncomfortable atmosphere of institute	25.158	0.000	-
Medium of Instruction	37.633	0.000	-
Gender Conscious	13.762	0.000	-

Binary Logistic Regression

The binary logistic regression was fitted considering all risk factors to estimate the effect of each factor on antisocial behavior. The forward stepwise (Likelihood Ratio) LR method was used to conclude the predictive strength of various risk factors.

Table 7
Variables in the Equation

	B	S.E	Wald	d.f	Sig	Exp(B)	95.0% C.I. for EXP(B)	
							Lower	Upper
X1D GPA	-.837	.218	14.747	1	.000	.433	.282	.664
Parent's low supervision	.733	.256	8.196	1	.004	2.081	1.260	3.436
Low Encouragement	.581	.264	4.852	1	.028	1.787	1.066	2.995
Complex with friends	.875	.261	11.213	1	.001	2.399	1.437	4.003
Medium of Instruction	.692	.264	6.882	1	.009	1.999	1.191	3.353
Constant	.082	.660	.015	1	.902	1.085		

As shown in Table 7 the variables GPA, parent's low supervision, low encouragement, complex with friends and medium of instruction appears as significant contributors towards antisocial behavior problems for adolescents as their p-values are less than 0.05.

The risk factor GPA is taken as numeric variable. The coefficient for GPA has negative association with antisocial behavior problems.

The risk factors parent's low supervision, low encouragement, complex with friends and medium of instruction were taken as categorical variable. Since the coefficient of these variables were positive therefore the students whose parents have low supervision on them, low encourage, they have complex with friends and they have poor medium of instruction have more chance of having antisocial behavior problems.

4. CONCLUSION

The antisocial behavior was found in students of public sector universities in Lahore. Antisocial behavior was prevalent in about one fifth of the population. It was more prevalent in males as compare to females. Logistic regression analysis was performed. The GPA was negatively associated with antisocial behavior problems. It shows that students have good GPA have less chance of suffering with antisocial behavior problem. Parent's low supervision causes antisocial behavior problem in students. The students whose parents have low supervision on them have more chance of having antisocial behavior problems than those whose parents have supervision on their children. The students suffering with problem of complex with friends have more chance of antisocial behavior problem. The students who have poor medium of instruction have more chance of having antisocial problems than those students who do not have poor medium of instruction.

REFERENCES

1. Agresti, A. (2002). *Categorical Data Analysis* (2nd eD.). John Wiley & Sons.
2. *Anti-social Behaviour Act 2003*. Available at https://www.legislation.gov.uk/ukpga/2003/38/pdfs/ukpga_20030038_en.pdf
3. Anti-social Behaviour, Crime and Policing Act July 2014. Available at https://www.gov.uk/government/collections/anti-social-behaviour-crime-and-police-bill
4. Bewick, V., Cheek, L. and Ball, J. (2005). Statistics Review 14: Logistic Regression. *Critical Care*, 9(1), 112-118.
5. Burt, S.A. and Donnellan, M.B. (2009). Development and validation of the Subtypes of Antisocial Behavior Questionnaire. *Aggressive behavior*, 35(5), 376-39.
6. Cresswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. New York, NY: Prentince Hall.
7. Day, D.M. and Wanklyn, S.G. (2012). *Identification and operationalization of the major risk factors for antisocial and delinquent behaviour among children and youth*. National Crime Prevention Centre, Public Safety Canada.
8. Harradine, S., Kodz, J., Lemetti, F. and Jones, B. (2004). *Defining and measuring anti-social behaviour*. Home Office, Research, Development and Statistics Directorate. Available at https://www.gov.uk/government/uploads/system/uploads/attachment.../dpr26.pdf
9. Johnson, E.L. (2012). Parenting Styles, Peer Pressure, and the formation of Antisocial Behavior. Honors Theses. Paper 101.
10. McGee, T.R., Wickes, R., Corcoran, J., Bor, W. and Najman, J. (2011). Antisocial behaviour: An examination of individual, family, and neighbourhood factors. *Trends and Issues in Crime and Criminal Justice*, (410), 1.
11. *Regression Diagnostics*. Available at http://www.oxfordjournals.org/our_journals/tropej/online/ma_chap5.pdf
12. Robinson, C. and Schumacker, R.E. (2009). Interaction Effects: Centering, Variance Inflation Factor, and Interpretation Issues. *Multiple Linear Regression Viewpoints*, 35(1), 6-11.
13. Schüppert, A. (2009). *Binomial (or Binary) Logistic Regression*. Statistics Seminar, Spring.
14. T'ng, S.T., Baharudin, R. and Ismail, Z. (2015). Antisocial behaviour in Malaysian adolescents: assessing measurement equivalence across gender differences. *Child Indicators Research*, 8(3), 537-550.
15. Tavakol, M. and Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55.
16. Wikipedia.org. *Cronbach's Alpha*: http://en.wikipedia.org/wiki/Cronbach's_alpha

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 169-174

A CROSS SECTIONAL STUDY OF ASSOCIATED FACTORS OF TOOTH DECAY

Mishal Naeem[1] and **Mahnaz Makhdum**[2]
Lahore College for Women University, Lahore, Pakistan
Email: [1]mishalbutt46@gmail.com
[2]minz_mak@hotmail.com

ABSTRACT

A cross-sectional epidemiological study of tooth decay was performed to determine the prevalence, factors influencing tooth decay and severity of tooth decay. A sample of 1087 student were selected randomly from 3 general public-sector universities of Lahore, Pakistan during the period 5 February, 2018 to 30 April, 2018. Data was collected through well designed questionnaire. Tests of association were used to test the significance of various risk factors with tooth decay. To see the effect of different risk factors on tooth decay, statistical technique of logistic regression was used. Over all tooth decay prevalence was found to be 431 (39.7%) in students. Logistics regression showed results that genetics, grinding teeth, rough bristles, brushing after eating, flossing, mouth wash, cleaning of molars, fluoride enriched toothpaste, sugary foods, soft drinks or beverages, medications, eating disorder, acidity, stomach problem, dry mouth and dental plaque significantly causing tooth decay. Milk, eating disorder, diabetes and heartburn had no effect on tooth decay. This study concludes that flossing, mouth wash, brushing after eating, cleaning of molars plays a protective role against acne prevalence.

1. INTRODUCTION

Tooth decay occurs when bacteria in mouth converts sticky food into acid know as plaque that causes cavities. When proper steps of oral hygiene are not undertaken to remove formed plaque than it leads to higher chances of having tooth decay. Tooth decay is very common problem worldwide and there is only little control over it.

The World Health Organization (WHO) has ranked it as number three among all chronic non-communicable diseases that require worldwide attention for prevention and treatment. Mulu et al. (2014) had studied the factors associated with dental caries among students of Bahir Dar City which concludes that children who did not clean their teeth properly were 2.6 times at more risk of dental caries compared to those who did it properly. The chance of having dental caries is 6.3 times higher among students who have toothache. Masson et al. (2010) had studied effects of sugar intake on tooth decay who concluded that non-dietary soft-drinks were found to be positively associated with increase chances of tooth decay treatment and tooth brushing were also found to be related with tooth decay treatment.

The factors that have taken into account for studying effects on tooth decay were age, gender, genetics, smoking, vape or shisha, habit of grinding teeth, visit to dentist, brushing, brushing after eating, flossing, mouth wash, tooth paste, cleaning of molars, junk food, sugary food, carbohydrate food, soft-drink or beverages, medications, milk, eating disorder, diabetes, acidity, stomach problem, dental plaque and dry mouth. The main objective of this study is to determine risk factors which contribute to the problem of tooth decay among youth of Lahore.

2. MATERIAL AND METHODS

The study was conducted in three general public-sector co-education universities of Lahore, Pakistan. The sample of 1087 had been collected through simple random sample from the population of 50336 students from co-education public sector universities which were Punjab University, Government College University, and Education University. A well-structured questionnaire comprising of socio-demographic factors, oral habits, eating habits, clinical factors designed by researcher was used to collect data.

Statistical analysis has been performed using SPSS 16.0. Cross tabulations of various risk factors with tooth decay are constructed and interpreted. In analytical section, bivariate analysis was performed. The binary logistic regression model was fitted using all risk factors to determine the effect of each risk factor on tooth decay.

3. RESULTS

Though cross tabulation it was obtained that 656 (60.3%) students were not suffering from tooth decay and 431 (39.7%) students had tooth decay. The students who did not have any family background of tooth decay were found to be 549 (50.5%) while the students who had tooth decay in their family background were found to be 538 (49.5%). Of 38 students who visit dentist once in few months, 23 (60.5%) did not have tooth decay whereas 16 (39.5%) students suffered from tooth decay. Among 1094 students who visit when they suffered from any dental problem, 628 (59.9%) were not suffering tooth decay whereas 421 (40.1%) had tooth decay. The result of bivariate analysis is shown in (Table 1).

Table 1
Bivariate Analysis

Predictor	Chi-Square	P-Value	Predictor	Chi-Square	P-Value
Age group	7.858	0.020	Sugary Food	24.422	0.000
Genetics	15.891	0.000	Carbohydrate Food	6.318	0.012
Smoking	4.465	0.035	Soft drinks or Beverages	11.227	0.000
Grinding Teeth	14.058	0.000	Milk	2.135	0.144
Visit to Dentist	0.007	0.935	Medications	4.903	0.027
Rough bristles of Brush	69.998	0.000	Eating Disorder	2.034	0.308
Brushing after Eating	11.749	0.001	Diabetes	0.082	0.642
Flossing	47.787	0.000	Acidity	17.020	0.000
Mouth Wash	38.016	0.000	Heart burn	1.011	0.315
Tooth Paste	2.022	0.155	Stomach problem	28.801	0.000
Clean Molars	23.39	0.000	Dental Plaque	14.362	0.000
Junk Food	7.691	0.006	Dry Mouth	12.692	0.000

The risk factors age, genetics, grinding teeth, bristles of brush, brushing after eating, flossing, mouth wash, cleaning of molars, junk food, carbohydrate enriches food, sugary food, consumption of soft drinks or beverages, acidity, heart burn, stomach problem, dental plaque, dry mouth were found to be statistically significantly associated with tooth decay. Age of the respondent and tooth decay were found to be statistically significantly associated and results show that as age of the respondent increases the risk of tooth decay increases.

The predictive logistic regression model is:

Tooth decay = 0.030 + 0.30.9 (genetics) + 0.529 (grinding teeth)
+ 0.815 (hard bristles) - 0.470 (brushing after eating) - 0.889 (flossing)
- 0.750 (mouthwash) - 0.500 (floruided toothpaste)
- 1.868 (clean molars) + 0.458 (sugary foods) + 0.499 (drinks)
- 0.857 (medications) + 3.004 (eating disorder) + 0.496 (acidity)
+ 1.162 (stomach problem) + 0.863 (dry mouth) 0.418 (dental plaque).

It can be seen from (Table 2) and significance is calculated based on critical region that $p<0.05$.

Table 2
Logistic Regression

	B	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**95.0% C.I. for EXP(B)**	
							Lower	**Upper**
Genetics	.309	.157	3.873	1	.049	1.363	1.001	1.854
Grinding teeth	.529	.190	7.760	1	.005	1.696	1.170	2.461
Bristles	.815	.163	25.092	1	.000	2.259	1.642	3.106
Brushing after eating	-.470	.163	8.297	1	.004	.625	.454	.860
Flossing	-.889	.169	27.613	1	.000	.411	.295	.573
Mouth wash	-.750	.182	17.026	1	.000	.472	.331	.675
Florided toothpaste	-.500	.206	5.900	1	.015	.606	.405	.908
Clean molars	-1.868	.160	135.871	1	.000	.154	.113	.211
Sugary foods	.458	.162	8.038	1	.005	1.581	1.152	2.170
Drinks	.499	.162	9.479	1	.002	1.647	1.199	2.263
Medications	-.857	.297	8.301	1	.004	.425	.237	.760
Eating disorder	3.004	1.427	4.430	1	.035	20.169	1.230	330.804
Acidity	.496	.196	6.396	1	.011	1.643	1.118	2.414
Stomach problem	1.162	.292	15.817	1	.000	3.196	1.803	5.666
Dry mouth	.863	.209	17.058	1	.000	2.369	1.573	3.568
Dental plaque	.418	.156	7.213	1	.007	1.520	1.120	2.062
Constant	.030	.301	.010	1	.920	1.031		

As shown in the above table, the risk factors genetics, grinding teeth, bristles, brushing after eating, flossing, mouth wash, toothpaste, cleaning of molars, sugary food, drinks, medications, eating disorder, acidity, stomach problem, dry mouth, dental plaque were found to significant against tooth decay for adolescents as their p-value were less than 0.05.

Since the coefficient of genetics or family history of tooth decay was found to be 0.309 indicating direct relationship between genetics and tooth decay. Odds ratio was found to be 1.363, indicating that student with family history of tooth decay were 1.363 times more chance of having tooth decay than those who did not have any family history of tooth decay.

Odds ratio for grinding teeth was found to be 1.696 representing that students who had the habit of grinding teeth were 1.696 times high chance of tooth decay as compared to those who did not have the habit of grinding teeth. Odds ratio for bristles of brush was found to be 2.259 indicating that students who uses rough bristle brush were 2.259 times higher risk of having tooth decay than those who use soft bristle brush. The odds ratio for brushing after eating was found to be 0.625 which means that the students who had the habit of brushing after eating were 1 - 0.625 = 0.485 = 48.5% protected against tooth decay.

As odds ratio for flossing was found to be 0.411, indicating that there was low chance of not having tooth decay in students who floss as compared to those students who did not floss. The students who floss had (1 – 0.411) = 0.699 = 69.9% protection against tooth decay. As 95% confidence interval did not include 1 (0.534, 0.897) indicating that flossing had statistically significantly negative association with tooth decay. Odds ratio for using mouth wash was found to be 0.472, indicating that the students who use mouth wash were at lower risk of having tooth decay as compared to those students who did not use mouth wash. The students who use mouth wash was 1 – 0.472 = 0.638 = 63.8% protected against tooth decay.

The odds ratio for toothpaste enriched with fluoride was found to be 0.606 representing that the students who use toothpaste with fluoride had a low risk of having tooth decay as compared to those who did not use tooth paste with fluoride. The students who use flourided toothpaste were 1 - 0.606 = 0.394 = 39.4% protected against tooth decay. The odds ratio for cleaning of molars was found to be 0.154, indicating that the students who did not clean their molars from inside had a high chance of having tooth decay as compared to those who clean their molars. The students who clean their molars from inside were 1 – 0.154 = 0.945 = 94.5% protected against tooth decay.

The odds ratio for sugary foods was found to be 1.581, indicating that there were increased risk of 1.581 times of tooth decay among those who frequently consumed sugary food as compared to those who did not frequently consume sugary food. The odds ratio for usage of beverages was found to be 1.647 indicating that there was increased risk of 1.647 times of tooth decay for those who used soft drinks or beverages frequently as compared to those who did not use drinks or beverages frequently.

The odds ratio for medication was found to be 1.425 indicating that there were 1.425 times higher chance of tooth decay among those who take medications than those students who did not take any medications. The odds ratio for eating disorder was found to be 2.063, indicating that there were 2.063 times higher chance of tooth decay among those students who had eating disorder than those who were not suffering from eating disorder. The odds ratio of acidity was found to be 1.643 indicating that there were 1.643 times higher chance of tooth decay among those who had the problem of acidity than those who were not suffering from the problem of acidity.

The odds ratio of stomach was found to be 3.196, indicating that there were 3.196 times higher chance of tooth decay among those who had the stomach problem than those who were not suffering from stomach problem. The odds ratio of plaque was found to be 2.369 which indicate that there were 1.693 times higher chance of tooth decay in students who had dental plaque. The odds ratio of dry mouth was found to be 1.520 indicating that there were 1.520 times high chance of tooth decay for dry mouth as compared to those who did not have dry mouth.

4. COMMENTS AND CONCLUSION

This study concluded that brushing after eating, flossing, mouth wash, molar cleaning habits of respondent were found to be negatively associated with tooth decay. Family tooth decay (genetics) and rough bristles of brush were found to be significantly positively associated with tooth decay. Students who had tooth decay in genes were at risk of developing tooth decay. Medications, acidity, stomach problem, dental plaque and dry mouth were also found to be positively associated with tooth decay.

Socio-demographic factors like age, smoking and eating habits like junk food, carbohydrate enriched food, soft drinks or beverages, were found to be positively associated with tooth decay which indicated that high consumption of these foods or drinks would result in high risk of tooth decay.

Binary logistic model was fitted to determine the predictive strength of different risk factors. This model was also fitted so that tooth decay could be predicted based on presence or absence of various risk factors. Genetics, grinding teeth, bristles of brush, brushing after eating, flossing, mouth wash, cleaning of molars, toothpaste, sugary foods, medications, acidity, eating disorder, stomach problem, dental plaque, dry mouth were included in the overall model as significant factors.

Traditionally in Pakistan, food enriched with carbohydrates and junk food are prepared and consumed in daily or frequently which is one of the major cause of tooth decay among adolescents. Regular or frequent uses of soft drinks or beverages in youngsters are greatly harmful and have a risk of causing tooth decay. Tooth decay can be prevented by adopting some healthy oral habits like brushing after eating, flossing, and using mouth wash. Further young generation tends to consume sugary food frequently which results in tooth decay. It can be seen from this study that visit to dentist is significantly associated with tooth decay, so people should pay regular visits to dentist. People should use toothpaste that is enriched in fluoride so that it could protect your enamel and tooth from decay. Also, people should clean their teeth particularly from inside where food sticks mostly.

People should increase their low-fat milk consumption so that their teeth have high prevention power from decay. Milk that is high in fat or full of cream have a high chance of suffering from tooth decay. People having dry mouth should increase their water or liquid consumption so that food or sticky particles from drinks could be erased from mouth and should brush their teeth after eating such things. People should brush their teeth properly so that plaque could be reduced, and proper measures should be taken after consulting doctor so that tooth decay could be avoided.

5. ACKNOWLEDGEMENT

My sincere and heartiest thanks to my parents whose continuous prayers, encouragement and support have enable me to achieve my goal in my research study. I would also like to thanks my siblings for their support. I felt thankful from bottom of my heart to Ms. Mahnaz Makhdum, Chief Executive of Statistics Department, who was my supervisor as well, who helped me out in every aspect of my research work with her inspiring supervision. I deem it as privilege to work under her guidance. I will always be grateful to her.

REFERENCES

1. Masson, L.F., Blackburn, A., Sheehy, C., Craig, L.C., Macdiarmid, J.I., Holmes, B.A. and McNeill, G. (2010). Sugar intake and dental decay: results from a national survey of children in Scotland. *British Journal of Nutrition*, 104(10), 1555-1564.
2. Mulu, W., Demilie, T., Yimer, M., Meshesha, K. and Abera, B. (2014). Dental caries and associated factors among primary school children in Bahir Dar city: a cross-sectional study. *BMC Research Notes*, 7(1), 949.
3. Khan, A.A., Jain, S.K. and Shrivastav, A. (2008). Prevalence of dental caries among the population of Gwalior (India) in relation of different associated factors. *European Journal of Dentistry*, 2, 81-85.
4. Gupta, A., Smithers, L.G., Harford, J., Merlin, T. and Braunack-Mayer, A. (2018). Determinants of knowledge and attitudes about sugar and the association of knowledge and attitudes with sugar intake among adults: A systematic review. *Appetite*, 126, 185-194.
5. Jette, A.M., Feldman, H.A. and Tennstedt, S.L. (1993). Tobacco use: a modifiable risk factor for dental disease among the elderly. American Journal of Public Health, 83(9), 1271-1276.
6. Al-Darwish, M., El Ansari, W. and Bener, A. (2014). Prevalence of dental caries among 12–14 year old children in Qatar. *The Saudi Dental Journal*, 26(3), 115-125.
7. Pohjola, V., Rekola, A., Kunttu, K. and Virtanen, J.I. (2016). Association between dental fear and oral health habits and treatment need among University students in Finland: a national study. *BMC Oral Health*, 16(1), 26.
8. Xu, W., Lu, H.X., Li, C.R. and Zeng, X.L. (2014). Dental caries status and risk indicators of dental caries among middle-aged adults in Shanghai, China. *Journal of Dental Sciences*, 9(2), 151-157.
9. Žemaitienė, M., Grigalauskienė, R., Vasiliauskienė, I., Saldūnaitė, K., Razmienė, J. and Slabšinskienė, E. (2016). Prevalence and severity of dental caries among 18-year-old Lithuanian adolescents. *Medicina*, 52(1), 54-60.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 175-182

INVESTIGATING LIFE SATISFACTION OF MICROCEPHALIC PERSONS

Kalsoom Akhtar Chaudhry[1], Tooba Khan[2] and **Kanwal Aslam[3]**
Kinnaird College for Women, Lahore, Pakistan
Email: [1]kalsoom.akhtar@kinnaird.edu.pk
[2]tooba.khan@kinnaird.edu.pk
[3]kanwalaslam20@gmail.com

ABSTRACT

Illness, particularly in the developing world, is considered a social burden when it takes a long time to recover. Microcephaly is a rare neurodevelopmental disorder in which the persons suffering from it have smaller heads compared to normal persons and it may be congenital or occur in early growing years of a child. People suffering from such disorder are often mistreated in Pakistan and they face numerous problems due to small heads and abnormalities. A cross-sectional study of Lahore city was carried out which aimed at examining life satisfaction of microcephalic persons. A total of 67 respondents comprised of children, teenagers and the elderly from different socio economic status. Life satisfaction turned out to be significantly associated with microcephalic family members, inherited microcephaly, friends and medical checkup. The results of the binary logistic regression revealed that hereditary factors, medical checkup and having friends were significant predictors of life satisfaction. Microcephalic persons must be paid extra attention as they are dissatisfied with their life due to stigmatisation and social exclusion.

KEYWORDS

Microcephaly, MPCH, Head circumference.

1. INTRODUCTION

Microcephaly is a neurodevelopmental disorder and babies born with this disorder have reduced head circumference compared to normal infants (more than 2-3 standard deviations below the mean) having the same gestation, age and gender. The characteristic features of this disorder are small brain, sloping head, delay in speech, difficulty in learning, spasticity, mental retardation, seizers and epilepsy, and thus the victims have delayed skills regarding motion, such as standing, sitting, and walking (Qazi & Reed, 1975). As age advances, a microcephalic baby grows and the body grows in size except for the size of head. Microcephaly can also be caused by maternal malnutrition, maternal alcoholism, and mercury poisoning and chromosomal aberration (Faheem et al., 2015). Microcephalic persons are physically much weaker and remain underweight than normal children of the same age and gender.

Microcephaly is categorized into two types: primary microcephaly and secondary microcephaly. Primary microcephaly is present at the time of birth and is non-syndromic in nature, however, in secondary microcephaly the individual is born with normal head circumference but later his head does not grow like normal individuals (Hussain et al., 2013).

The trend of consanguineous marriages is highlighted in the Middle East, West Asia and North America where the high rate of such marriages depends upon different factors including socioeconomic status, local tradition, religion and education level of the people in these countries. Consanguineous marriages favoured the inheritance of genetic disorders as autosomal recessive in carriers and they caused congenital aberration (CA) in their offspring. World Health Organization suggested that about 3% of neonates are born with CA that is equal to 3 million fetuses and infants per year. About 70% of infants died in their first month of birth due to CA and it was also seen that 30-40% of CA are genetic (Naibkhil & Chitkara, 2016).

As microcephaly is very rare autosomal recessive genetic disorder, its incidence is varies from one population to another. Its incidence rate is 1 in 250,000 in Holland, 1 in 200,000 in Scotland, 1 in 30,000 in Japan (Faheem et al., 2015). It is also estimated that the incidence rate of microcephaly is 1.3-150 in 100,000 which depends on the consanguinity level in a specific population (Hanzlik & Gigante, 2017). Microcephaly is more common in Arab and Asian families than the European families due to consanguinity (Abuelo, 2007). Consanguineous marriages are supported by society and culture in Arab, Asia, North Africa and South India. Consanguineous or cousin marriages are commonly practiced in several countries, such as Pakistan, Saudi Arabia, Iraq, Turkey, North Africa and Australia. Consanguinity increases the risk of autosomal recessive conditions (Hamamy, Masri, Al-Hadidy, & Ajlouni, 2007). In Pakistan, the rate of consanguinity has been studied to be as high as 70% and thus the incidence of microcephaly is 1 per 100,000 in Pakistan (Hussain et al., 2013).

As it is not the skull disorder, it is the brain disorder instead, but there are several myths concerning this disorder in Pakistan and the victims are commonly regarded as “rat people” or “chuhas”. According to one of those myths, microcephalic persons got kidnapped in their childhood and the masters deformed their skulls by placing iron pots or clamps, hence causing growth retardation of their brains and heads. (Irshad, Anwar, & Shehzad; Miles, 1996).

Though microcephaly is inherited in an autosomal recessive pattern but one gene WDFY3 (MCPH18) shows autosomal dominant inheritance. The chance to transmit the disease for this locus is 50% in every pregnancy whereas the chance is reduced to half as 25% in the case of rest of the gene.

Scientists have identified 24 genetic loci so far which are MCPH1-24 with associated genes, such as MCPH1, WDR62, CDK5RAP2, CASC5 (CASC6), ASPM, CENPJ, STIL, CEP135, CEP152, ZNF335, PHC1, CDK4, CENPE, SASS6, MFSD2A, ANKLE, CIT,

WDFY, COPB2, KIF14, NCAPD2, NCAPD3, NCAPH and NUP37 (Braun et al., 2018; De Baere, Speleman, Van Roy, De Paepe, & Messiaen, 1998; Gruneberg et al., 2006; Khan et al., 2017; Loïodice et al., 2004; Martin et al., 2016; Simonsen et al., 2004).

Microcephaly is not a curable disease. As autosomal primary microcephaly is a neurodevelopment genetic disorder and is developed before birth, so its treatment is not available. Secondary microcephaly can be prevented in some cases but no specific interventions are available for primary microcephaly. Specific motor training programs can be promoted to train the persons affected with this disorder. Early intervention programs, occupational and developmental physical therapies as well as genetic counseling might be helpful to prevent the incidence of this disorder (Harris, 2015).

2. DATA AND METHODOLOGY

A close ended well-structured questionnaire was used for this cross sectional study. Data were collected from 67 microcephalic persons in the city of Lahore and its surrounding areas, such as shrine Daata Darbar, Pattoki, Bauwala, Walton, Gulberg and Kasur. The chi-square test, Fisher's exact test and binary logistic regression analysis were carried out.

3. RESULTS

Table 1
Percent Distribution of Lifestyle Factors

Variable	Categories	Percent
Lifestyle Satisfaction	No	40.3
	Yes	59.7
Gender	Female	34.3
	Male	65.7
Age	Less than 5 years	13.4
	5-10 years	34.3
	10-15 years	20.9
	15-20 years	14.9
	Over 20	16.4
Socioeconomic Status	Low	80.6
	Medium	17.9
	High	1.5

Variable	Categories	Percent
Qualification	Uneducated	77.6
	Primary	14.9
	Middle	6.0
	Intermediate	1.5
Family System	Nuclear	47.8
	Joint	40.3
	Others	11.9
Inherited Microcephaly	No	50.7
	Yes	49.3
Number of Microcephalic Family members	0	50.7
	1	25.4
	2	17.9
	3	6.0
Formal Schooling	No	80.6
	Yes	19.4
Pocket Money	No	62.7
	Yes	37.3
Getting Gifts	No	50.7
	Yes	49.3
Friends	No	62.7
	Yes	37.3
Outing	No	35.9
	Yes	64.1
Hobby	Sports	28.4
	Playing with dolls or friends	25.4
	Stitching / knitting	3.0
	Drawing / painting	4.5
	Others	38.8

Variable	Categories	Percent
Favorite Task	Working at home	7.5
	Spending time with family	4.5
	Sleeping	10.4
	Playing	44.8
	Worship	14.9
	Others	17.9
Watching Television	No	56.7
	Yes	43.3
Mobile Usage	No	73.1
	Yes	26.9
Difficulty Interacting with Strangers	No	38.8
	Yes	61.2
Family Income	Below 15000	58.2
	15000-25000	29.9
	25000-35000	7.5
	Over 35000	4.5
Medical Checkup	Never	38.8
	Sometimes	40.3
	Yearly	7.5
	Monthly	13.4

This research explored that 65.7% of the microcephalic persons were males and 34.3% were females out of the sample of 67 microcephalic persons. Of the total microcephalic persons, 80.6% belonged to low socioeconomic status and 77.6% were uneducated whereas 50.7% microcephalic persons inherited this disorder. Almost 51% of the microcephalic persons have not ever got gifts and 62.7% were not getting pocket money. Approximately 40% of the microcephalic persons were dissatisfied with their life and 62.7% reported that they had no friends. The hobbies of 38.8% of microcephalic persons were reading, singing and watching television whereas 4.5% preferred to spend their time with their families.

Table 2
Bivariate Analysis

Variable	P-Value
Microcephalic Family Members	0.001
Inherited Microcephaly	0.046
Having Friends	0.000
Medical Checkup	0.011

The results of the Pearson's chi square test and Fisher's exact test presented evidence that life satisfaction has a significant association with microcephalic family members, inherited microcephaly, having friends and medical checkup.

Table 3
Classification Table

Observed		Predicted		
		Life Satisfaction		Percentage Correct
		Un-satisfied	Satisfied	
Life Satisfaction	Un-satisfied	23	4	85.2
	Satisfied	4	36	90.0
Overall Percentage				88.1

A logistic regression model was developed to observe the factors that were significant in explaining life satisfaction of microcephalic persons. According to the results, 88.1% of the total cases were correctly classified by the model.

Table 4
Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Microcephalic Family Members	1.609	.610	6.971	1	.008	5.000
Inherited Microcephaly	2.707	.909	8.867	1	.003	14.983
Having Friends	3.660	1.159	9.970	1	.002	38.867
Medical Checkup			8.467	3	.037	
Medical Checkup (1)	-.090	.974	.009	1	.926	.914
Medical Checkup (2)	-3.832	1.685	5.171	1	.023	.022
Medical Checkup (3)	-2.332	1.206	3.743	1	.053	.097
Constant	-2.126	.776	7.510	1	.006	.119

Genetic factors, such as inherited microcephaly and number of microcephalic family members, and having friends turned out to be significant predictors of life satisfaction whereas the rest of the related factors were insignificant.

With one more microcephalic family member, the estimated odds of life satisfaction for microcephalic persons increase by a factor of 5. The estimated odds of being satisfied with life for those persons who inherited microcepahly equal 14.983 times the estimated odds for those who have secondary microcephaly. The likelihood of being satisfied with life for those microcephalic persons who have friends is 38.867 times the likelihood for those who do not have friends. Microcephalic persons who have an annual medical checkup are 0.023 times as likely to be satisfied with their life as those who do not go for a medical checkup annually.

4. COMMENTS AND CONCLUSION

This study aimed at investigating life satisfaction of microcephalic persons in the capital of Punjab province, Pakistan. The total number of microcephalic persons being investigated was 67 and about one-third were females. The majority belonged to low socioeconomic status and was uneducated. More than half the respondents had no other microcephalic members in their families and a vast majority had not been visiting doctors regularly. Nearly two-fifths were not satisfied with their life whereas those respondents who reported satisfaction with their life preferred to spend their time with their friends. Genetic factors and medical checkups also played a significant role.

No researches have thus far been carried out on assessing life satisfaction of persons having primary or secondary microcephaly in Pakistan. Researchers may attempt to conduct studies on lifestyle of persons having this disorder in other areas so that numerous other factors can also be taken into account.

Microcephaly is a rare neurodevelopmental disorder causing its victims to feel discriminated against and stimatised in the society. This research put forth the fact that female respondents were less satisfied with their life compared to males. Microcephalic persons who spent time with friends were satisfied with their life compared to others. Having investigated the lifestyle of microcephalic persons and their problems, this is imperative to impart awareness among the masses concerning the difficulties faced by microcephalic persons since they are part of our society and therefore they must be treated with extra care and affection.

REFERENCES

1. Abuelo, D. (2007). Microcephaly Syndromes. Paper presented at the *Seminars in pediatric neurology*. 14. 118-27. 10.1016/j.spen.2007.07.003.
2. Braun, D.A., Lovric, S., Schapiro, D., Schneider, R., Marquez, J., Asif, M. and Rao, J. (2018). Mutations in multiple components of the nuclear pore complex cause nephrotic syndrome. *Journal of Clinical Investigation,* 128(10), 4313-4328.
3. De Baere, E., Speleman, F., Van Roy, N., De Paepe, A. and Messiaen, L. (1998). Assignment1 of the cellular retinol-binding protein 1 gene (RBP1) and of the

coatomer betasubunit gene (COPB2) to human chromosome band 3q23 by in situ hybridization. *Cytogenetic and Genome Research*, 82(3-4), 226-227.

4. Faheem, M., Naseer, M.I., Rasool, M., Chaudhary, A.G., Kumosani, T.A., Ilyas, A.M. and Al-Qahtani, M.H. (2015). Molecular genetics of human primary microcephaly: an overview. *BMC Medical Genomics,* 8(1), S4.
5. Gruneberg, U., Neef, R., Li, X., Chan, E.H., Chalamalasetty, R.B., Nigg, E.A. and Barr, F.A. (2006). KIF14 and citron kinase act together to promote efficient cytokinesis. *The Journal of Cell Biology*, 172(3), 363-372.
6. Hamamy, H.A., Masri, A.T., Al-Hadidy, A.M. and Ajlouni, K.M. (2007). Consanguinity and genetic disorders. Profile from Jordan. *Saudi Medical Journal*, 28(7), 1015-1017.
7. Hanzlik, E. and Gigante, J. (2017). Microcephaly. *Children*, 4(6), 47. doi: 10.3390/children4060047
8. Harris, S.R. (2015). Measuring head circumference: Update on infant microcephaly. *Canadian Family Physician*, 61(8), 680-684.
9. Hussain, M.S., Bakhtiar, S.M., Farooq, M. Anjum, I., Janzen, E., Toliat, M.R. and Noegel, A. (2013). Genetic heterogeneity in Pakistani microcephaly families. *Clinical Genetics*, 83(5), 446-451.
10. Irshad, S., Anwar, S. and Shehzad, B. (2011). Linkage Study of Primary Microcephaly in Pakistani Kindred. *Research Journal of Biology*, 01(01), 1-6.
11. Khan, M.A., Windpassinger, C., Ali, M.Z., Zubair, M., Gul, H., Abbas, S. and Nawaz, Z. (2017). Molecular genetic analysis of consanguineous families with primary microcephaly identified pathogenic variants in the ASPM gene. *Journal of Genetics*, 96(2), 383-387.
12. Loïodice, I., Alves, A., Rabut, G., van Overbeek, M., Ellenberg, J., Sibarita, J.-B. and Doye, V. (2004). The entire Nup107-160 complex, including three new members, is targeted as one entity to kinetochores in mitosis. *Molecular Biology of the Cell*, 15(7), 3333-3344.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 183-198

A GENERALIZED REGRESSION-PRODUCT ESTIMATOR IN PRESENCE OF NON-RESPONSE

Muhammad Zubair[1] and **Asad Ali**[2§]
[1] Department of Statistics, University of Sargodha, Sargodha, Pakistan.
[2] Department of Quantitative Methods, SBE, University of Management and Technology, Lahore, Pakistan
[§] Corresponding Author Email: ranaasadali23@gmail.com

ABSTRACT

In this paper, we propose a generalized regression-product estimator for population mean, taking two auxiliary variables into consideration using two phase sampling scheme with sub sampling technique in presence of non-response. The two situations with known population means of auxiliary variables, incomplete information on study variable and incomplete/complete information on auxiliary variables have been considered. Expressions for bias and mean square error have been derived. The proposed estimator has been compared with usual unbiased estimator and some existing estimators. An empirical study has also been carried out.

KEY WORDS

Non-Response, regression-product Estimator, Two Phase Sampling, Sub Sampling Technique.

1. INTRODUCTION

Survey statisticians have to face problem of non-response in surveys related to human beings. Failure in collecting information from any number of sampling units on study or auxiliary variable(s) or both is known as non-response. The non-response reduces effective sample size which causes increase in variance, bias etc. Hansen and Hurwitz (1946) introduced sub-sampling technique to overcome the problem. Several estimators for population parameters utilizing this technique have been suggested by survey statisticians like Cochran (1977), Rao (1986, 1990), Khare and Srivastava (1993, 1995, 1997, 2010), Sodipo and Obisesan (2007), Singh and Kumar (2008, 2009, 2011), Sing et al. (2010), Zubair (2012), Ismail et al. (2013), Pal and Singh (2017) and Ismail and Shahbaz (2017).

Suppose a population of size 'N' is considered to draw a simple random sample of 'n' units without replacement. From these 'n' units, 'r_1' units respond and 'r_2' are non-respondents. The population is divided in respondent and non-respondent groups consisting of 'N_1' and 'N_2' units corresponding to sample respondents and non-respondents. A sub-sample of size 'k' $\left(k=\frac{r_2}{h}, h>1\right)$ is drawn from 'r_2' non-respondents units to collect information about study variable from these 'k' units.

Hansen and Hurwitz (1946), proposed following unbiased estimator of population mean by introducing sub-sampling technique to deal with problem of non-response:

$$t_1 = \bar{y}^* = \left(\frac{r_1}{n}\right)\bar{y}_{r_1 1} + \left(\frac{r_2}{n}\right)\bar{y}_{k2}, \tag{1.1}$$

where $\bar{y}_{r_1 1} = \frac{\sum_{i=1}^{r_1} y_i}{r_1}$ is mean of responding part of sample and $\bar{y}_{k2} = \frac{\sum_{i=1}^{k} y_i}{k}$ is mean of sub sample taken from non-responding part of sample. The estimator (1.1) is unbiased with variance:

$$Var(t_1) = \lambda_2 S_y^2 + \theta S_{y(2)}^2 = V_y^*, \tag{1.2}$$

where

$$S_y^2 = \frac{\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^2}{N-1},\ S_{y(2)}^2 = \frac{\sum_{i=1}^{N_2}\left(y_i - \bar{Y}_2\right)^2}{N_2 - 1},\ \bar{Y} = \frac{\sum_{i=1}^{N_1} Y_i}{N},\ \bar{Y}_2 = \frac{\sum_{i=1}^{N_2} Y_i}{N_2},\ \lambda_2 = \frac{1}{n_2} - \frac{1}{N},$$

$$\theta = \frac{W_2(h-1)}{n_2},\quad W_2 = \frac{N_2}{N}.$$

2. TWO-PHASE SAMPLING PROCEDURE IN THE PRESENCE OF NON-RESPONSE

The two-phase sampling procedure in presence of non-response using sub-sampling technique is as follows:-

In first phase, a large sample of size 'n_1' units is drawn using simple random sampling without replacement to collect information on auxiliary variable(s).

In second phase, from 'n_1' first phase sample units, another sample of size 'n_2' using simple random sampling without replacement is drawn. From these n_2 units, r_1 units respond and r_2 units do not. Information on study variable 'y' is collected from these responding units. Then a sub sample of size k $\left(k = \frac{r_2}{h}, h > 1\right)$ is selected from r_2 non-responding units using simple random sampling without replacement to collect information on study variable.

Several Statisticians have proposed estimators of population mean using two-phase sampling scheme with sub-sampling techniques taking auxiliary variable(s) into consideration in presence of non-response. Some of them have suggested ratio, product and regression estimators under following two situations.

Situation I: Known mean(s) of auxiliary variable(s) and incomplete information on both study and auxiliary variable(s).

Situation II: Known mean(s) of auxiliary variable(s) and incomplete information on study variable but complete information on auxiliary variable(s).

Some estimators of population mean for two phase sampling scheme in presence of non-response under situation I and II are presented below. Following notations are used to express the mean square error.

$$V_y^* = \lambda_2 S_y^2 + \theta S_{y(2)}^2, \quad V_x^* = \lambda_2 S_x^2 + \theta S_{x(2)}^2, \quad V_z^* = \lambda_2 S_z^2 + \theta S_{z(2)}^2, \quad V_{xy}^* = \lambda_2 S_{xy} + \theta S_{xy(2)},$$
$$V_{yz}^* = \lambda_2 S_{yz} + \theta S_{yz(2)}, \quad V_{xz}^* = \lambda_2 S_{xz} + \theta S_{xz(2)}, \quad V_x = \lambda_2 S_x^2, \quad V_z = \lambda_2 S_z^2,$$
$$V_{xy} = \lambda_2 S_{xy}, \quad V_{yz} = \lambda_2 S_{yz}, \quad V_{xz} = \lambda_2 S_{xz} \quad \lambda_2 = 1/n_2 - 1/N.$$

Cochran (1977) proposed following estimators of population mean for situation I.

$$t_2 = \left(\frac{\bar{y}^*}{\bar{x}^*}\right)\bar{X}, \tag{2.1}$$

$$t_3 = \bar{y}^*\left(\frac{\bar{x}^*}{\bar{X}}\right), \quad (2.2)\; t_4 = \bar{y}^* + \hat{\beta}_{yx}^*\left(\bar{X} - \bar{x}^*\right), \tag{2.3}$$

where $\hat{\beta}_{yx}^* = s_{xy}^* / s_x^{*2}$ and s_{xy}^* & s_x^{*2} are estimates, calculated from available information.

Expressions for mean square error upto first order approximation are given respectively.

$$MSE(t_2) \approx V_y^* + R_1^2 V_x^* - 2R_1 V_{xy}^*, \tag{2.4}$$

$$MSE(t_3) \approx V_y^* + R_1^2 V_x^* + 2R_1 V_{xy}^*, \tag{2.5}$$

$$MSE(t_4) \approx V_y^* + \beta_{yx}^2 V_x^* - 2\beta_{yx} V_{xy}^*, \tag{2.6}$$

where $R_1 = \bar{Y} / \bar{X}$.

Rao (1986) proposed following estimators of population mean for situation II.

$$t_5 = \bar{y}^* \frac{\bar{X}}{\bar{x}}, \tag{2.7}$$

$$t_6 = \bar{y}^* \frac{\bar{x}}{\bar{X}}, \tag{2.8}$$

$$t_7 = \bar{y}^* + \hat{\beta}_{yx}\left(\bar{X} - \bar{x}\right), \tag{2.9}$$

where $\hat{\beta}_{yx}$ is an estimate for coefficient of regression y on x.

Expressions for mean square error are given respectively.

$$MSE\left(t_5\right) \approx V_y^* + R_1^2 V_x - 2R_1 V_{xy} \tag{2.10}$$

$$MSE\left(t_6\right) \approx V_y^* + R_1^2 V_x + 2R_1 V_{xy} \tag{2.11}$$

$$MSE\left(t_7\right) \approx V_y^* + \beta_{yx}^2 V_x - 2\beta_{yx} V_{xy} \tag{2.12}$$

Khare and Srivastava (1997) proposed following transformed ratio type estimators.

$$t_8 = \bar{y}^* \frac{\bar{X} + L}{\bar{x}^* + L}, \tag{2.13}$$

$$t_9 = \frac{\bar{y}^* (\bar{X} + L_1)}{(\bar{x} + L_1)}, \tag{2.14}$$

where L and L_1 are positive constants.

Expressions for mean square error with optimum values of constants are given below.

$$MSE\left(t_8\right)_{opt.} \approx V_y^* - \frac{V_{xy}^{*2}}{V_x^*}. \tag{2.15}$$

$$MSE\left(t_9\right)_{opt.} \approx V_y^* - \frac{V_{xy}^2}{V_x}. \tag{2.16}$$

Sodipo and Obisesan (2007) proposed difference cum ratio estimator for situation II.

$$t_{10} = \left(\bar{y}^* - \alpha\left(\bar{x} - \bar{X}\right)\right)\left(\frac{\bar{X}}{\bar{x}}\right), \tag{2.17}$$

where α is suitable constant.

Expression for mean square error with optimum value of α is given below.

$$MSE\left(t_{10}\right)_{opt.} \approx V_y^* + \beta_{yx}^2 V_x - 2\beta_{yx} V_{xy}. \tag{2.18}$$

Singh and Kumar (2008) proposed following ratio type estimator for situation I.

$$t_{11} = \bar{y}^* \left(\frac{\bar{X}}{\bar{x}^*}\right)^{\alpha_1} \left(\frac{\bar{X}}{\bar{x}}\right)^{\alpha_2}, \tag{2.19}$$

where $\bar{x}^*$ and $\bar{x}$ are unbiased estimators of population mean for auxiliary variable and α_1 & α_2 are constants.

Expression for mean square error with optimum values of constants is given below.

$$MSE\left(t_{11}\right)_{opt.} \approx V_y^* - \left(\frac{V_{xy}^2}{V_x} + \frac{\left(V_{xy}^* - V_{xy}\right)^2}{V_x^* - V_x} \right). \tag{2.20}$$

Singh and Kumar (2011) proposed following combined regression-cum-ratio estimators using two auxiliary variables.

$$t_{12} = \left(\bar{y}^* + \hat{\beta}_{yx}^* \left(\bar{X} - \bar{x}^*\right)\right)\left(\frac{\bar{Z}}{\bar{Z} + \alpha\left(\bar{z}^* - \bar{Z}\right)} \right), \tag{2.21}$$

$$t_{13} = \left(\bar{y}^* + \hat{\beta}_{yx} \left(\bar{X} - \bar{x}\right)\right)\left(\frac{\bar{Z}}{\bar{Z} + \alpha^*\left(\bar{z} - \bar{Z}\right)} \right), \tag{2.22}$$

where $\hat{\beta}_{yx}^* = s_{xy}^* / s_x^{*2}$, $\hat{\beta}_{yx} = s_{yx} / s_x^2$, and α and $\alpha*$ are constants.

Expressions for mean square error using optimum values of constants are given as under.

$$MSE\left(t_{12}\right)_{opt.} \approx MSE\left(t_4\right) - N^{*2} / D^*, \tag{2.23}$$

$$MSE\left(t_{13}\right)_{opt.} \approx MSE\left(t_7\right) - A^2 V_z, \tag{2.24}$$

where $N^* = \lambda_2 A S_z^2 + \theta B S_{z(2)}^2$, $D^* = V_z^*$, $A = \beta_{yz} - \beta_{yx}\beta_{xz}$ and $B = \beta_{yz(2)} - \beta_{yx}\beta_{xz(2)}$.

Zubair (2012) proposed following regression-cum-ratio estimators with two auxiliary variables.

$$t_{14} = \alpha\left(\bar{y}^* + \hat{\beta}_{yx}\left(\bar{X} - \bar{x}^*\right)\right)\frac{\bar{Z}}{\bar{z}^*} + \left(1-\alpha\right)\left(\bar{y}^* + \hat{\beta}_{yz}\left(\bar{Z} - \bar{z}^*\right)\right)\frac{\bar{X}}{\bar{x}^*}, \tag{2.25}$$

$$t_{15} = \alpha_1\left(\bar{y}^* + \hat{\beta}_{yx}\left(\bar{X} - \bar{x}\right)\right)\frac{\bar{Z}}{\bar{z}} + \left(1-\alpha_1\right)\left(\bar{y}^* + \hat{\beta}_{yz}\left(\bar{Z} - \bar{z}\right)\right)\frac{\bar{X}}{\bar{x}}, \tag{2.26}$$

where α and α_1 are constants.

Expressions for mean square error using optimum values of constants are given as under.

$$MSE\left(t_{14}\right)_{opt.} \approx MSE\left(t_2\right) + \beta_{yz}\left(\beta_{yz}V_z^* - 2\left(V_{yz}^* - R_1 V_{xz}^*\right)\right) - \frac{A^{*2}}{B^*}, \tag{2.27}$$

where

$$A^* = \left(R_1 - \beta_{yx}\right)\left(R_1 V_x^* - V_{xy}^* + \beta_{yz} V_{xz}^*\right) - \left(R_2 - \beta_{yz}\right)\left(\beta_{yz} V_z^* - V_{yz}^* + R_1 V_{xz}^*\right),$$

$$B^* = \left(R_1 - \beta_{yx}\right)^2 V_x^* + \left(R_2 - \beta_{yz}\right)^2 V_z^* - 2\left(R_1 - \beta_{yx}\right)\left(R_2 - \beta_{yz}\right) V_{xz}^*, \text{ and } R_2 = \bar{Y} / \bar{Z}.$$

$$MSE\left(t_{15}\right)_{opt.} \approx MSE\left(t_5\right) + \beta_{yz}\left(\beta_{yz} V_z - 2\left(V_{yz} - R_1 \beta_{yz} V_{xz}\right)\right) - \frac{A^2}{B} \tag{2.28}$$

where

$$A = \left(R_1 - \beta_{yx}\right)\left(R_1 V_x - V_{xy} + \beta_{yz} V_{xz}\right) - \left(R_2 - \beta_{yz}\right)\left(\beta_{yz} V_z - V_{yz} + R_1 V_{xz}\right), \text{ and}$$

$$B = \left(R_1 - \beta_{yx}\right)^2 V_x + \left(R_2 - \beta_{yz}\right)^2 V_z - 2\left(R_1 - \beta_{yx}\right)\left(R_2 - \beta_{yz}\right) V_{xz}.$$

Ismail and Shahbaz (2017) proposed following combined ratio and regression estimator with two auxiliary variables.

$$t_{16} = \left(\bar{y}^* + \hat{\beta}_{yx}^*\left(\bar{X} - \bar{x}^*\right)\right)\left(\frac{\bar{z}^*}{\bar{Z}}\right)^{\alpha_d}, \tag{2.29}$$

$$t_{17} = \left(\bar{y}^* + \hat{\beta}_{yx}\left(\bar{X} - \bar{x}\right)\right)\left(\frac{\bar{z}}{\bar{Z}}\right)^{\alpha_d^*}, \tag{2.30}$$

where α_d and α_d^* are constants.

Expressions for mean square error using optimum values of constants are given as under.

$$MSE\left(t_{16}\right)_{opt.} \approx MSE\left(t_4\right) - \frac{V_{yz}^* - \beta_{yx} V_{xz}^*}{V_z^*}, \tag{2.31}$$

$$MSE\left(t_{17}\right)_{opt.} \approx MSE\left(t_7\right) - \frac{V_{yz} - \beta_{yx} V_{xz}}{V_z}. \tag{2.32}$$

3. PROPOSED GENERALIZED REGRESSION-PRODUCT ESTIMATOR

Following Ismail and Shahbaz (2017), we proposed generalized regression-product type estimator for population mean using two auxiliary variables in the presence of non-response in two phase sampling scheme under situation I-II.

The proposed estimator under situation I is

$$t_{ng1} = \left(\bar{y}^* + \hat{\beta}_{yx}\left(\bar{X} - \bar{x}^*\right) + \hat{\beta}_{yz}\left(\bar{Z} - \bar{z}^*\right)\right)\left(\frac{\bar{x}^*}{\bar{X}}\right)^a \left(\frac{\bar{z}^*}{\bar{Z}}\right)^b, \tag{3.1}$$

where $\hat{\beta}_{yx}$ and $\hat{\beta}_{yz}$ are estimates of regression coefficients and a and b are constants.

To find out bias and mean square error of (3.1) we describe some well-known relations as:

$$\bar{y}^* = \bar{Y} + \bar{e}_y^*, \quad \bar{x}^* = \bar{X} + \bar{e}_x^*, \quad \bar{z}^* = \bar{Z} + \bar{e}_z^*,$$
$$\hat{\beta}_{yx} = \beta_{yx} + \bar{e}_{r1}, \quad \hat{\beta}_{yz} = \beta_{yz} + \bar{e}_{r2}. \tag{3.2}$$

Such that

$$E\left(\bar{e}_y^*\right) = E\left(\bar{e}_x^*\right) = E\left(\bar{e}_z^*\right) = 0$$

and

$$\left.\begin{aligned} E\left(\bar{e}_y^{*2}\right) &= \lambda_2 S_y^2 + \theta S_{y(2)}^2 = V_y^*, \quad & E\left(\bar{e}_x^{*2}\right) &= \lambda_2 S_x^2 + \theta S_{x(2)}^2 = V_x^*, \\ E\left(\bar{e}_z^{*2}\right) &= \lambda_2 S_z^2 + \theta S_{z(2)}^2 = V_z^*, \quad & E\left(\bar{e}_y^*\bar{e}_x^*\right) &= \lambda_2 S_{xy} + \theta S_{xy(2)} = V_{xy}^*, \\ E\left(\bar{e}_y^*\bar{e}_z^*\right) &= \lambda_2 S_{yz} + \theta S_{yz(2)} = V_{yz}^*, \quad & E\left(\bar{e}_x^*\bar{e}_z^*\right) &= \lambda_2 S_{xz} + \theta S_{xz(2)} = V_{xz}^*. \end{aligned}\right\} \tag{3.3}$$

Using the above relations (3.1) may be written as

$$t_{ng1} = \left(\bar{Y} + \bar{e}_y^* - \left(\beta_{yx} + \bar{e}_{r1}\right)\bar{e}_x^* - \left(\beta_{yz} + \bar{e}_{r2}\right)\bar{e}_z^*\right)\left(1 + \frac{\bar{e}_x^*}{\bar{X}}\right)^a \left(1 + \frac{\bar{e}_z^*}{\bar{Z}}\right)^b. \tag{3.4}$$

Expanding the right hand side of (3.4) and neglecting terms order higher than two, we have

$$\begin{aligned} t_{ng1} - \bar{Y} \approx{}& \bar{e}_y^* + \left(aR_1 - \beta_{yx}\right)\bar{e}_x^* + \left(bR_2 - \beta_{yz}\right)\bar{e}_z^* - \bar{e}_r\bar{e}_x^* - \bar{e}_r\bar{e}_z^* \\ &+ \frac{a}{2\bar{X}}\left(\left(a-1\right)R_1 - 2\beta_{yx}\right)\bar{e}_x^{*2} + \frac{b}{2\bar{Z}}\left(\left(b-1\right)R_2 - \beta_{yz}\right)\bar{e}_z^{*2} \\ &+ \frac{a}{\bar{X}}\bar{e}_y^*\bar{e}_x^* + \frac{b}{\bar{Z}}\bar{e}_y^*\bar{e}_z^* - \left(\frac{a\beta_{yz}}{\bar{X}} + \frac{b\beta_{yx}}{\bar{Z}} - \frac{abR_1}{\bar{Z}}\right)\bar{e}_x^*\bar{e}_z^*. \end{aligned} \tag{3.5}$$

Applying expectation on both sides of (3.5) we get the expression for bias of (3.1)

$$\begin{aligned} B\left(t_{ng1}\right) \approx{}& \frac{a}{2\bar{X}}\left(\left(a-1\right)R_1 - 2\beta_{yx}\right)V_x^* + \frac{b}{2\bar{Z}}\left(\left(b-1\right)R_2 - \beta_{yz}\right)V_z^* \\ &+ \frac{a}{\bar{X}}V_{xy}^* + \frac{b}{\bar{Z}}V_{yz}^* - \left(\frac{a\beta_{yz}}{\bar{X}} + \frac{b\beta_{yx}}{\bar{Z}} - \frac{abR_1}{\bar{Z}}\right)V_{xz}^*. \end{aligned} \tag{3.6}$$

Squaring both sides of (3.5) neglecting terms of order higher than two, we get

$$\begin{aligned} \left(t_{ng1} - \bar{Y}\right)^2 \approx{}& \bar{e}_y^{*2} + \left(aR_1 - \beta_{yx}\right)^2 \bar{e}_x^{*2} + \left(bR_2 - \beta_{yz}\right)^2 \bar{e}_z^{*2} + 2\left(aR_1 - \beta_{yx}\right)\bar{e}_x^*\bar{e}_y^* \\ &+ 2\left(bR_2 - \beta_{yz}\right)\bar{e}_y^*\bar{e}_z^* + 2\left(aR_1 - \beta_{yx}\right)\left(bR_2 - \beta_{yz}\right)\bar{e}_x^*\bar{e}_z^*. \end{aligned} \tag{3.7}$$

Applying expectation on both sides of (3.7) and simplifying we get mean square error of (3.1) upto first degree of approximation.

$$MSE\left(t_{ng1}\right) \approx V_y^* + \left(aR_1 - \beta_{yx}\right)^2 V_x^* + \left(bR_2 - \beta_{yz}\right)^2 V_z^* + 2\left(aR_1 - \beta_{yx}\right)V_{xy}^*$$
$$+2\left(bR_2 - \beta_{yz}\right)V_{yz}^* + 2\left(aR_1 - \beta_{yx}\right)\left(bR_2 - \beta_{yz}\right)V_{xz}^*. \tag{3.8}$$

The optimum values of *a* and *b* are

$$a = \frac{AV_z^* - BV_{xz}^*}{R_1\left(V_x^*V_z^* - V_{xz}^{*2}\right)}, \tag{3.9}$$

and

$$b = \frac{BV_x^* - AV_{xz}^*}{R_2\left(V_x^*V_z^* - V_{xz}^{*2}\right)}, \tag{3.10}$$

where

$$A = \beta_{yx}V_x^* - V_{xy}^* + \beta_{yz}V_{xz}^*, \qquad B = \beta_{yz}V_z^* - V_{yz}^* + \beta_{xy}V_{xz}^*.$$

Putting the values of a and b in (3.8), we get

$$MSE\left(t_{ng1}\right)_{opt.} \approx V_y^* - \frac{V_x^*V_{yz}^{*2} + V_z^*V_{xy}^{*2} - 2V_{xy}^*V_{yz}^*V_{xz}^*}{V_x^*V_z^* - V_{xz}^{*2}}. \tag{3.11}$$

Some special cases of tng1 along with their respective mean square error taking a few possible values of constants *a* and *b* into consideration are presented in following table.

Table 1
Special Cases of Generalized Estimator t_{ng1}
For Different Values of a and b

a	*b*	Estimator	MSE
0	0	$t_{n1(1)} = \bar{y}^* + \hat{\beta}_{yx}\left(\bar{X} - \bar{x}^*\right) + \hat{\beta}_{yz}\left(\bar{Z} - \bar{z}^*\right)$	$MSE\left(t_4\right) + \beta_{yz}^2 V_z^* - 2\beta_{yz}V_{yz}^* + 2\beta_{yx}\beta_{yz}V_{xz}^*$
1	0	$t_{n2(1)} = \left(\bar{y}^* + \hat{\beta}_{yx}\left(\bar{X} - \bar{x}^*\right) + \hat{\beta}_{yz}\left(\bar{Z} - \bar{z}^*\right)\right)\left(\dfrac{\bar{x}^*}{\bar{X}}\right)$	$V_y^* + \phi^2 V_x^* + \beta_{yz}^2 V_z^* + 2\phi V_{xy}^* - 2\beta_{yz}V_{yz}^* - 2\phi\beta_{yz}V_{xz}^*$
0	1	$t_{n3(1)} = \left(\bar{y}^* + \hat{\beta}_{yx}\left(\bar{X} - \bar{x}^*\right) + \hat{\beta}_{yz}\left(\bar{Z} - \bar{z}^*\right)\right)\left(\dfrac{\bar{z}^*}{\bar{Z}}\right)$	$MSE\left(t_4\right) + \psi^2 V_z^* + 2\psi V_{yz}^* - 2\psi\beta_{yx}V_{xz}^*$
1	1	$t_{n4(1)} = \left(\bar{y}^* + \hat{\beta}_{yx}\left(\bar{X} - \bar{x}^*\right) + \hat{\beta}_{yz}\left(\bar{Z} - \bar{z}^*\right)\right)\left(\dfrac{\bar{x}^*}{\bar{X}}\right)\left(\dfrac{\bar{z}^*}{\bar{Z}}\right)$	$V_y^* + \phi^2 V_x^* + \psi^2 V_z^* + 2\phi V_{xy}^* + 2\psi V_{yz}^* + 2\phi\psi V_{xz}^*$

$\phi = R_1 - \beta_{yx}$, and $\psi = R_2 - \beta_{yz}$.

The proposed estimator under situation II is

$$t_{ng2} = \left(\bar{y}^* + \hat{\beta}_{yx}\left(\bar{X} - \bar{x}\right) + \hat{\beta}_{yz}\left(\bar{Z} - \bar{z}\right)\right)\left(\frac{\bar{x}}{\bar{X}}\right)^a \left(\frac{\bar{z}}{\bar{Z}}\right)^b, \tag{3.12}$$

where $\hat{\beta}_{yx}$ and $\hat{\beta}_{yz}$ are estimates of regression coefficients and *a* and *b* are constants.

To find out bias and mean square error of (3.12) we describe some well-known relations as:

$$\bar{y}^{*} = \bar{Y} + \bar{e}_y^{*}, \quad \bar{x} = \bar{X} + \bar{e}_x, \quad \bar{z} = \bar{Z} + \bar{e}_z, \quad \hat{\beta}_{yx} = \beta_{yx} + \bar{e}_{r1}, \quad \hat{\beta}_{yz} = \beta_{yz} + \bar{e}_{r2}. \tag{3.13}$$

Such that

$$E\left(\bar{e}_y^{*}\right) = E\left(\bar{e}_x\right) = E\left(\bar{e}_z\right) = 0$$

and

$$\left.\begin{aligned} &E\left(\bar{e}_y^{*2}\right) = \lambda_2 S_y^2 + \theta S_{y(2)}^2 = V_y^{*}, && E\left(\bar{e}_x^{2}\right) = \lambda_2 S_x^2 = V_x, \\ &E\left(\bar{e}_z^{2}\right) = \lambda_2 S_z^2 = V_z, && E\left(\bar{e}_y \bar{e}_x\right) = \lambda_2 S_{xy} = V_{xy}, \\ &E\left(\bar{e}_y \bar{e}_z\right) = \lambda_2 S_{yz} = V_{yz}, && E\left(\bar{e}_x \bar{e}_z\right) = \lambda_2 S_{xz} = V_{xz}. \end{aligned}\right\} \tag{3.14}$$

Using the above relations (3.12) may be written as

$$t_{ng2} = \left(\bar{Y} + \bar{e}_y^{*} - \left(\beta_{yx} + \bar{e}_{r1}\right)\bar{e}_x - \left(\beta_{yz} + \bar{e}_{r2}\right)\bar{e}_z\right)\left(1 + \frac{\bar{e}_x}{\bar{X}}\right)^{a}\left(1 + \frac{\bar{e}_z}{\bar{Z}}\right)^{b}. \tag{3.15}$$

Expanding the right hand side of (3.15) and neglecting terms order higher than two, we have

$$\begin{aligned} t_{ng2} - \bar{Y} \approx\; & \bar{e}_y^{*} + \left(aR_1 - \beta_{yx}\right)\bar{e}_x + \left(bR_2 - \beta_{yz}\right)\bar{e}_z - \bar{e}_r\bar{e}_x - \bar{e}_r\bar{e}_z \\ & + \frac{a}{2\bar{X}}\left(\left(a-1\right)R_1 - 2\beta_{yx}\right)\bar{e}_x^2 + \frac{b}{2\bar{Z}}\left(\left(b-1\right)R_2 - \beta_{yz}\right)\bar{e}_z^2 \\ & + \frac{a}{\bar{X}}\bar{e}_y\bar{e}_x + \frac{b}{\bar{Z}}\bar{e}_y\bar{e}_z - \left(\frac{a\beta_{yz}}{\bar{X}} + \frac{b\beta_{yx}}{\bar{Z}} - \frac{abR_1}{\bar{Z}}\right)\bar{e}_x\bar{e}_z. \end{aligned} \tag{3.16}$$

Applying expectation on both sides of (3.16) we get the expression for bias of (3.12)

$$\begin{aligned} B\left(t_{ng2}\right) \approx\; & \frac{a}{2\bar{X}}\left(\left(a-1\right)R_1 - 2\beta_{yx}\right)V_x + \frac{b}{2\bar{Z}}\left(\left(b-1\right)R_2 - \beta_{yz}\right)V_z \\ & + \frac{a}{\bar{X}}V_{xy} + \frac{b}{\bar{Z}}V_{yz} - \left(\frac{a\beta_{yz}}{\bar{X}} + \frac{b\beta_{yx}}{\bar{Z}} - \frac{abR_1}{\bar{Z}}\right)V_{xz}. \end{aligned} \tag{3.17}$$

Squaring both sides of (3.16) neglecting terms of order higher than two, we get

$$\begin{aligned} \left(t_{ng2} - \bar{Y}\right)^2 \approx\; & \bar{e}_y^{*2} + \left(aR_1 - \beta_{yx}\right)^2\bar{e}_x^2 + \left(bR_2 - \beta_{yz}\right)^2\bar{e}_z^2 + 2\left(aR_1 - \beta_{yx}\right)\bar{e}_x\bar{e}_y \\ & + 2\left(bR_2 - \beta_{yz}\right)\bar{e}_y\bar{e}_z + 2\left(aR_1 - \beta_{yx}\right)\left(bR_2 - \beta_{yz}\right)\bar{e}_x\bar{e}_z. \end{aligned} \tag{3.18}$$

Applying expectation on both sides of (3.18) and simplifying we get mean square error of (3.12) upto first degree of approximation.

$$MSE\left(t_{ng2}\right) \approx V_y^* + \left(aR_1 - \beta_{yx}\right)^2 V_x + \left(bR_2 - \beta_{yz}\right)^2 V_z + 2\left(aR_1 - \beta_{yx}\right)V_{xy} + 2\left(bR_2 - \beta_{yz}\right)V_{yz} + 2\left(aR_1 - \beta_{yx}\right)\left(bR_2 - \beta_{yz}\right)V_{xz}. \quad (3.19)$$

The optimum values of a and b are

$$a = \frac{AV_z - BV_{xz}}{R_1\left(V_xV_z - V_{xz}^2\right)}, \quad (3.20)$$

and

$$b = \frac{BV_x - AV_{xz}}{R_2\left(V_xV_z - V_{xz}^2\right)}, \quad (3.21)$$

where

$$A = \beta_{yx}V_x - V_{xy} + \beta_{yz}V_{xz}, \qquad B = \beta_{yz}V_z - V_{yz} + \beta_{xy}V_{xz}.$$

Putting the values of a and b in (3.19), we get

$$MSE\left(t_{ng2}\right)_{opt.} \approx V_y^* - \frac{V_xV_{yz}^2 + V_zV_{xy}^2 - 2V_{xy}V_{yz}V_{xz}}{V_xV_z - V_{xz}^2}. \quad (3.22)$$

Some special cases of t_{ng1} along with their respective mean square error taking a few possible values of constants a and b into consideration are presented in following table.

Table 2

Special Cases of Generalized Estimator t_{ng2} For Different Values of a and b

a	*b*	Estimator	MSE
0	0	$t_{n1(2)} = \bar{y}^* + \hat{\beta}_{yx}\left(\bar{X} - \bar{x}\right) + \hat{\beta}_{yz}\left(\bar{Z} - \bar{z}\right)$	$MSE\left(t_7\right) + \beta_{yz}^2V_z - 2\beta_{yz}V_{yz} + 2\beta_{yx}\beta_{yz}V_{xz}$
1	0	$t_{n2(2)} = \left(\bar{y}^* + \hat{\beta}_{yx}\left(\bar{X} - \bar{x}\right) + \hat{\beta}_{yz}\left(\bar{Z} - \bar{z}\right)\right)\left(\frac{\bar{x}}{\bar{X}}\right)$	$V_y^* + \phi^2V_x + \beta_{yz}^2V_z + 2\phi V_{xy} - 2\beta_{yz}V_{yz} - 2\phi\beta_{yz}V_{xz}$
0	1	$t_{n3(2)} = \left(\bar{y}^* + \hat{\beta}_{yx}\left(\bar{X} - \bar{x}\right) + \hat{\beta}_{yz}\left(\bar{Z} - \bar{z}\right)\right)\left(\frac{\bar{z}}{\bar{Z}}\right)$	$MSE\left(t_4\right) + \psi^2V_z + 2\psi V_{yz} - 2\psi\beta_{yx}V_{xz}$
1	1	$t_{n4(2)} = \left(\bar{y}^* + \hat{\beta}_{yx}\left(\bar{X} - \bar{x}\right) + \hat{\beta}_{yz}\left(\bar{Z} - \bar{z}\right)\right)\left(\frac{\bar{x}}{\bar{X}}\right)\left(\frac{\bar{z}}{\bar{Z}}\right)$	$V_y^* + \phi^2V_x + \psi^2V_z + 2\phi V_{xy} + 2\psi V_{yz} + 2\phi\psi V_{xz}$

$\phi = R_1 - \beta_{yx}$, and $\psi = R_2 - \beta_{yz}$.

5. EFFICIENCY COMPARISON

The proposed estimators are compared with other estimators taken into consideration under two situations prescribed in section 2.

Situation I

i) $MSE\left(t_{ng2}\right)_{opt.} < Var\left(t_1\right)$, if $\dfrac{V_x V_{yz}^2 + V_z V_{xy}^2 - 2V_{xy}V_{yz}V_{xz}}{V_x V_z - V_{xz}^2} > 0,$

ii) $MSE\left(t_{ng1}\right)_{opt.} < MSE\left(t_2\right)$, if $\dfrac{V_x^* V_{yz}^{*2} + V_z^* V_{xy}^{*2} - 2V_{xy}^* V_{yz}^* V_{xz}^*}{V_x^* V_z^* - V_{xz}^{*2}} + R_1^2 V_x^* - 2R_1 V_{xy}^* > 0,$

iii) $MSE\left(t_{ng1}\right)_{opt.} < MSE\left(t_3\right)$, if $\dfrac{V_x^* V_{yz}^{*2} + V_z^* V_{xy}^{*2} - 2V_{xy}^* V_{yz}^* V_{xz}^*}{V_x^* V_z^* - V_{xz}^{*2}} + R_1^2 V_x^* + 2R_1 V_{xy}^* > 0,$

iv) $MSE\left(t_{ng1}\right)_{opt.} < MSE\left(t_4\right)$, if $\dfrac{V_x^* V_{yz}^{*2} + V_z^* V_{xy}^{*2} - 2V_{xy}^* V_{yz}^* V_{xz}^*}{V_x^* V_z^* - V_{xz}^{*2}} + \beta_{yx}^2 V_x^* - 2\beta_{yx} V_{xy}^* > 0,$

v) $MSE\left(t_{ng1}\right)_{opt.} < MSE\left(t_8\right)_{opt.}$, if $\dfrac{V_x^* V_{yz}^{*2} + V_z^* V_{xy}^{*2} - 2V_{xy}^* V_{yz}^* V_{xz}^*}{V_x^* V_z^* - V_{xz}^{*2}} - \dfrac{V_{xy}^{*2}}{V_x^*} > 0,$

vi) $MSE\left(t_{ng1}\right)_{opt.} < MSE\left(t_{11}\right)_{opt.}$, if

$$\frac{V_x^* V_{yz}^{*2} + V_z^* V_{xy}^{*2} - 2V_{xy}^* V_{yz}^* V_{xz}^*}{V_x^* V_z^* - V_{xz}^{*2}} - \left(\frac{V_{xy}^2}{V_x} + \frac{\left(V_{xy}^* - V_{xy}\right)^2}{V_x^* - V_x}\right) > 0,$$

vii) $MSE\left(t_{ng1}\right)_{opt.} < MSE\left(t_{12}\right)_{opt.}$, if

$$\frac{V_x^* V_{yz}^{*2} + V_z^* V_{xy}^{*2} - 2V_{xy}^* V_{yz}^* V_{xz}^*}{V_x^* V_z^* - V_{xz}^{*2}} + \beta_{yx}^2 V_x^* - 2\beta_{yx} V_{xy}^* - N^{*2} / D^* > 0,$$

viii) $MSE\left(t_{ng1}\right)_{opt.} < MSE\left(t_{14}\right)_{opt.}$, if

$$\frac{V_x^* V_{yz}^{*2} + V_z^* V_{xy}^{*2} - 2V_{xy}^* V_{yz}^* V_{xz}^*}{V_x^* V_z^* - V_{xz}^{*2}} + R_1\left(R_1 V_x^* - 2V_{xy}^*\right) + \beta_{yz}\left(\beta_{yz} V_z^* - 2\left(V_{yz}^* - R_1 \beta_{yz} V_{xz}^*\right)\right) - \frac{A^{*2}}{B^*}, > 0,$$

ix) $MSE\left(t_{ng1}\right)_{opt.} < MSE\left(t_{16}\right)_{opt.}$, if

$$\frac{V_x^* V_{yz}^{*2} + V_z^* V_{xy}^{*2} - 2V_{xy}^* V_{yz}^* V_{xz}^*}{V_x^* V_z^* - V_{xz}^{*2}} + \beta_{yx}^2 V_x^* - 2\beta_{yx} V_{xy}^* - \frac{V_{yz}^* - \beta_{yx} V_{xz}^*}{V_z^*} > 0.$$

Situation II

i) $MSE\left(t_{ng1}\right)_{opt.} < Var\left(t_1\right)$, if $\dfrac{V_x^* V_{yz}^{*2} + V_z^* V_{xy}^{*2} - 2V_{xy}^* V_{yz}^* V_{xz}^*}{V_x^* V_z^* - V_{xz}^{*2}} > 0,$

ii) $MSE\left(t_{ng2}\right)_{opt.} < MSE\left(t_5\right)$, if $\dfrac{V_x V_{yz}^2 + V_z V_{xy}^2 - 2V_{xy} V_{yz} V_{xz}}{V_x V_z - V_{xz}^2} + R_1^2 V_x - 2R_1 V_{xy} > 0,$

iii) $MSE\left(t_{ng2}\right)_{opt.} < MSE\left(t_6\right)$, if $\dfrac{V_x V_{yz}^2 + V_z V_{xy}^2 - 2V_{xy} V_{yz} V_{xz}}{V_x V_z - V_{xz}^2} + R_1^2 V_x + 2R_1 V_{xy} > 0,$

iv) $MSE\left(t_{ng2}\right)_{opt.} < MSE\left(t_7\right)$, if $\dfrac{V_x V_{yz}^2 + V_z V_{xy}^2 - 2V_{xy} V_{yz} V_{xz}}{V_x V_z - V_{xz}^2} + \beta_{yx}^2 V_x - 2\beta_{yx} V_{xy} > 0,$

v) $MSE\left(t_{ng2}\right)_{opt.} < MSE\left(t_9\right)_{opt.}$, if $\dfrac{V_x V_{yz}^2 + V_z V_{xy}^2 - 2V_{xy} V_{yz} V_{xz}}{V_x V_z - V_{xz}^2} - \dfrac{V_{xy}^2}{V_x} > 0,$

vi) $MSE\left(t_{ng2}\right)_{opt.} < MSE\left(t_{10}\right)_{opt.}$, if

$$\frac{V_x V_{yz}^2 + V_z V_{xy}^2 - 2V_{xy} V_{yz} V_{xz}}{V_x V_z - V_{xz}^2} + \beta_{yx}^2 V_x - 2\beta_{yx} V_{xy} > 0,$$

vii) $MSE\left(t_{ng2}\right)_{opt.} < MSE\left(t_{13}\right)_{opt.}$, if

$$\frac{V_x V_{yz}^2 + V_z V_{xy}^2 - 2V_{xy} V_{yz} V_{xz}}{V_x V_z - V_{xz}^2} + \beta_{yx}^2 V_x - 2\beta_{yx} V_{xy} - A^2 V_z > 0,$$

viii) $MSE\left(t_{ng2}\right)_{opt.} < MSE\left(t_{15}\right)_{opt.}$, if

$$\frac{V_x V_{yz}^2 + V_z V_{xy}^2 - 2V_{xy} V_{yz} V_{xz}}{V_x V_z - V_{xz}^2} + R_1\left(R_1 V_x - 2V_{xy}\right) + \beta_{yz}\left(\beta_{yz} V_z - 2\left(V_{yz} - R_1 \beta_{yz} V_{xz}\right)\right) - \frac{A^2}{B} > 0,$$

ix) $MSE\left(t_{ng2}\right)_{opt.} < MSE\left(t_{17}\right)_{opt.}$, if

$$\frac{V_x V_{yz}^2 + V_z V_{xy}^2 - 2V_{xy} V_{yz} V_{xz}}{V_x V_z - V_{xz}^2} + \beta_{yx}^2 V_x - 2\beta_{yx} V_{xy} - \frac{V_{yz} - \beta_{yx} V_{xz}}{V_z} > 0.$$

The proposed estimators are more efficient than other estimators taken into consideration if above conditions are satisfied.

6. EMPIRICAL COMPARISON BETWEEN PROPOSED ESTIMATORS WITH SOME OF THE EXISTING ESTIMATORS

To compare performance of proposed estimators with other estimators taken into consideration, following three data sets from literature which are utilized by several authors recently are used with different values of *h i-e* 2, 3, 4 and 5.

i. Data Set 1 {Source: Khare and Sinha (2007)}
   y: weight in kg of the children,
   x: skull circumference in cm of the children,

z: chest circumference in cm of the children.

$\bar{Y}$ = 19.4968, $\bar{X}$ = 51.1726, $\bar{Z}$ = 55.8611, S_y^2 = 9.2662, S_x^2 =2.3662,

S_z^2 =10.7155, $S_{y(2)}^2$ = 5.5424, $S_{x(2)}^2$ = 1.6080, $S_{z(2)}^2$ = 9.1060, S_{xy} = 1.5359,

S_{yz} = 8.4300, S_{xz} = 1.4955, $S_{xy(2)}$ = 1.4240, $S_{yz(2)}$ = 5.1790, $S_{xz(2)}$ = 2.1811,

N = 95, n_1 = 35, n_2 = 24, W_2 = 0.25.

ii. Data Set 2 {Source: Khare and Sinha (2009)}
y: Number of agricultural labors in the village,
x: Area (in hectares) of the village,
z: Number of cultivators in the village.

$\bar{Y}$ = 137.92, $\bar{X}$ = 144.87, $\bar{Z}$ = 185.21, S_y^2 = 33143.8046, S_x^2 = 13769.7786,

S_z^2 = 37818.7754, $S_{y(2)}^2$ = 82296.4624, $S_{x(2)}^2$ = 18544.3932, $S_{z(2)}^2$ = 75136.7307, S_{xy} = 16449.6016, S_{yz} = 27615.284, S_{xz} = 18484.2668, $S_{xy(2)}$ = 28127.3873,

$S_{yz(2)}$ = 61335.4186, $S_{xz(2)}$ = 26876.0201, N = 96, n_1 = 35, n_2 = 24, W_2 = 0.25.

iii. Data Set 3 {Source: Khare and Sinha (2012)}
y: Number of literate persons in the village,
x: Number of literate workers in the village,
z: Number of non-workers in the village.

$\bar{Y}$ = 145.30, $\bar{X}$ = 165.26, $\bar{Z}$ = 259.08, S_y^2 = 12194.3432, S_x^2 =12628.5452,

S_z^2 =38769.925, $S_{y(2)}^2$ = 9762.2304, $S_{x(2)}^2$ = 8873.3009, $S_{z(2)}^2$ = 19572.9054,

S_{xy} = 10051.7317, S_{yz} = 19569.0254, S_{xz} = 17922.9363, $S_{xy(2)}$ = 7259.5840, $S_{yz(2)}$ = 12026.0063, $S_{xz(2)}$ = 9752.1859, N = 109, n_1 = 35, n_2 = 24, W_2 = 0.25.

Table 3
Percent Relative Efficiencies (PREs) of Estimators with Respect to t_1 Considering Situation I for Data Set 1

Estimators	**1/*h***			
	1/2	**1/3**	**1/4**	**1/5**
t_1	100	100	100	100
t_2	111.1080	112.0632	112.7904	113.3626
t_3	85.0489	84.4118	83.9403	83.5773
t_4	114.2695	115.9052	117.1629	118.1601
t_8	114.3350	116.1003	117.5027	118.6418
t_{11}	114.6239	116.5316	118.0045	119.1760
t_{12}, t_{16}	264.2326	246.1211	235.468	228.4956
t_{14}	273.5685	250.6667	237.5283	229.0813
t_{ng1}	**312.1812**	**288.1801**	**273.7346**	**264.0870**

Table 4
Percent Relative Efficiencies (PREs) of Estimators with respect to t_1 Considering Situation I for Data Set 2

Estimators	1/*h*			
	1/2	1/3	1/4	1/5
t_1	100	100	100	100
t_2	205.4402	197.0242	192.8781	190.4098
t_3	47.3872	49.2439	50.2777	50.9363
t_4	221.5195	213.6198	209.6958	207.3497
t_8	223.1413	216.814	213.9888	212.4308
t_{11}	226.8408	220.4861	217.2919	215.3699
t_{12}	256.3163	258.2567	260.2764	261.9061
t_{14}	276.2881	278.3386	279.7765	280.8165
t_{16}	256.3163	258.2567	260.2764	261.9061
t_{ng1}	**289.9974**	**289.4980**	**289.8362**	**290.2776**

Table 5
Percent Relative Efficiencies (PREs) of Estimators with respect to t_1 Considering Situation I for Data Set 3

Estimators	1/*h*			
	1/2	1/3	1/4	1/5
t_1	100	100	100	100
t_2	277.7405	273.2430	270.1373	267.8639
t_3	31.2396	31.5586	31.7891	31.9633
t_4	282.6989	277.5995	274.0893	271.5255
t_8	282.7123	277.6359	274.1483	271.6047
t_{11}	282.7719	277.7165	274.2356	271.6926
t_{12}, t_{16}	345.6866	342.4363	340.4789	339.2234
t_{14}	127.3882	134.2365	139.5634	143.8253
t_{ng1}	**551.5244**	**533.9889**	**523.5421**	**516.7765**

Table 6
Percent Relative Efficiencies (PREs) of Estimators with respect to t_1 Considering Situation II for Data Set 1

Estimators	1/*h*			
	1/2	1/3	1/4	1/5
t_1	100	100	100	100
t_5	108.0328	106.8068	105.9055	105.2149
t_6	88.0180	89.5513	90.7367	91.6805
t_7, t_9, t_{10}	109.8475	108.3232	107.2075	106.3556
t_{13}, t_{17}	225.5286	191.2266	171.6480	158.9883
t_{15}	235.0479	197.0256	175.7097	162.0728
t_{ng2}	**251.1119**	**206.5172**	**182.2457**	**166.9827**

Table 7
Percent Relative Efficiencies (PREs) of Estimators with respect to t_1 Considering Situation II for Data Set 2

Estimators	1/*h*			
	1/2	**1/3**	**1/4**	**1/5**
t_1	100	100	100	100
t_5	145.1430	127.2391	119.5038	115.1901
t_6	58.0357	66.7693	72.4939	76.5359
t_7, t_9, t_{10}	148.0170	128.7475	120.5148	115.9477
t_{13}, t_{17}	151.0046	130.2908	121.5422	116.7146
t_{15}	153.1981	131.4082	122.2816	117.2648
t_{ng2}	**157.0523**	**133.3402**	**123.5516**	**118.2063**

Table 8
Percent Relative Efficiencies (PREs) of Estimators with respect to t_1 Considering Situation II for Data Set 3

Estimators	1/*h*			
	1/2	**1/3**	**1/4**	**1/5**
t_1	100	100	100	100
t_5	206.7779	175.0737	157.8863	147.1026
t_6	35.8359	40.2117	44.0288	47.3877
t_7, t_9, t_{10}	209.2502	176.5402	158.9041	147.8733
t_{13}, t_{17}	232.2563	189.7053	167.8695	154.5831
t_{15}	114.0070	111.3616	109.5567	108.2466
t_{ng2}	**293.9047**	**221.1716**	**188.1186**	**169.2332**

7. CONCLUSION

Mathematical comparison made with usual unbiased estimator and some other exponential estimator cited in section 2 shows that our proposed estimators are more efficient under the conditions given in section 5. Furthermore, results of empirical study, presented in tables 3 to 8 considering three different data sets revealed that our proposed estimators are more efficient. The percentage relative efficiency of our proposed estimators increases as the sub sampling fraction increases.

REFERENCES

1. Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. New York: John Wiley and Sons.
2. Hansen, M.H. and Hurwitz, W.N. (1946). The problem of non-response in sampling surveys. *Journal of American Statistical Association.* 41, 517-529.
3. Ismail, M. and Shahbaz, M.Q. (2017). Combination of Ratio and Regression Estimator of Population Mean in Presence of Non–response. *GU J. Sci.*, 30(4), 634-642
4. Ismail, M., Hanif, M. and Shahbaz, M.Q. (2013). A New Cost Effective Estimator in the Presence of Non-response for Two-phase Sampling. *Jurnal Teknologi*, 63(2), 1-4.

5. Khare, B.B. and Srivastava, S. (1993). Estimation of population mean using auxiliary character in presence of non-response. *Nat. Acad. Sci. Lett. India*, 16, 111-114.
6. Khare, B.B. and Srivastava, S. (1995). Study of conventional and alternative two-phase sampling ratio, product and regression estimators in presence of non-response. *Nat. Acad. Sci. Lett. India*, 65, 195-203.
7. Khare, B.B. and Srivastava, S. (1997). Transformed ratio type estimators for the population mean in presence of non-response. *Communication in Statistics Theory and Method*, 17, 1779-1791.
8. Khare, B.B. and Srivastava, S. (2010). Generalized two phase estimators for the population mean in the presence of non-response. *The Aligarh journal of Statistics*, 30, 39-54.
9. Khare, B.B. and Sinha, R.R. (2007). Estimation of the ratio of the two population means using multi auxiliary characteristics in the presence of non-response. *In Statistical Techniques in Life Testing, Reliability, Sampling Theory and Quality Control*, 1, 63-171.
10. Khare, B.B. and Sinha, R.R. (2009). On the class of estimators for population mean using multi auxiliary characters in the presence of non-response. *Statistics in Transition-New Series*, 10(1), 3-14.
11. Khare, B.B. and Sinha, R.R. (2012). Improved classes of estimators for ratio of two means with double sampling the non-respondents. *Statistica*, 49(3), 75-83.
12. Pal, S.K. and Singh, H.P. (2017). Estimation of finite population mean using auxiliary information in presence of non-response. *Communications in Statistics - Simulation and Computation*. 47(1), 143-165.
13. Rao, P.S.R.S. (1986). Ratio estimation with sub sampling the non-respondents. *Survey Methodology*, 12(2), 217-230.
14. Rao, P.S.R.S. (1990). Regression estimators with sub sampling of non-respondents, In-*Data Quality Control.* Theory and Pragmatics, (Gunar E. Liepins and V.R.R. Uppuluri, eds) Marcel Dekker, New York, 191-208.
15. Singh, H.P. and Kumar, S. (2008). A regression approach to the estimation of the finite population mean in the presence of non-response. *Aust. N.Z.J. Stat*., 50(4), 395-408.
16. Singh, H.P. and Kumar, S. (2009). A general procedure of estimating the population mean in the presence of non-response under double sampling using auxiliary information. *SORT*, 33(1), 71-84.
17. Singh, H.P. and Kumar, S. (2011). Combination of regression and ratio estimate in presence of non-response. *Braz. J. Probab. and Statist.*, 25(2), 205-217.
18. Singh, H.P., Kumar, S. and Kozak, M. (2010). Improved estimation of finite population mean using sub-sampling to deal with non-response in two phase sampling scheme. *Communication in Statistics Theory and Method*, 39(5), 791-802.
19. Sodipo, A.A. and Obisesan, K.O. (2007). Estimation of the population mean using difference cum ratio estimator with full response on the auxiliary character. *Res. J. App. Scien.*, 2, 769-772.
20. Zubair, M. (2012). *Some regression cum ratio estimators using two phase sampling scheme in case of non-response*. Unpublished M.Phil. Thesis. National College of Business Administration and Economics Lahore, Pakistan.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 199-212

DETERMINANTS OF GROSS DOMESTIC PRODUCT: A COMPARATIVE STUDY OF PAKISTAN AND CHINA

Asad Ali[1§], Muhammad Zubair[2] and **Muhammad Iqbal[2]**

[1] Department of Quantitative Methods, SBE, University of Management and Technology, Lahore, Pakistan

[2] Department of Statistics, University of Sargodha, Sargodha, Pakistan.

[§] Corresponding Author Email: ranaasadali23@gmail.com

ABSTRACT

Gross domestic product (GDP) is a vital indicator for economy of a country. Economy of every country heavily depends on GDP for its growth. There are many determinants who performed their role in increase or decrease of GDP and ultimately on growth of an economy. In this study, effect of some determinants (FDI, EDS, GNE, EGS, IGS, GDS, and NIA) on GDP of Pakistan has been statistically quantified for the period of 1982-2017. For the purpose of comparison, the effect of these determinants has also been measured on the GDP of China for the period under consideration. Multiple Regression has been utilized for the said purpose. Multiple Regression Model suggest that IGS is most significant determinant for the case of Pakistan having direct relationship. While for China, NIA is statistically most significant determinant having inverse relation. Results of Granger Causality Test depicts that, GDP of Pakistan has bi-causality relationship with GNE, IGS, and GDS while it has single causality relation with FDI, EDS, EGS, and NIA. In case of China results depicts that, GDP has bi-causality relationship with all determinants other than GNE.

KEY WORDS

GDP, Regression, Economy, Bi-Causality, CLRM, Granger Causality.

1. INTRODUCTION

Gross domestic product (GDP) is a vital indicator for economy of a country. Economy of every country heavily depends on GDP for its growth. There are many determinants who performed their role in increase or decrease of GDP and ultimately on growth of an economy. Several researchers found the impact of different factors on GDP like Barro (1995), Gregoriou and Ghosh (2009), Kira (2013), Lane (2016), and Shahbaz et al. (2016). Barro (1995) operated regression model and found the impact of inflation on economic growth of 100 countries. The results revealed that as inflation mounting then economic growth rate of real capita GDP decline. Gregoriou and Ghosh (2009) determined that government expenditure had strong growing effects on 15 developing countries' economic growth by utilizing generalized method of moments but had different amount in different countries. Kira (2013) adopted Keynes model to find out the effect of some indicators on Gross Domestic Product of Tanzania. Country's consumption, Net Exports, Imports and Investment are considered as independent variables and found that only Consumption and Export significantly positively influenced the GDP. Lane (2016) examined the relationship of Greenhouse gases with GDP and concluded that extensive decarbonisation would create many difficulties especially for the countries depending on organic fuels. Shahbaz et al. (2016) found the casual relationship for the period of 1985-2010 between economic growth, carbon dioxide emissions and FDI for 117 countries. The results showed that carbon dioxide emission had relationship with FDI as well as economic growth.

OBJECTIVE OF THE STUDY

Important features of study are as follows:

i) To find out efficient determinants of GDP in Pakistan and China.
ii) Suggestions for improvement of GDP in Pakistan and China.

HYPOTHESIS TO BE TESTED

The study focus on testing the following hypothesis for both countries as Pakistan and China:

In case of Pakistan

H_1 : The Model is best fit. ($P<0.05$)
H_0 : The Model is not best fit. ($P>0.05$)

In case of China

H_2 : The Model is best fit. ($P<0.05$)
H_0 : The Model is not best fit. ($P>0.05$)

ALTERNATIVE HYPOTHESIS ABOUT COEFFICIENTS OF REGRESSION OF INDEPENDENT VARIABLES

As it is mentioned in the study that total seven independent variables are used to investigate the effect on GDP, so

H_1 : Foreign Direct Investment affect the GDP
H_2 : External Debt Stocks affect the GDP
H_3 : Gross National Expenditure affect the GDP
H_4 : Exports of Goods and Services affect the GDP
H_5 : Imports of Goods and Services affect the GDP
H_6 : Gross Domestic Savings affect the GDP
H_7 : Net Income from Abroad affect the GDP

2. LITERATURE REVIEW

Literature review is a fundamental part of the research. It tells about the past work related to topic under consideration. With the help of literature review, a researcher comes to know that how to deal with problem under study, how much work has been done by utilizing various methodologies as well as unanswered questions relating to the problem. There are many researches in literature on gross domestic product and economic growth of Pakistan. Some of these are case study of Pakistan and remaining are on comparison of Pakistan with different countries. For the case of Pakistan, Iqbal (1994) is pioneer who found the effectiveness of different variables on GDP by utilizing multiple regression model for the period 1970-91. He concluded that inflation rate, investment, imports of goods, government consumption increased the economic growth during this period while impact of exports of goods was insignificant. Khilji and Mahmood (1997) explored the relationship between Pakistan economic growth and defence expenditures by using granger causality for the period of 1972-1995. Results revealed that, defence expenditures had inverse relationship with Pakistan's economic growth. Iqbal and Zahid (1998) examined effect of some indicators on growth of Pakistan's economy by using multiple regression model for the period 1959-60 to 1996-97. The results showed that primary school enrolment-labour force ratio, stock of physical capital, exports and imports of goods of Pakistan were directly proportional while budget deficit, external debit, and per capita real income were inversely proportional to growth. Shah and Ahmed (2003) used Foreign Direct Investment as a dependent variable and Political Environment as a Dummy Variable of Democratic Government versus Military rule, GDP, Gross National Product, and Effective Import Duty as a Proxy of Tariff Rate, Cost of Capital, and Expenditure on Transport and Communication as independent variables. The data on the above mentioned variables were collected from 1960-61 to 1999-00 for Pakistan. Results of their study revealed that, Foreign Direct Investment had statistically significant long run relation with all variables.

Shirazi and Manap (2004), Iqbal and Sattar (2005), Mubarik (2005), Falki (2009), Rahman (2009), Awan et al. (2010), and Jamil and Ahmad (2010) also study the importance of GDP for Pakistan and effect of different factors on it. Mahmood et al. (2011) utilized the general autoregressive conditional heteroskedasticity model to found the relationship of Exchange rate with four economic variables GDP, Trade Openness, FDI, and Growth rate in Pakistan. The results revealed that, Exchange rate had direct relationship with FDI, and inverse relationship with GDP, Trade Openness, and Growth rate. Ahmad et al. (2012) compared the growth of economy in Pakistan and Bangladesh with the help of development in stock market. They analyzed the results by using regression and found that, in Pakistan stock market increased the economic growth due to its large size but in Bangladesh it had due to its liquidity. They also found that, the economic growth of Bangladesh was in better position than Pakistan. Mehmood (2012) compared the economy of Pakistan and Bangladesh on the basis of Gross Domestic Product and found the effect of 13 different economic variables on it. The study found that some variables like gross national expenditure (GNE), goods exports (GE), gross saving (GS), and final consumption expenditure (FCE) were directly proportional to GDP in Pakistan and gross national expenditure, external debt stock total, goods import and export in case of Bangladesh. Moreover, study also showed that external debts total stock and services exports had inverse relation in Pakistan but in Bangladesh final consumption expenditure had inverse relation with gross domestic product. Batool and Zulfiqar (2013) made a simple regression model by using Gross Domestic Product as an independent variable and Interest Payment on Domestic Debt, Exports, Private Savings, and Total Domestic Debts as dependent variables. The results explained that in Pakistan Total Domestic Debt had inversely proportional to Gross Domestic Product and other variables directly proportional with it. Kumar (2016) studied the relationship of population growth with Pakistan's economic development and found that there had a weak relationship between them.

Rawski (2001) elaborated that, from 1998 the official claims about China's GDP had based on exaggerations. He also claimed that for the period of 1997-2001, the GDP had not greater than the one third of officially announced value. Zhang and Cheng (2009) utilized multivariate regression model and found that energy consumption and carbon emissions had no leading effect on China's economic growth for the period of 1960-2007. Agrawal and Khan (2011) used simple regression to find out the effect of FDI on GDP for China and India. By using data from 1993-2009, results of regression model indicated that FDI had more impact on GDP of China as compare to India. By using Regression model Gao et al. (2016) examined that, transportation freight index had direct relation with GDP of China. Wang et al. (2016) found the casual relation between growth of economy, consumption of energy and carbon dioxide emission. The results depicted that, there had long run relation between the variables for china in 1990 to 2012.

Gaps in the Literature

We see that how different researchers found the efficient factors to improve the gross domestic product of country. Many economists also compared the gross domestic product of one country with another. But, there has been one portion which is still neglected "A comparison of efficient determinants of gross domestic product of Pakistan and China". So, in current study we are not only going to find efficient indicators but also to check that which of them will play important role in prediction of gross domestic product of Pakistan as well as China.

3. MATERIAL AND METHODOLOGY

With the help of literature review, we select seven variables as stimulus for GDP in the case of Pakistan and China for the period of 1982-2017. The seven stimulus (Independent) variables are:

- i) Foreign Direct Investment (FDI) (current US$)
- ii) External Debt Stocks (EDS) (current US$)
- iii) Gross National Expenditure (GNE) (current US$)
- iv) Exports of Goods and Services (EGS) (current US$)
- v) Imports of Goods and Services (IGS) (current US$)
- vi) Gross Domestic Savings (GDS) (current US$)
- vii) Net Income from Abroad (NIA) (current US$)

To investigate the effect of these seven variables on the GDP in Pakistan and China, the data were taken from the World Bank Development Indicators for the period of 1982-2017. The figures of the GDP were taken in current US$. In the study following empirical Multiple Regression Model is utilize to investigate the effect of these seven variables on the GDP in Pakistan and China.

In Case of Pakistan

SPSS and E-views softwares have been utilized for the purpose of analysis, the following model considered as a best fit model for this study. This model is selected from numerous other statistical models which obtained by Multiple Regression Analysis with stepwise method. Results of model are given in Table 1.

$$\mathrm{GDP} = \beta_0 + \beta_1\ \mathrm{FDI} + \beta_2\ \mathrm{EDS} + \beta_3\ \mathrm{EGS} + \beta_4\ \mathrm{IGS} + \beta_5\ \mathrm{GDS} + \beta_6\ \mathrm{NIA} + e$$

Here

- β_0 = Intercept of the model
- β_1 = Coefficient regression of Foreign Direct Investment
- β_2 = Coefficient regression of External Debt Stock
- β_3 = Coefficient regression of Exports of Goods and Services
- β_4 = Coefficient regression of Imports of Goods and Services
- β_5 = Coefficient regression of Gross Domestic Savings
- β_6 = Coefficient regression of Net Income from Abroad
- e = Error Term

After putting the values of slope coefficients in model, the Multiple Regression Line is,

$$\mathrm{GDP} = -1.85\ \mathrm{FDI} + 0.51\ \mathrm{EDS} - 0.94\ \mathrm{EGS} + 3.51\ \mathrm{IGS} + 2.86\ \mathrm{GDS} + 2.26\ \mathrm{NIA}$$

In Case of China

The following model considered as a best fit model for the case of China. This model is selected from numerous other statistical models which obtained by Multiple Regression Analysis with stepwise method. Results of model are given in Table 2.

$$\mathrm{GDP} = \beta_0 + \beta_1\mathrm{FDI} + \beta_2\mathrm{EDS} + \beta_3\mathrm{EGS} + \beta_4\mathrm{GDS} + \beta_5\mathrm{NIA} + \mathrm{e}$$

Here

- β_0 = Intercept of the model
- β_1 = Coefficient regression of Foreign Direct Investment
- β_2 = Coefficient regression of External Debt Stock
- β_3 = Coefficient regression of Exports of Goods and Services
- β_4 = Coefficient regression of Gross Domestic Savings
- β_5 = Coefficient regression of Net Income from Abroad
- e = Error Term

After putting the values of slope coefficients in model, the Multiple Regression Line is,

$$\mathrm{GDP} = -2.49\mathrm{FDI} + 1.22\mathrm{EDS} + 0.42\mathrm{EGS} + 1.69\mathrm{GDS} - 4.26\mathrm{NIA}$$

Granger Causality Test

Granger causality test is utilized to check whether a time series is helpful to forecast other time series. It also used to find bi-causal relationship between two variables. A time series Y is said to granger cause X, if by using series of F test on lagged values of Y and X it shown that Y values provide statistically significant information about next values of X. Granger test only finds the predictive causality.

For the forecasting purpose, Granger Causality test is utilized to check the bi-directional relationship of the indicators with GDP. Hypothesis used in this test are:

First Direction

H0: indicator does not granger cause the gross domestic product

H1: indicator granger causes the gross domestic product

Second Direction

H0: gross domestic product does not granger cause the indicator

H1: gross domestic product granger causes the indicator

To test the above mentioned hypothesis following vector auto-regressions (VAR) are used:

$$X_t = \xi_o \sum_{i=1}^{n} \vartheta_i Y_{t-i} + \sum_{j=1}^{n} \xi_j X_{t-j} + \varepsilon_{1t}$$

$$Y_t = \varphi_o \sum_{i=1}^{m} \varphi_i Y_{t-i} + \sum_{j=1}^{m} \phi_j X_{t-j} + \varepsilon_{2t}$$

If the coefficients $\sum \vartheta_i$ and $\sum \varphi_i$ are confirmed to be not equal to zero by using F test, then null hypothesis proved is false. Which conclude that both indicators are in bi-causal relationship.

4. RESULTS AND DISCUSSION

Gross domestic product of a country includes all final goods as well as services provided in the year under consideration. Its value shows that a county's economy is in progress or not. Figure 1 & 2 shows that gross domestic product of Pakistan and China increase with the passage of time. In this study we take gross domestic product as a dependent variable for formulating an econometric model.

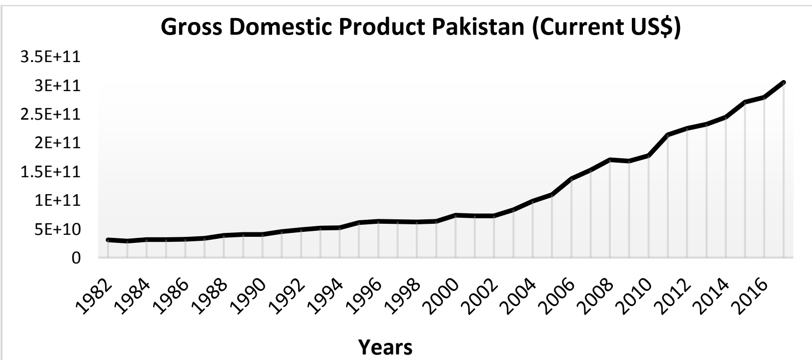


Figure 1: Gross Domestic Product of Pakistan

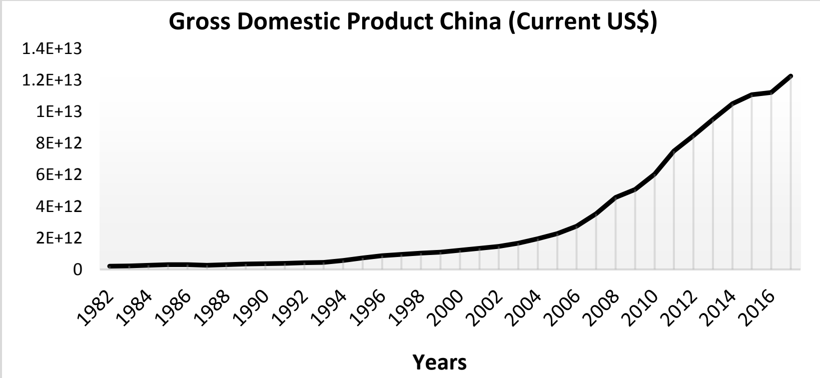


Figure 2: Gross Domestic Product of China

First time in 2002, Pakistan's gross domestic product shows high progress while it reflect its steady growth during the period 1982-2001. In the period 2002-2008, government of Pakistan took many radical steps to improve the economy of Pakistan in which they are fully successful. Some radical steps of government are as follows:

- Reduction in foreign loans from $38 billion to $36 billion.
- Rise in foreign exchange reserves.
- Improvement in balance of payments.
- Deficit up to only Rs. 508 million, which is far less than the average deficit of $3 billion in the past.
- Purchase of over $3 billion by the State Bank from open market is one of the main reason in increase of foreign exchange reserves in start of 2002.
- Another reason in increase of forex reserves is the increase in remittances sent by overseas Pakistanis.
- Pakistan's commitment to reduce poverty was first reflected when in December 2003 poverty Reduction Strategy Paper (PRSP) was finalized, it also plays important role in growth of economy.
- Real GDP reached to 8.4% in 2004-2005 due to increase in large scale manufacturing i.e. 15.4%, impressive recovery in agriculture sector i.e. 7.5% and strong growth in service sector i.e. 7.9%.

But after the elections of 2008, government of Pakistan changed. During the period of new government, Pakistan gross domestic product shows dump in its growth for the period of 2008-2010 but after 2010 there is again an increasing trend. On the other hand, China's GDP is steady for the period of 1982-1994. From 1994, gross domestic product shows minor increase but after 2003 there is rising trend.

In Case of Pakistan

The study obtains the following best fit model for Pakistan:

$$\mathrm{GDP} = -1.85\ \mathrm{FDI} + 0.51\ \mathrm{EDS} - 0.94\ \mathrm{EGS} + 3.51\ \mathrm{IGS} + 2.86\ \mathrm{GDS} + 2.26\ \mathrm{NIA} + \mathrm{e}$$

R-square and Adjusted R-square are touching their maximum possible value i.e. 1, so we can say that the model under consideration explain maximum variability in dependent variable. The values of AIC and SIC are also less than all other possible models during stepwise regression.

Table 1
Best Fit Regression Model for Pakistan

Variable	Coefficient	Std. Error	t-Statistic	Prob.
β_0	0.00	0.00	-4.488	0.000*
β_1	-1.85	1.47	-1.256	0.221
β_2	0.51	0.22	2.261	0.033**
β_3	-0.94	1.22	-0.773	0.447
β_4	3.51	0.60	5.837	0.000*
β_5	2.86	0.85	3.373	0.002*
β_6	2.26	0.69	3.248	0.003*
R-squared			0.99	
Adjusted R-squared			0.99	
AIC			47.23	
SIC			47.55	
F-significant			0.00*	
Durbin-Watson			1.20	
Significant (White Heteroskedasticity Test)			0.05**	

*, ** show significance at 1%, and 5% respectively.

In the test of overall significance of model, F-significant value (p-value) is 0.00 so we don't reject our H_1 in the case of Pakistan that model is best fitted. We also examine the other assumptions of Classical Linear Regression Model (CLRM) i.e. autocorrelation and homoscedasticity. For examining autocorrelation, we have utilized Durbin-Watson d test. The d statistic 1.20 is laying in the no decision zone. So, there are not much evidences for the autocorrelation in the model. White Heteroskedasticity test has been utilized for examining the constant position of variance. Its significant value (p-value) 0.05 indicates that there is homoscedasticity in the model.

In Table 1, results also shows that, the value of regression coefficient for FDI $\beta_1 = -1.85$ which measures the slope of regression line. It indicates that when the value of FDI increases by one unit, the decrease in the value of GDP on the average will be 1.85 units due to its negative sign, while it is assumed that all other coefficients of regression remains constant. Further the value of regression coefficient for EDS β_2 = 0.51, which shows that as the one unit of EDS increases, the increase in the value of GDP on the average will be 0.51 units while all other coefficients of regression remain constant. The value of regression coefficient for EGS $\beta_3 = -0.94$, which shows that as the one unit of EGS increases, the decrease in the value of GDP on the average will be 0.94 units due to its negative sign, while it is assumed that all other coefficients of regression remain constant. Furthermore, the value of regression coefficient for IGS β_4 = 3.51, which shows that if one unit of IGS increase, then the estimated increase in the value of GDP will be 3.51 units, while it is assumed that all other coefficients of regression remain constant. The value of regression coefficient for GDS β_5 = 2.86, which shows that if one unit of

GDS increases, then the increase in the value of GDP on the average will by 2.86 units, while it is assumed that all other coefficients of regression remain constant. On the same pattern, the value of regression coefficient for NIA β_6 = 2.26, which shows that when the increase in NIA is one unit, then the increase in the value of GDP on the average will be 2.26 units while all other coefficients of regression kept constant.

Accept or Reject Hypothesis about Coefficients of Regression in Case of Pakistan

The significant values of these coefficients of regressions show that, except FDI and EGS all other indicators are statistically important determinant for the gross domestic product of Pakistan. Therefore, on the basis of these significant values we reject our H_1 and H_4 while H_2, H_5, H_6 and H_7 are accepted in case of Pakistan.

In Case of China

We obtain the following best fit model:

$$\text{GDP} = -2.49\text{FDI} + 1.22\text{EDS} + 0.42\text{EGS} + 1.69\text{GDS} - 4.26\text{NIA} + \text{e}$$

R-square and Adjusted R-square are touching their maximum possible value i.e. 1, so we can say that the model under consideration explain maximum variability in dependent variable. The values of AIC and SIC are also less than all other possible models during stepwise regression. In the test of overall significance of model, F-significant value (p-value) is 0.00 so we don't reject our H_2 in the case of China that model is best fitted. For examining autocorrelation, we have utilized Durbin-Watson d test. The d statistic 1.43 is laying in the no decision zone. So, there are not much evidences for the autocorrelation in the model. White Heteroskedasticity test has been utilized for examining the constant position of variance. Its significant value (p-value) 0.06 indicates that there is homoscedasticity in the model with 10% level of significance.

Table 2
Best Fit Regression Model for China

Variable	Coefficient	Std. Error	t-Statistic	Prob.
β_0	0.00	0.00	3.589	0.0014*
β_1	-2.49	1.05	-2.363	0.0259**
β_2	1.22	0.05	2.576	0.016**
β_3	0.42	0.02	2.119	0.0438**
β_4	1.69	0.06	27.102	0.000*
β_5	-4.26	1.58	-2.690	0.0123**
R-squared			0.99	
Adjusted R-squared			0.99	
AIC			52.62	
SIC			52.89	
F-significant			0.00*	
Durbin-Watson			1.43	
Significant (White Heteroskedasticity Test)			0.06***	

*, **, *** show significance at 1%, 5% and 10% respectively.

In Table 2, estimated results also show that, the value of regression coefficient for FDI $\beta_1 = -2.49$ which measures the slope of regression line. It indicates that when the value of FDI increases by one unit, the decrease in the value of GDP on the average will be 2.49 units due to its negative sign, while it is assumed that all other coefficients of regression remains constant. The value of regression coefficient for EDS $\beta_2 = 1.22$, indicates that when the value of EDS increases by one unit, the increase in the value of GDP on the average will be 1.22 units while it is assumed that all other coefficients of regression remain constant. Further, the value of regression coefficient for EGS $\beta_3 = 0.42$, indicates that when the value of EGS increases by one unit, the estimated increases in the value of GDP on the average will be 0.42 units while it is assumed that all other coefficients of regression remain constant. The value of regression coefficient for GDS $\beta_4 = 1.69$, indicates that when the value of GDS increases by one unit, the increase in the value of GDP on the average will be 1.69 units while it is assumed that all other coefficients of regression remain constant. On the same pattern, the value of regression coefficient for NIA $\beta_5 = -4.26$, indicates that when the value of NIA increases by one unit, the decrease in the value of GDP on the average will be 4.26 units due to its negative sign, while it is assumed that all other coefficients of regression remain constant.

Accept or Reject Hypothesis about Coefficients of Regression in case of China

The significant values of these coefficients of regressions show that, all indicators are statistically important determinant for the gross domestic product of China. Therefore, on the basis of these significant values we don't reject our H_1, H_2, H_4, H_6 and H_7 in case of China.

Granger Causality Test

In Table 3, by using vector auto-regressions (VAR), results of granger causality test show that which indicator has significant impact in the prediction of gross domestic product of Pakistan and vice versa. On the basis of past studies, we know that granger causality test has sensitivity problem on the lag length of variables (Chontanawat et al. (2006), Fatai et al. (2004), Hassan & Kalim (2012), Hossain & Hossain (2012), Mubarik (2005), Wang et al. (2016) and Zhang & Cheng (2009)). So, we find the results of granger causality test on 4 different lag length values i.e. 2, 3, 4 and 5. Results reveal that, there is very strong evidence for bi-directional relationship of GDP with GNE, IGS and GDS. GDP shows single directional relationship with FDI, EDS, EGS and NIA. From these results we can see that FDI granger cause the GDP but GDP does not granger cause the FDI. So when we have to predict the future value of GDP, FDI plays important role but in the prediction of FDI, GDP does not play important role. In the Same way, for the prediction of GDP there is an important role of NIA but in the prediction of NIA, there in not significant role of GDP. Opposite to these, in the prediction process of GDP, EDS does not play important role but, for the prediction of EDS, GDP is important factor. EGS is not important for the prediction of future values of GDP but GDP is important factor in the forecasting of EGS.

Table 3
Granger Causality Test for Pakistan

Null Hypothesis	Lag Length			
	2	3	4	5
GDP does not Granger Cause FDI	1.254	1.779	1.165	9.831*
FDI does not Granger Cause GDP	2.319	4.203**	3.432**	7.154*
GDP does not Granger Cause EDS	2.703***	2.921***	5.821*	11.833*
EDS does not Granger Cause GDP	0.418	0.166	0.414	0.408
GDP does not Granger Cause GNE	4.538**	2.517***	3.567**	4.365*
GNE does not Granger Cause GDP	2.336	1.107	2.392***	3.718**
GDP does not Granger Cause EGS	4.814**	2.379***	2.576***	1.644
EGS does not Granger Cause GDP	4.788**	2.012	2.054	1.504
GDP does not Granger Cause IGS	7.289*	6.649*	6.157*	4.485*
IGS does not Granger Cause GDP	1.604	0.803	3.339**	2.591***
GDS does not Granger Cause GDP	2.879***	3.061**	2.391***	3.378**
GDP does not Granger Cause GDS	1.179	2.962***	1.897	2.435***
NIA does not Granger Cause GDP	0.998	0.175	0.649	0.910
GDP does not Granger Cause NIA	6.663*	7.279*	4.608*	2.968**

*, **, *** show significance at 1%, 5% and 10% respectively.

To check the bi-directional relation of GDP with other indicators, granger causality test is utilized and its results are displayed in Table 4. The results show that, there is bi-directional relationship of GDP with all indicators except GNE. In the prediction of future values of GDP, GNE does not have much importance. But for the prediction of GNE, GDP will play important role.

Table 4
Granger Causality Test for China

Null Hypothesis	Lag Length			
	2	3	4	5
FDI does not Granger Cause GDP	7.651*	4.868*	2.263***	1.491
GDP does not Granger Cause FDI	10.677*	4.011**	3.629**	2.857**
EDS does not Granger Cause GDP	2.800***	2.328***	3.075**	3.548**
GDP does not Granger Cause EDS	5.633*	6.371*	9.188*	6.165*
GNE does not Granger Cause GDP	2.467***	7.047*	2.002	1.229
GDP does not Granger Cause GNE	4.674**	13.616*	5.095*	2.259***
EGS does not Granger Cause GDP	5.201**	17.245*	6.597*	4.857*
GDP does not Granger Cause EGS	4.960**	10.178*	9.035*	6.093*
IGS does not Granger Cause GDP	3.767**	25.496*	13.192*	9.485*
GDP does not Granger Cause IGS	5.479*	7.539*	14.700*	12.102*
GDS does not Granger Cause GDP	1.142	27.260*	10.852*	11.807*
GDP does not Granger Cause GDS	0.069	21.195*	13.035*	17.409*
NIA does not Granger Cause GDP	1.844	4.797**	0.911	4.078**
GDP does not Granger Cause NIA	6.466*	5.063*	2.292***	2.469***

*, **, *** show significance at 1%, 5% and 10% respectively.

5. CONCLUSION

The first objective of this study is to find important determinants of GDP for Pakistan and China separately. In this study, FDI, EDS, GNE, EGS, IGS, GDS and NIA have been utilized as possible determinants for GDP in the period of 1982 to 2017. Best classical linear regression models (CLRM) are derived for both countries by using these determinants.

In case of Pakistan, GNE is going to drop due to the problem of multicollinearity and remaining six stimulus variables have used in best fitted classical linear regression model (CLRM). On the basis of results in Table 1, we conclude that FDI and EGS are not statistically significant important determinants for GDP. So, hypothesis H_1 and H_4 in case of Pakistan are rejected. Results also reveal that IGS has major impact on GDP of Pakistan in the period of 1982 to 2017. The impact on GDP shows increase of 3.51 units, for the one-unit increase in IGS. While EDS has most minor impact on GDP out of all understudy stimulus variables in case of Pakistan. The impact of one-unit increase in EDS, cause the increase of 0.51 units of GDP.

For China, GNE and IGS are going to drop due to the issue of multicollinearity and remaining five stimulus variables have used in best fitted classical linear regression model (CLRM). On the basis of results provided in Table 2, all the five variables are statistically significantly important determinants for GDP so, hypothesis H_1, H_2, H_4, H_6 and H_7 for China are accepted. Results also show that out of these five variables, NIA and EGS have major and minor impact on GDP of China in the period of 1982 to 2017 respectively. NIA has inverse relation with GDP, but on the other hand, EGS has positive relation with it. Results indicate that for change of one unit in NIA and EGS, the impact on GDP is 4.26 and 0.42 units change respectively.

Granger causality test is also utilized on Pakistan and China separately for the period of 1982 to 2017. Results describe that for the prediction of Pakistan's GDP; FDI, GNE, IGS, GDS and NIA plays vital role. While in the case of China, FDI, EDS, EGS, IGS, GDS and NIA plays important role for the prediction of GDP.

POLICY IMPLICATIONS AND RECOMMENDATIONS

From this study, we suggest that if Pakistan's governments carry the economic policies of previous governments in a steady way then, there will be less structural break points in the economic growth of Pakistan. As compare to Pakistan, China's governments follow a steady plan to increase the economic growth so, Pakistan should also follow the same strategy to enhance its economy. China has main focus on FDI to boost the economy so, Pakistan should also concentrate on this area to increase its economy. Pakistan also ignores the EDS while it is one of the major determinants of GDP in case of China. China has negative NIA which significantly affect its economy so, there is need to make improvements on this side.

Government policies could be framed in the light of this study because the impact of determinants have been statistically quantified. The trend of GDP of both countries will be helpful for the investors to prepare their future plan. This study will be helpful for the policymakers and investors to realize that which determinants are important for the growth of GDP.

For further research, it is suggested to conduct a comparative study on short run and long run behavior of these determinants on gross domestic product for the countries under consideration.

REFERENCES

1. Agrawal, G. and Khan, M.A. (2011). Impact of FDI on GDP: A Comparative Study of China and India. *International Journal of Business and Management*, 6(10), 71-79.
2. Ahmad, Z., Khan, A.A. and Tariq, A. (2012). Stock market development and economic growth: A comparative study of Pakistan and Bangladesh. *African Journal of Business Management*, 6(8), 2985-2989.
3. Awan, M.Z., Khan, B. and Zaman, K. (2010). A Nexus between foreign direct investment & Pakistan's economy. *International Research Journal of Finance and Economics*, 52, 17-27.
4. Barro, R.J. (1995). *Inflation and economic growth*. National Bureau Economic Research (No. 5326, pp. 1-36). Working Paper.
5. Batool, S.A. and Zulfiqar, A. (2013). Domestic debt: boon or curse? A case of Pakistan. *International Journal Trade and Global Markets*, 6(4), 374-383.
6. Chontanawat, J., Hunt, L.C. and Pierse, R. (2006). Causality between energy consumption and GDP: evidence from 30 OECD and 78 non-OECD countries (No. 113). *Surrey Energy Economics Centre (SEEC)*, School of Economics, University of Surrey.
7. Wikipedia.org. *Economy of China*. Retrieved from https://en.wikipedia.org/wiki/Economy_of_China
8. Wikipedia.org. *Economy of Pakistan*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Economy_of_Pakistan
9. Daily Dawn (2005). *Executive summary of Economc Survey 2004-05*. Retrieved from http://www.dawn.com/news/142102/executive-summary-of-economic-survey-2004-05
10. Falki, N. (2009). Impact of Foreign Direct Investment on Economic Growth in Pakistan. *International Review of Business Research Papers*, 5(5), 110-120.
11. Fatai, K., Oxley, L. and Scrimgeour, F.G. (2004). Modelling the causal relationship between energy consumption and GDP in New Zealand, Australia, India, Indonesia, The Philippines and Thailand. *Mathematics and Computers in Simulation*, 64(3-4), 431-445.
12. Gao, Y., Zhang, Y., Li, H., Peng, T. and Hao, S. (2016). Study on the relationship between comprehensive transportation freight index and GDP in China. *Procedia Engineering*, 137, 571-580.
13. Gregoriou, A. and Ghosh, S. (2009). The Impact of Government Expenditure on Growth: Empiricle Evidence from A Heterogeneous Panel. *Bulletin of Economic Research*, 61(1), 95-102.
14. Hassan, M.S. and Kalim, R. (2012). The Triangular Causality Among Education, Health and Economic Growth: A Time Series Analysis of Pakistan. *World Applied Sciences Journal*, 18(2), 196-207.
15. Hossain, A. and Hossain, M. K. (2012). Empirical Relationship between Foreign Direct Investment and Economic Output in South Asian Countries: A Study on Bangladesh, Pakistan and India. *International Business Research*, 5(1), 9-21.
16. Iqbal, Z. (1994). Macroeconomic Effects of Adjustment Kending in Pakistan. *The Pakistan Development Review*, 33(4), 1011-1031.

17. Iqbal, Z. and Sattar, A. (2005). The Contribution of Workers' Remittances to Economic Growth in Pakistan. *Pakistan Institute of Development Economics, Research Paper* No. 187, 1-25.
18. Iqbal, Z. and Zahid, G.M. (1998). Macroeconomic Determinants of Economic Growth in Pakistan. *The Pakistan Development Review*, 37(2), 125-148.
19. Jamil, F. and Ahmad, E. (2010). The relationship between electricity consumption, electricity prices and GDP in Pakistan. *Energy Policy,* 38, 6016-6025.
20. Khilji, N.M. and Mahmood, A. (1997). Military Expenditures and Economic Growth in Pakistan. *The Pakistan Development Review*, 36(4), 791-808.
21. Kira, A.R. (2013). The Factors Affecting Gross Domestic Product (GDP) in Developing Countries: The Case of Tanzania. *European Journal of Business and Management*, 5(4), 148-158.
22. Kumar, M. (2016). Importance of Population Growth in Future Economic Development of Pakistan. *Advances in Social Sciences Research Journal*, 3(2), 42-59.
23. Lane, J.E. (2016). The COP21 Agreement: Greenhouse Gases and the GDP. *International Journal of Research in Humanities and Social Studies*, 3(1), 1-17.
24. Mahmood, I., Ehsanullah, M. and Habib, A. (2011). Exchange rate volatility & macroeconomic variables in Pakistan. *Business Management Dynamics*, *1*(2), 11-22.
25. Mehmood, S. (2012). Effect of Different Factors on Gross Domestic Product: A Comparative Study of Pakistan and Bangladesh. *Academy of Contemporary Research Journal*, 1(1), 11-22.
26. Mubarik, Y.A. (2005). Inflation and Growth: An Estimate of the Threshold Level of Inflation in Pakistan. *SBP-Research Bulletin*, 1(1), 35-44.
27. Institute of Policy Studies (2002). *Pakistan's Economy and Budget 2002-03*. Retrieved from Industry Economy http://www.pakistaneconomist.com/issue2002/issue27/i&e4.htm
28. Rahman, M. (2009). Contributions of Exports, FDI and Expatriates'remittances to Real GDP of Bangladesh, India, Pakistan and Sri Lanka. *Southwestern Economic Review*, 36, 141-153.
29. Rawski, T.G. (2001). What os happening to China's statistics? *China Economic Review*, 12, 347-354.
30. Shah, Z. and Ahmed, Q.M. (2003). The Determinants of Foreign Direct Investment in Pakistan: An Empiricle Investigation. *The Pakistan Development Review*, 42(4), 697-714.
31. Shahbaz, M., Nasreen, S. and Ozturk, I. (2016). FDI, growth and CO2 emissions relationship: Evidence from high, middle and low income countries. *Bulletin of Energy Economics (BEE)*, 4(1), 54-69.
32. Shirazi, N.S. and Manap, T.A.A. (2004). Exports and Economic Growth Nexus: The Case of Pakistan. *The Pakistan Development Review*, 43(4), 563-581.
33. Wang, S., Li, Q., Fang, C. and Zhou, C. (2016). The relationship between economic growth, energy consumption, and CO2 emissions: Empirical evidence from China. *Science of the Total Environment*, 542, 360-371.
34. Zhang, X.P. and Cheng, X.M. (2009). Energy consumption, carbon emissions, and economic growth in China. *Ecological Economics*, 68, 2706-2712.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 213-222

PREDICTION OF RAINFALL IN SARGODHA: A COMPARISON OF VARIOUS TIME SERIES METHODOLOGIES

Asad Ali[1§], **Kanwal Iqbal**[2] and **Madiha Fatima**[3]
Department of Quantitative Methods, SBE,
University of Management and Technology, Lahore, Pakistan
Email: [1] ranaasadali23@gmail.com
[2] F2017204003@umt.edu.pk
[3] F2017204004@umt.edu.pk
[§] Corresponding Author

ABSTRACT

Rainfall prediction is one of the most important and challenging task in the modern world. In this study, different time series methodologies have been utilized to predict monthly rainfall in district Sargodha, Pakistan. Monthly data have been obtained from Pakistan Meteorological Department for the period of January 1981 to December 2017. Data have been divided into two mutually exclusive parts ranging from January 1981 to December 2005 and January 2006 to December 2017. Neural Networking (NN), Autoregressive Integrated Moving Average (ARIMA), and Fourier Series Models have been utilized on in sample observations (January 1981 to December 2005) for predicting the rain fall of out of sample observations (January 2006 to December 2017). On the basis of prediction errors (Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE)), proposed NN model found to be the best among them.

KEYWORDS

Neural Networking, Bipolar Sigmoid, Fourier Transformation, Periodogram, Correlogram, Prediction.

1. INTRODUCTION

Global warming is a very serious, special and hot issue for the meteorologists, scientists and hydrologists from the past decades. Temperature, rain fall and weather are badly affected due to this phenomenon. Rain fall shows very unusual pattern from last few years. So, researchers try to forecast the future values and its pattern. Takasao and Shiiba (1984) introduced different procedures for forecasting. These procedures were; an on-line rainfall-forecasting procedure and on-line runoff-forecasting procedure. In rainfall forecasting system digital radar was used while stochastic or deterministic state-space runoff model was used in runoff procedure. French et al. (1992) proposed a neural network (NN) in space and time for rainfall forecasting by using current field as input. Bremaud and Pointin (1993) stated that at short range rainfall forecasting radar echoes can be used. This was the best way of forecasting for urban hydrological aim. They present a new forecasting method of heavy rainfall as a focal point and observe its motion; associated with convective cells. After comparing this with three other methods ("persistence, global extrapolation and the SCOUT II.0"), it was found that this gives best and reliable results. Burlando et al. (1993) said that there was a process autoregressive moving average (ARMA), followed by rainfall forecasting. They present two approaches for estimation, continuous estimation approach and event-based estimation approach, both follow autoregressive moving average (ARMA) model which was useful in a non-linear process and forecasting was much easier through ARMA model.

There are extensive studies showing that ANNs (Artificial Neural Networks) can be used in order to solve problems in hydrology. Zhang and Scofield (1994) introduced a method for estimation of heavy rainfall and also posed a system for estimation which was ANSER (Artificial Neural network expert system for satellite-derived Estimation of Rainfall). It also described that by using this technique, clouds merger, rainfall amount calculation was ten times faster, and error in estimating rainfall was less than ten percent. Toth et al. (2000) employed three models auto-regressive moving average (ARMA), Artificial Neural Network (ANN) and k-nearest-neighbors (KNN) for forecasting. After applying models ANN considered as a best model in accurate runoff-forecasting. Wong, et al. (2003), presented a soft computing technique which was used for model prediction of rainfall. This technique uses ANN (Artificial Neural Networks) logic. In this technique SOM was used to reduce the complexity of data by dividing the data in different groups, after dividing population BPNNs is used within each cluster to learn about population. Christodoulou et al. (2004) concluded that Weather radar can be used to measure rainfall rate with 23% average error, for homogeneous grouping of data SOM and KNN can be used. Krishnankutty (2006) described that an ANN model gave better result and can be used for rainfall monsoon prediction for long range.

Kumarasiri and Sonnadara (2006) developed a new method based on feed-forward back-ward propagation in which Artificial Neural Networks (ANNs) was used for rainfall forecasting. Paras et al. (2008) introduced a model for weather forecasting, which was used to predict weather parameters humidity, temperature etc. He introduced model by using artificial neural networks (ANNs), which can be used for further prediction of weather. A feature based model gave reliable and accurate results of prediction. Chattopadhyay and Chattopadhyay (2007) derived a new model based on Artificial Neural Network (ANN), consisting of three-layer network. They presented an artificial neural networks (ANNs) model which was used for prediction of summer-monsoon rainfall. He used approximately 75% available data of the year of monsoon months for average prediction of monsoon rainfall. After comparing artificial neural networks (ANN) model with conventional persistence forecast, they showed that artificial neural networks (ANN) was quick and skillful model for monsoon rainfall prediction in India. Reddy and Maity (2007) also posed an approach for Orissa state for regional forecasting of rainfall named as Artificial Intelligence approach. He showed that this was highly useful for non-linear variables for forecasting the rainfall.

Chattopadhyay and Chattopadhyay (2008) developed nineteen neural network (ANN) models by using the total rainfall amount of a given year in the summer-monsoon months as input and average monsoon rainfall as output. For average summer-monsoon rainfall eleven nodes of a neural network were found more technical in forecasting. They concluded that eleven hidden-nodes and three-layered neural network (ANN) has more power to produce a desire effects than asymptotic regression. According to Chen et al. (2010) land falling tropical cyclones (LTCs) was the main cause of flood, fatalities, and heavy rainfall; which can be improved with the advancement of LTCs. They showed in their research that there was association between heavy rainfall and mechanism such as "boundary layer energy transfer, moisture transport, topographic effects" etc... Quantitative pre-capitation estimation (QPE) and quantitative precipitation forecasting (QPF) were recently bettered technologies, which plays a vital role in improving rainfall forecasting.

Chen et al. (2012) presented a new SD scheme (statistical Downloading Scheme) which was used for rainfall's prediction. This scheme was used to search for optimal domain which was the central element. Optimal predictors and rainfall both were closely related. This SD scheme was useful when coupled models were used for forecasting rainfall. Arslan et al. (2013) used quality control chart for checking rainfall and climate changes and also posed a scheme which was used to detect major, minor and gradual (floods) shifts. It was not only used on other determinants i.e. snow melt, temperature, and sea level rise but also used in monsoon duration to measure abrupt changes in rainfall. Nayak et al. (2013) said that back propagation network was useful to get significant result of rainfall prediction and conclude that MLP, BPN, SOM, RBFN, and SVM used in different methods of forecasting gave desired results than other techniques.

Objective

The main objectives of this study are:

- To check the pattern of Rainfall for last 37 years in Sargodha.
- To compare different Time Series methodologies for prediction of Rainfall in Sargodha.

2. MATERIAL AND METHODOLOGY

The main objective of this study is to develop a significant model for predicting rain fall in district Sargodha, Pakistan. For this purpose, monthly data of rain fall (mm) have been collected from Pakistan Meteorological Department for the period of January 1981 to December 2017. Further, data have been divided into two parts i-e in sample part ranging from January 1981 to December 2005 and out of sample part ranging from January 2006 to December 2017. Time plot has been utilized to study the pattern of rain fall. Neural Networking (NN), Autoregressive Integrated Moving Average (ARIMA), and Fourier Series Models have been utilized on in sample observations (January 1981 to December 2005) for predicting the rain fall of out of sample observations (January 2006 to December 2017) in district Sargodha, Pakistan. On the basis of prediction errors (Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE)), all the proposed models have been compared. The model with least prediction errors have been recommended for the policy makers to utilize in future policy making.

3. RESULTS AND DISCUSSION

Time series plot of rain fall in district Sargodha has been constructed. Figure 1 shows that there exist extensive oscillations in the series. The series shows that there is decreasing pattern in rain fall from last decade.

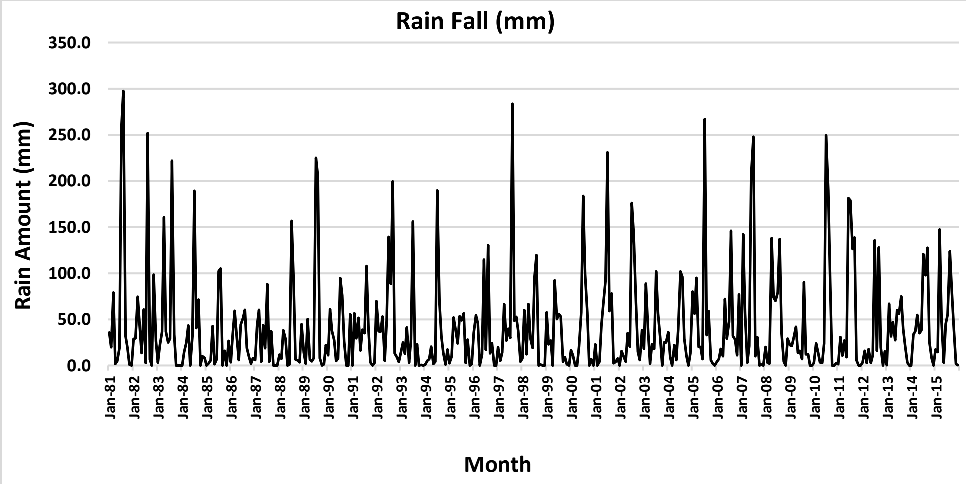


Figure 1: Original Series of Rain Fall in District Sargodha

Neural Networking (NN)

Back Propagation NN with bipolar sigmoid function, 12 input layer neurons, 12 hidden layer neurons, and 1 output layer neuron has been utilized on in sample data. In this method, there is no parametric model to represent the time series. The NN is applied for prediction of monthly rain fall in Sargodha and plot of the original and its prediction is given in Figure 2.

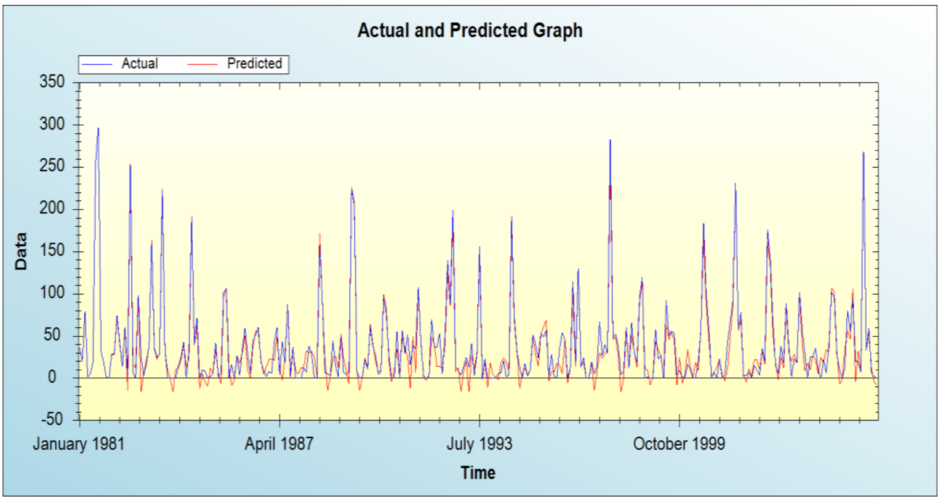


Figure 2: Prediction using NN

Autoregressive Integrated Moving Average (ARIMA)

For ARIMA modeling, the stationarity of data is basic requirement. In order to check stationarity of data, Augmented Dickey-Fuller (ADF) test has been utilized at level as well as at 1st difference.

Table 1
Augmented Dickey-Fuller Test

ADF	Statistic	Sig.
Level	-1.226179	0.2019
1st Difference	-17.36668	0.0000*

*Significant at 1%.

Results of ADF test show that time series (Rain Fall (RF_t)) has been stationary at 1st difference (see Table 1). Therefore, following I(1) ARIMA(p,d,q) has been utilized for further analysis:

$$D(RF_t) = A_t = \nabla^1 RF_t = (1 - B)^1 RF_t$$
$$A_t = \alpha_1 A_{t-1} + \cdots + \alpha_p A_{t-p} + Z_t + \cdots + \beta_q Z_{t-q} \quad (3.1)$$

where A_t and Z_t are 1st difference series of RF_t and error term respectively. P and q are parameters of autoregressive and moving average parts respectively. Correlogram of time series at 1st difference has been generated for quantifying the parameters p and q of ARIMA model (see Figure 3).

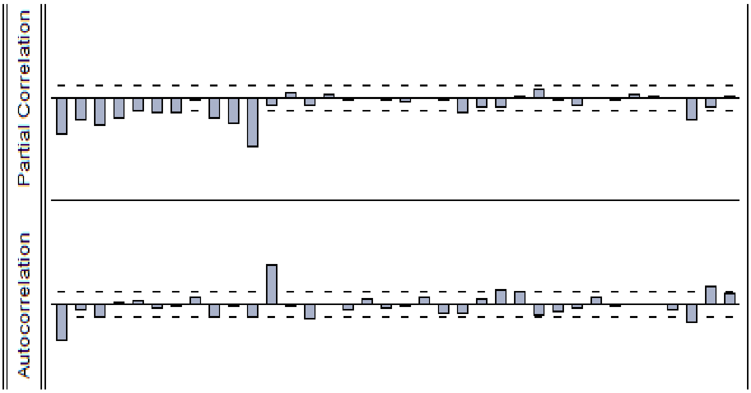


Figure 3: Correlogram at 1st Difference

By observing above correlogram, following models have been suggested and compared on the basis of Schwarz Criterion (SC) (see Table 2).

Table 2
Suggested I(1) ARIMA Models and their SC

Model	SC	Model	SC
1,1,0	11.184	1,1,11	10.971
0,1,1	10.867	12,1,1	10.695*
1,1,1	10.846	12,1,2	10.713**
1,1,2	10.863	12,1,3	10.733
1,1,3	10.878	12,1,4	10.752
1,1,4	10.860	12,1,9	10.733
1,1,9	10.933	12,1,10	10.727
1,1,10	10.961	12,1,11	10.721

*lowest SC value, **second lowest SC value.

Two models having least value of SC have been opted for further analysis, and their prediction errors have also been measured (see Table 3).

Table 3
Prediction errors ARIMA (12,1,1) and ARIMA (12,1,2)

	RMSE	MAE	MSE
12,1,1	65.205	41.605	4251.692
12,1,2	65.776	42.061	4326.482

Results of prediction errors have been suggested that, ARIMA (12,1,1) has less error in predicting the monthly rain fall as compare to ARIMA (12,1,2). Therefore, ARIMA (12,1,1) has been utilized for predicting the monthly rain fall in District Sargodha. Figure 4 shows the pattern of rain fall forecasted (RFF) values from ARIMA (12,1,1).

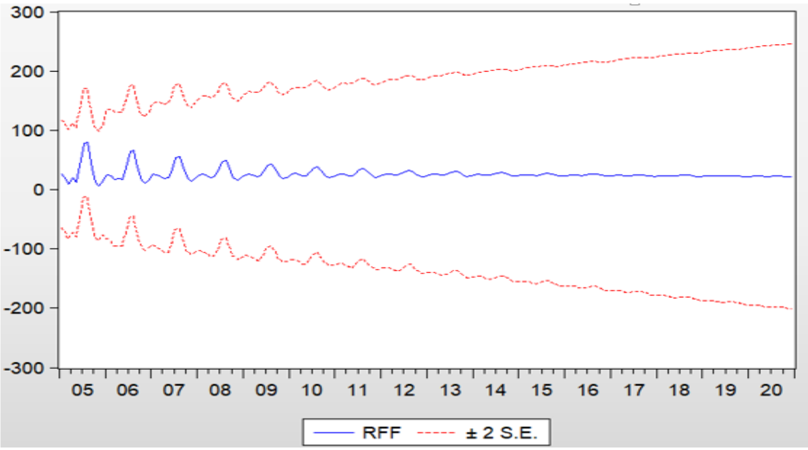


Figure 4: Rain Fall Forecasted (RFF) by ARIMA(12,1,1)

Fourier Series Models

Fourier series has been utilized to transform time series domain into frequency domain series for determining the importance of each frequency in the original series. The general Fourier series model is written as:

$$RF_t = T_t + \sum_{i=1}^{k}(\alpha_i cos_i wt + \beta_i sin_i wt) + \varepsilon_t \quad (3.2)$$

where T_t is trend component, k = n/2 (in case of even number) and k = (n-1)/2 (in case of odd number), w is angular frequency (= 2πf), α_i and β_i are coefficients, and ε_t is error term. The coefficients will have been calculated as:

$$\alpha_i = \frac{2}{N}\sum_{t=1}^{N}\left(RF_t - \hat{T}_t\right) cos w_i t \quad (3.3)$$

$$\beta_i = \frac{2}{N}\sum_{t=1}^{N}\left(RF_t - \hat{T}_t\right) sin w_i t \quad (3.4)$$

Estimation of Trend Component

The estimated trend equation for rain fall in Sargodha is given as:

$$T_t = 39.89 - 0.007146t \tag{3.5}$$

The adjusted R^2 of trend model has been equal to 0.9989, which shows that following model is 99.89% good fitted.

Periodogram Analysis

Periodogram has been utilized to partition the total variance of a rain fall time series into component variances because period in cycles are not exactly known. The estimated frequencies (freq), periods (p), and their intensities (pdg) are given in Table 4. Results showed that, period and frequency corresponding to largest intensity are n = 12 and f = 0.08333. Here, frequency is just the inverse of period. Thus the period of cycle or long term rain fall cycle is 12 months.

Table 4
Frequencies, Periods and Intensities

freq	p	pdg	freq	p	pdg
0		0	0.17	5.88235	11.07116
0.00333	300	4922.288	0.17333	5.76923	319.1614
0.00667	150	3896.887	0.17667	5.66038	1633.156
0.01	100	7166.023	0.23333	4.28571	1109.504
0.01333	75	839.0844	0.23667	4.22535	10264.07
0.01667	60	4846.107	0.24	4.16667	10497.77
0.06667	15	1900.004	0.24333	4.10959	3232.038
0.07	14.28571	5807.714	0.24667	4.05405	5408.023
0.07333	13.63636	1979.215	0.25	4	60679.04
0.07667	13.04348	295.0308	0.25333	3.94737	5501.59
0.08	12.5	2037.916	0.25667	3.8961	146.4201
0.08333	**12**	**164682.6**	0.26	3.84615	1957.431
0.08667	11.53846	2122.998	0.29667	3.37079	291.5723
0.09	11.11111	433.435	0.39333	2.54237	7551.971
0.09333	10.71429	2396.082	0.39667	2.52101	12374.76
0.09667	10.34483	2186.025	0.4	2.5	6835.5
0.1	10	76.47541	0.40333	2.47934	3861.078
0.12667	7.89474	5720.855	0.40667	2.45902	18749.08
0.13	7.69231	59.64194	0.41	2.43902	5716.999
0.13333	7.5	4866.835	0.41333	2.41935	128.8946
0.13667	7.31707	2413.819	0.48333	2.06897	4434.535
0.14	7.14286	109.9116	0.48667	2.05479	14575.99
0.14333	6.97674	1457.191	0.49	2.04082	5154.252
0.16	6.25	12088.64	0.49333	2.02703	10517.58
0.16333	6.12245	5145.393	0.49667	2.01342	729.2228
0.16667	6	139143.5	0.5	2	839.2203

Estimation of Seasonal Component

As the number of observations is even therefore,

$$k = \frac{n}{2} = \frac{12}{2} = 6$$

and

$$w = 2\pi f = 2 * 3.14 * 0.08333 = 0.523$$

The estimated parameters of seasonal component indicates that at 5% level of significance coswt, sinwt, sin2wt, cos3wt, sin3wt, cos4wt are significant (see Table 5).

Table 5
Parameter Estimation in Seasonal Component

Predictor	Coef	SE	T	P
coswt	-29.363	3.31	-8.87	0
sinwt	-15.162	3.308	-4.58	0
cos2wt	-2.272	3.31	-0.69	0.493
sin2wt	30.038	3.308	9.08	0
cos3wt	12.354	3.31	3.73	0
sin3wt	-15.474	3.309	-4.68	0
cos4wt	-8.405	3.309	-2.54	0.012
sin4wt	3.937	3.31	1.19	0.235
cos5wt	3.183	3.307	0.96	0.337
sin5wt	-0.677	3.313	-0.2	0.838
cos6wt	6.096	4.652	1.31	0.191
sin6wt	27.61	15.3	1.8	0.072

Therefore, the estimated seasonal model is given as:

$$\begin{aligned} RF_t = &-29.363coswt - 15.162sinwt + 30.038sin2wt \\ &+12.354cos3wt - 15.474sin3wt - 8.405cos4wt \end{aligned} \tag{3.6}$$

General Fourier series Model

The general model which consist on estimated trend (3.5) and estimated seasonal (3.6) components maybe written as:

$$\begin{aligned} RF_t = &\ 39.89 - 0.007146t - 29.363coswt \\ &-15.162sinwt + 30.038sin2wt + 12.354cos3wt \\ &-15.474sin3wt - 8.405cos4wt \end{aligned} \tag{3.7}$$

Comparison

By utilizing above three methodologies, rain fall in district Sargodha has been predicted for out of sample period ranging from January 2006 to December 2017. For comparing these methodologies, prediction errors have also been measured (see Table 6).

Table 6
Prediction Errors

Model	RMSE	MAE	MSE
NN	8.805	6.047	77.518
Fourier Model	45.321	28.975	2053.991
ARIMA (12,1,1)	65.205	41.605	4251.692

Results in Table 6 revealed that, neural networking with back propagation bipolar sigmoid function has least prediction error for rain fall in Sargodha. Where, Fourier series model stand at position second among three methodologies.

4. CONCLUSION

The study has been carried out for observing the pattern of Rainfall for last 37 years in Sargodha and to compare different Time Series methodologies for this purpose. Monthly data have been obtained from Pakistan Meteorological Department for the period of January 1981 to December 2017. Further, data have been divided into two mutually exclusive parts ranging from January 1981 to December 2005 and January 2006 to December 2017. Time plot of rain fall depict that there are too many oscillations in the series. Neural Networking (NN), Autoregressive Integrated Moving Average (ARIMA), and Fourier Series Models have been utilized on in sample observations (January 1981 to December 2005) for predicting the rain fall of out of sample observations (January 2006 to December 2017) in district Sargodha, Pakistan. On the basis of prediction errors (Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE)), proposed NN model found to be the best among all. Therefore, we recommend NN methodology to Pakistan Meteorological Department for prediction of rain fall in future. As the accurate prediction will help the department to make precise policies.

REFERENCES

1. Arslan, M., Mehmood, R., Riaz, M. and Baqar, M. (2013). Monitoring of changes in climate during blocking event by using run rule based quality control charts. In the *International Conference on Regional Climate-CORDEX*, Brussels, Belgium.
2. Bremaud, P.J. and Pointin, Y.B. (1993). Forecasting heavy rainfall from rain cell motion using radar data. *Journal of Hydrology*, 142, 373-389.
3. Burlando, P., Rosso, R., Cadavid, L.G. and Salas, J.D. (1993). Forecasting of short-term rainfall using ARMA models. *Journal of Hydrology*, 144(1-4), 193-211.
4. Chattopadhyay, S. and Chattopadhyay, G. (2008). Identification of the best hidden layer size for three-layered neural net in predicting monsoon rainfall in India. *Journal of Hydroinformatics*, 10(2), 181-188.
5. Chattopadhyay, S. and Chattopadhyay, M. (2007). A Soft Computing Technique in rainfall forecasting. *Proceedings of International Conference on IT*, HIT, India (pp. 523-526). arXiv preprint nlin/0703042.
6. Chen, H., Sun, J. and Wang, H. (2012). A Statistical Downscaling Model for Forecasting Summer Rainfall in China from DEMETER Hindcast Datasets. *American Meterological Society*, 27(3), 608-628.

7. Christodoulou, C.I., Michaelides, S.C., Gabella, M. and Pattichis, C.S. (2004). Prediction of rainfall rate based on weather radar measurements. In *2004 IEEE International Joint Conference on Neural Networks*, 2, 1393-1396.
8. French, M.N., Krajewski, W.F. and Cuykendall, R.R. (1992). Rainfall Forecasting in Space and Time using a Neural Network. *Journal of Hydrology*, 137(1-4), 1-31.
9. Krishnankutty, N. (2006). Long-range monsoon rainfall prediction of 2005 for the districts and sub-division Kerala with artificial neural network. Current Science, 90(6), 773-779.
10. Reddy, M.J. and Maity, R. (2007). Regional rainfall forecasting using large scale climate teleconnections and artificial intelligence techniques. *Journal of Intelligent Systems*, 16(4), 307-322.
11. Kumarasiri, A.D. and Sonnadara, D.U.J. (2006). Rainfall Forecasting: An Artificial Neural Network Approach. *Proceedings of the Technical Sessions, Institute of Physics – Sri Lanka*, 22, 1-16.
12. Chen, L., Li, Y. and Cheng, Z. (2010). An overview of research and forecasting on rainfall associated with landfalling tropical cyclones. *Advances in Atmospheric Sciences*, 27(5), 967-976.
13. Nayak, D.R., Mahapatra, A. and Mishra, P. (2013). A Survey on Rainfall Prediction using Artificial Neural Network. *International Journal of Computer Applications*, 72(16), 32-40.
14. Takasao, T. and Shiiba, M. (1984). Development of Techniques for On-Line Forecasting of Rainfall and Flood Runoff. *Natural Disaster Science*, 6(2), 83-112.
15. Toth, E., Brath, A. & Montanari, A. (2000). Comparison of short-term rainfall prediction models for real-time flood forecasting. *Journal of Hydrology*, 239, 132-147.
16. Wong, K.W., Wong, P.M., Gedeon, T.D. and Fung, C.C. (2003). Rainfall prediction model using soft computing technique. *Soft Computing*, 7(6), 434-438.
17. Zhang, M. and Scofield, R.A. (1994). Artificial neural network techniques for estimating heavy convective rainfall and recognizing cloud mergers from satellite data. *International Journal of Remote Sensing*, 15(16), 3241-3261.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 223-228

A STUDY OF RELATIONSHIP BETWEEN PHYSICAL FACILITIES IN CLASSROOM AND STUDENT'S PERFORMANCE: A SECONDARY LEVEL STUDY

Farah Kausar
University of Education, Lahore, Pakistan
Email: farahkausar222@gmail.com

ABSTRACT

The study was conducted to know the relationship of physical classroom facilities and student's performance at secondary level. The main objective of the study was to investigate the relationship of physical facilities and student's performance. The study was descriptive in nature and survey in type Questionnaire was designed, for the students to get their opinion about the relationship of physical classroom facilities and students performance. The sample of the study was 200 students of public and private school at secondary level. These were drawn by using convenient sampling technique. The result of the questionnaire has been analyzed by SPSS percentage and it was presented in the form of tables which was followed by interpretation. Study showed that the classroom physical facilities play a vital role in student's performance. It was concluded that providing physical facilities in classroom can enchant a learner in a best way and cover the gap of the study. There was highly significant positive relationship between physical facilities in classroom and student's performance. It was recommended that classroom physical facilities should be well structured, managed, equipped and facilitated during teaching learning process.

KEY WORDS

Relationship, Physical classroom facilities, Student's performance, Secondary level.

INTRODUCTION

Class room environment is a major aspect in the teaching and learning process and is crucial to determine student's performance. Physical classroom environment stimulates student's engagement in the learning process and influences their behavior. All classrooms have an environment or climate that could play a serious role in the success of the students. Real learning does not happen without an environment that fosters that possibility. A good classroom environment always demands such strategies, methods and measures which should facilitate productive work in teaching learning Process.

Various physical dimensions of classroom environment are such as seating position, classroom furnishing arrangement, spatial density, privacy, noise and acoustics, climate and thermal control and window classroom. Research studies on the classroom environment have related that physical arrangement plays a vital role in teaching learning process. It can affect the performance of both teachers and students (Savage, 1999; Stewart and Evans, 1997).

The physical classroom learning setting is comprised of different things i.e. size of the room, lighting, degree of temperature, condition of classroom environment refers to the physical room in which teacher and learners are the main element including its spatial elements i.e., floor, windows, walls as well as other classroom equipments i.e., desks, chairs, rugs, chalkboards, tack boards, counters and computer equipment but not limited to these things (Fisher, 2008).

Physical environment plays a central role in any activity and makes it more conductive, successful and achievable. According to Oni (1992) and Hallak (1990), physical facilities compose a strategic factor in the operation and functioning of an organization as they determine the excellent performance of any social organization or system including education.

Taylor and Vlastos (2009) found the relationship between environment and design within the classroom from a theoretical perspective. They found that physical environment of the classroom acts as ‘silent curriculum’. It means that classroom environment design can facilitate and improve the learning process like the overt curriculum. Unfortunately, in Pakistan, educational institutions lack of physical facilities which results in malfunctioning of these institutions. Poor and inadequate facilities affect the overall performance of the institutions.

Sufficient facilities promote academic performance and ensure to strengthen the overall institutional achievement. While unattractive and old school buildings; cracked classroom walls and floors; lack of toilets; lack of desks and benches; lack of transport facility; lack of proper security system; lack of drinking water; lack of power supply; lack of educational technology; lack of first aids facility etc. negatively affect academic achievement of the institutions. Therefore, it is right to say that academic achievement has a close link with the availability of educational facilities (Hussain, et al. 2012).

Poor classroom acoustics can adversely affect learning environment for many students. Constant noise exposure can damage cognitive performance and functioning (Higgins et al, 2004).

Physical environment can affect student’s comfort and also their ability to learn to some extent. Students who are comfortable are likely to get much information as compared to those who are uncomfortable. Besides, the physical atmosphere can also affect the morale of the learners. Unfavorable classroom environment can discourage the learners and they become less willing to learn (www.enotes.com).

OBJECTIVE OF THE STUDY

The main objective of the study was to investigate the relationship of physical facilities and student’s performance.

RESEARCH QUESTIONS

1. To find out the relationship of physical classroom facilities and student performance.
2. To find out the students perception about physical classroom facilities and student performance.

RESEARCH METHODOLOGY

The study was designed to find the relationship of physical classroom facilities and students performance at secondary level. A quantitative research was used in this study. Philosophical Paradigm of this research is positivism. The researchers used descriptive survey to measure the relationship of physical classroom facilities and students performance. Descriptive survey included to collect data from students of secondary school level. For the purpose of collection data questionnaire were developed and filled by students. The questionnaire was self developed with the help of literature review and experts. The population of the study was students of secondary school of 1 public and 1 private in district Lahore. The sample of the study was 200 students at secondary level. The researchers used convenient sampling in this research.

DATA COLLECTION PROCEDURE

Data was collected by administering the questionnaire. Questionnaires were administered by the researchers personally the response rate for the questionnaire was 100%.

DATA ANALYSIS PROCEDURE

Descriptive statistics is fundamental in organizing research data as it serves to summarize the information. Collected data were analyzed by applying Pearson’s correlations coefficients to see any relation between the variables under study. Data collected were categorized, coded and subjected to statistical package of social sciences for analysis. Data analysis was done using Quantitative technique. Quantitative data were analyzed using descriptive statistics including mean scores, and percentages that were presented using table. Qualitative data was analyzed thematically by arranging responses according to the research questions and objectives. Thereafter, similar responses were tallied and inferences, conclusion and recommendation were drawn.

FINDINGS AND CONCLUSIONS

The findings of the study were discussed according to the questions

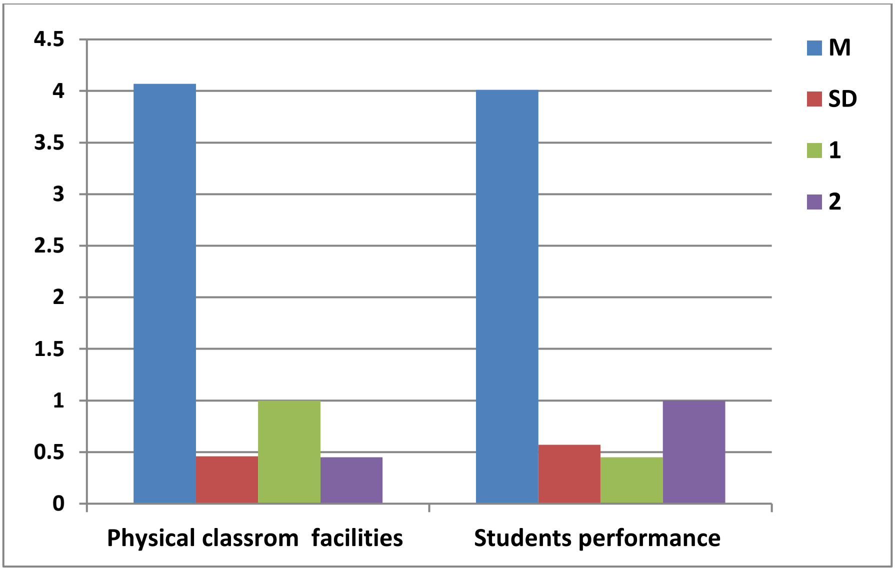
Research Question 1:

What is the relationship of physical classroom facilities and students performance?

Table 1
Pearson’s correlations between Physical Classroom Facilities and Students Performance

Variables	M	SD	1	2
Physical Classroom Facilities	4.07	.46	1	.45**
Student Performance	4.01	.57	.45**	1

**p<0.01, N=200.



The table 1 shows that is positive correlation between physical classroom facilities and students performance ($p < 0.01$, sig = .000), therefore, it has been proved that there is a correlation between physical classroom facilities and students performance. Analysis was carried out further in order to see the relationship between various characteristics of classroom environment and student performance.

Research Question 2:

What is the teachers and students perception about physical classroom facilities and student performance?

Table 2
Means of Students Opinion about Classroom Physical Facilities and Students Performance at Secondary Level

Statement	N	Strongly Agree	Agree	Undecided	Disagree	Strongly	Mean
Friendly atmosphere of the class room environment improve the learning.	200	66.0%	32.0%	2.0%	.0%	.0%	4.64
Small and ventilated room affects the student performance.	200	74.0%	20.0%	2.0%	4.0%	.0%	4.58
Bad systems of cleaning in classroom affect the students learning.	200	62.0%	30.0%	6.0%	2.0%	.0%	4.50
Uncomfortably furniture affects the classroom environment.	200	54.0%	36.0%	8.0%	2.0%	.0%	4.40
The unavailability of electricity affects the learning.	200	58.0%	30.0%	6.0%	6.0%	.0%	4.34
External noise causing distraction in the classroom.	200	49.0%	40.8%	6.1%	2.0%	2.0%	4.33
Seating arrangement that put pupils closer to their teacher during class motivate the learning process.	200	58.0%	32.0%	2.0%	4.0%	4.0%	4.36
The use of different activities, teaching resources and technology upgrade our performance.	200	51.0%	32.7%	8.2%	6.1%	2.0%	4.24

It is evident from the table 2 that students opinion about; friendly atmosphere of the classroom (4.64), small and ventilated room (4.58), bad system of cleaning (4.50), uncomfortable of furniture (4.40), and seating arrangement (4.36) show high mean score. Medium mean score on external noise (4.33), unavailability of electricity (4.34) and different activities of teaching (4.24). It is concluded that students strongly agree that physical facilities effect the student's performance.

CONCLUSIONS

Following conclusion was drawn from the findings of the study. The findings of the study revealed that there is a strong relationship between physical classroom facilities and student's performance. Small and unventilated room, poor sanitation system, uncomfortable furniture, non-availability of electricity, noisy classroom, congested sitting of students, effects on student learning and creates the unfavorable environment which is not conducive for students' learning. It is concluded that the use of different teaching methods and A.V aids, friendly atmosphere, Proper seating arrangement of classroom improve the student learning and make environment conducive for learning.

RECOMMENDATIONS

Based on findings and conclusions, the researchers made the following recommendations:

- As the study revealed that classroom physical classroom facilities plays a vital role in rise students' academic achievement therefore, it is strongly recommended that classroom physical should be well organized, managed ,and physical facilities should be provided to each school on main concern.
- This research may be expanded by enhancing population and sample. It may be conducted further another levels by using mixed methods of research.

REFERENCES

1. Hussain, I., Ahmed, M., Ahmad, S., Suleman, Q. and Khalid, N. (2012). A Study to Investigate the Availability of Educational Facilities at the Secondary School Level in District Karak. *Language in India*, 12(10), 234-25.
2. Stewart, S.C. and Evans, W.H. (1997) . Setting the stage for success: Assessing the instructional environment. *Preventing School Failure*, 41(2), 53-56.
3. Stewart, S.C., Evans, W.H. and Kaczynski, D.J. (1997). Setting the stage for success: Assessing the instructional environment. Preventing School Failure. *Alternative Education for Children and Youth*, 41(2), 53-56.
4. Savage, T.V. (1999). *Teaching self-control through management and discipline*. Boston: Allynand Bacon.
5. Oni, J.O. (1992). *Resource and Resource Utilization as Correlates of School Academic Performance*. Ph.D. Thesis, University of Ibadan.
6. Fisher, E.S. (2008). *The Effect of the Physical Classroom Environment on Literacy Outcomes: How 3rd Class Teachers use the Physical Classroom to Implement a Balanced Literacy Curriculum*. A Thesis presented to the Faculty of the Graduate School University of Missouri.
7. Laslett, R. and Smith, C. (1984). *Effective Classroom Management: A Teacher's Guide*, Croon Helm Ltd, London, UK.
8. Basit, A. (2005). *Classroom Management Techniques at Secondary Level and Developing a Model for Urban Schools for District Peshawar*. M.Phil. Thesis, Faculty of Education, Allama Iqbal Open University Islamabad. pp. 16-17.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 229-241

ON THE BAYESIAN ESTIMATION FOR TWO COMPONENT MIXTURE OF REFLECTED WEIBULL DISTRIBUTION, ASSUMING TYPE I CENSORED DATA

Wajiha Nasir[1] and **Muhammad Zubair**[2]

[1] Department of Statistics, Govt. College Women University Sialkot, Pakistan. Email: wajiha.nasir@gcwus.edu.pk

[2] Department of Statistics, University of Sargodha, Sargodha, Pakistan. Email: zubair.bakhsh@uos.edu.pk

ABSTRACT

In this paper, two component mixture model of reflected Weibull distribution has been studied under paradigm of Bayesian analysis. Posterior distribution has been estimated by using informative priors. Prior predictive distributions have been derived for elicitation of hyperparameters. Bayes estimators and their corresponding risks has been estimated by using different loss functions. Performance of Bayes estimators have been determined by Monte-Carlo Simulation as well as by real life example.

1. INTRODUCTION

Weibull distribution is well known for using in life testing and reliability experiments. When the Weibull distribution is reflected on its axis, then a new form of distribution is obtained known as reflected Weibull distribution. When a random variable X, is said to have Weibull distribution then if $Y=-X$ is said to have reflected Weibull distribution by Cohen (1973). This paper deals with the estimation of mixture model of reflected Weibull distribution when scale parameter is unknown with known shape. Danish and Aslam (2012) has studied the randomly censored Weibull distribution by Bayesian inference. Ng et al. (2012) has also studied three parameter of Weibull distribution based on progressively type-II censored samples. Singh et al. (2013) has studied inverse Weibull distribution under type-II hybrid sampling via Bayesian approach. Noor and Aslam (2013) has studied mixture model of inverse weibull distribution under type-I censoring. Almaki and Nadrarajah (2014) has presented different form of Weibull distribution including reflected Weibull distribution.

In this paper, mixture model of reflected Weibull distribution has been studied using type-I censoring scheme. Mixture model of reflected Weibull distribution has not been studied earlier by using Bayesian analysis.

A random variable Y has said to be distributed as reflected Weibull distribution having following pdf

$$f(y)=\alpha\beta(-y)^{\beta-1}e^{-\alpha(-y)\beta},\ -\infty<y<0,$$

where $\alpha,\beta>0$.

2. POPULATION AND MODEL

Distribution function with characterization of convex combination of other specific probability distribution function. A mixture may be contain of a finite number of base elements, where usually a relatively small number of individual distributions are combined together, or an infinite number of base elements. Two component mixture model for Frechet distribution having two parameters and with mixing weight is

$$F(y) = \omega F_1(y) + (1-\omega) F_2(y) \quad 0 < \omega < 1, \tag{1}$$

where $F(y_i) = e^{-\alpha(-y)^{\beta}}$. Now eq. 1 can be written as

$$F(y) = \omega e^{-\alpha(-y)^{\beta}} + (1-\omega) e^{-\alpha(-y)^{\beta}} \tag{2}$$

Now, the mixture density will be

$$f(y) = \omega f_1(y) + (1-\omega) f_2(y) \tag{3}$$

The reflected Weibull model is

$$f_i(y) = \alpha_i \beta (-y)^{\beta-1} e^{-\alpha_i(-y)^{\beta}}, i = 1,2, \ -\infty < y < 0 \tag{4}$$

where $\alpha, \beta > 0$.

Now, Assume that a random sample of size n is selected for reliability of units i.e. $(y_1, y_2, ..., y_n)$.Let we assume r units in sample $[0, t_0]$ in which termination time is fixed which is t_0 and $n-r$ the remaining samples are survive for the rest when the termination time is over. Some situation is taken for two sub population model i.e. from n_i units have lifetime in the interval and $n_i - r_i$ have survived time in which $r = r_1 + r_2$ are uncensored units. Suppose x_{ij} be the failure time of j_{th} unit which is associated with i_{th} sub population i.e. $i = 1, 2$, $j = 1, 2, ..., r_i$, $0 < y_{1j}, y_{2j} \le t_0$. The likelihood for above stated situation is

$$\begin{aligned} L(\alpha_1, \alpha_2, \omega) = & \prod_{j=1}^{r_1} \omega \alpha_1 \beta (-y_j)^{\beta-1} e^{-\alpha_1(-y_{2j})^{\beta}} \\ & \prod_{j=1}^{r_2} (1-\omega) \alpha_2 \beta (-y_j)^{\beta-1} e^{-\alpha_2(-y_{2j})^{\beta}} \left[F(T) \right]^{n-r} \end{aligned} \tag{5}$$

where $F(T) = 1 - F(T)$ and known as survival function and x=$[x_{1i}, x_{2i}]$

$$\begin{aligned} L(\alpha_1, \alpha_2, \omega) = & \prod_{j=1}^{r_1} \omega \alpha_1 \beta (-y_j)^{\beta-1} e^{-\alpha_1(-y_j)^{\beta}} \prod_{j=1}^{r_2} (1-\omega) \alpha_2 \beta (-y_j)^{\beta-1} e^{-\alpha_2(-y_j)^{\beta}} \\ & \left[1 - \left\{ \omega e^{-\alpha_1(-t_0)^{\beta}} + (1-\omega) e^{-\alpha_2(-t_0)^{\beta}} \right\} \right]^{n-r} \end{aligned} \tag{6}$$

By simplifying eq.6 we get

$$L(\alpha_1,\alpha_2,\omega)=\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^l\,\omega^{r_1+m-l}(1-\omega)^{r_2-m}$$
$$\exp\left\{-\alpha_1\left(\sum_{i=1}^{r_1}(-y_i)^\beta+(-t_0)^\beta\right)\right\}\exp\left\{-\alpha_2\left(\sum_{i=1}^{r_2}(-y_i)^\beta+(-t_0)^\beta\right)\right\} \tag{7}$$

3. POSTERIOR DISTRIBUTION

In this section, posterior distribution for mixture model have been derived by using informative priors (Exponential and Inverted Chi Square priors).

Suppose $\alpha_1 \sim \text{Chi Square}(v_1), \alpha_1 \sim \text{Chi Square}(v_2)$ and $w \sim U(0,1)$. We assume that joint distribution of prior is independent, so joint prior will become

$$h(\alpha_1,\alpha_1,w)\propto \alpha_1^{\frac{v_1}{2}-1}\alpha_1^{\frac{v_1}{2}-1}e^{-\left(\frac{\alpha_1+\alpha_2}{2}\right)} \tag{8}$$

Now, the posterior distribution using eq. 7 and eq. 8 is

$$p(\alpha_1,\alpha_2,\omega)\propto\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^l B(r_1+m-l+1,r_2-m+1)$$
$$\times\alpha_1^{r_1+\frac{v_1}{2}-1}e^{-\alpha_1 G_1}\alpha_2^{r_2+\frac{v_2}{2}-1}e^{-\beta_2 G_2},\alpha_1,\alpha_2>0,0<\omega<1. \tag{9}$$

where $G_1=\sum_{i=1}^{r_1}(-y_i)^\beta+(-t_0)^\beta+\frac{1}{2}$ and $\text{G}_2=\sum_{i=1}^{r_2}(-y_i)^\beta+(-t_0)^\beta+\frac{1}{2}$.

Suppose $\alpha_1 \sim \text{Exponential}(\lambda_1), \alpha_2 \sim \text{Exponential}(\lambda_2)$ and $w \sim U(0,1)$. Under the assumption of joint independence the prior distribution will be

$$h(\alpha_1,\alpha_1,w)\propto e^{-(\alpha_1\lambda_1+\alpha_2\lambda_2)} \tag{10}$$

Now, the posterior distribution using eq. 10 is

$$p(\alpha_1,\alpha_2,\omega)\propto\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^l B(r_1+m-l+1,r_2-m+1)$$
$$\alpha_1^{r_1-1}e^{-\alpha_1 K_1}\alpha_2^{r_2-1}e^{-\beta_2 K_2},\alpha_1,\alpha_2>0,0<\omega<1. \tag{11}$$

where $K_1=\sum_{i=1}^{r_1}(-y_i)^\beta+(-t_0)^\beta+\lambda_1$ and $K_2=\sum_{i=1}^{r_2}(-y_i)^\beta+(-t_0)^\beta+\lambda_2$.

Now Bayes estimators and their posterior risks will be estimated by different loss functions. We use square error loss function (SELF) and precautionary loss function (PLF) for deriving BEs and their cross ponding PRs.

Bayes estimators and posterior risk for α_1, α_2 and w using SELF for inverted chi-square are as follows

$$\hat{\alpha}_{1,SELF} = \frac{1}{C_1} \sum_{m=0}^{n-r} \sum_{l=0}^{m} \binom{n-r}{m} \binom{m}{l} (-1)^j B(r_1 + m - l + 1, r_2 - m + 1) \frac{\Gamma\left(r_1 + \frac{v_1}{2} + 2\right)}{G_1^{r_1 + \frac{v_1}{2} + 2}} \frac{\Gamma\left(r_2 + \frac{v_2}{2}\right)}{G_2^{r_2 + \frac{v_2}{2}}} \tag{12}$$

$$\rho\left(\hat{\alpha}_{1,SELF}\right) = \frac{1}{C_1} \sum_{m=0}^{n-r} \sum_{l=0}^{m} \binom{n-r}{m} \binom{m}{l} (-1)^j B(r_1 + m - l + 1, r_2 - m + 1) \frac{\Gamma\left(r_1 + \frac{v_1}{2} + 3\right)}{G_1^{r_1 + \frac{v_1}{2} + 3}} \frac{\Gamma\left(r_2 + \frac{v_2}{2}\right)}{G_2^{r_2 + \frac{v_2}{2}}} - \left(\hat{\alpha}_{1,SELF}\right)^2 \tag{13}$$

$$\hat{\alpha}_{2,SELF} = \frac{1}{C_1} \sum_{m=0}^{n-r} \sum_{l=0}^{m} \binom{n-r}{m} \binom{m}{l} (-1)^j B(r_1 + m - l + 1, r_2 - m + 1) \frac{\Gamma\left(r_1 + \frac{v_1}{2}\right)}{G_1^{r_1 + \frac{v_1}{2}}} \frac{\Gamma\left(r_2 + \frac{v_2}{2} + 2\right)}{G_2^{r_2 + \frac{v_2}{2} + 2}} \tag{14}$$

$$\rho\left(\hat{\alpha}_{2,SELF}\right) = \frac{1}{C_1} \sum_{m=0}^{n-r} \sum_{l=0}^{m} \binom{n-r}{m} \binom{m}{l} (-1)^j B(r_1 + m - l + 1, r_2 - m + 1) \frac{\Gamma\left(r_1 + \frac{v_1}{2}\right)}{G_1^{r_1 + \frac{v_1}{2}}} \frac{\Gamma\left(r_2 + \frac{v_2}{2} + 3\right)}{G_2^{r_2 + \frac{v_2}{2} + 3}} - \left(\hat{\alpha}_{2,SELF}\right)^2 \tag{15}$$

$$\hat{w}_{,SELF} = \frac{1}{C_1} \sum_{m=0}^{n-r} \sum_{l=0}^{m} \binom{n-r}{m} \binom{m}{l} (-1)^j B(r_1 + m - l + 2, r_2 - m + 2) \frac{\Gamma\left(r_1 + \frac{v_1}{2}\right)}{G_1^{r_1 + \frac{v_1}{2}}} \frac{\Gamma\left(r_2 + \frac{v_2}{2}\right)}{G_2^{r_2 + \frac{v_2}{2}}} \tag{16}$$

$$\rho\left(w_{SELF}\right)=\frac{1}{C_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^j\;B\left(r_1+m-l+3,r_2-m+3\right)\frac{\Gamma\left(r_1+\frac{v_1}{2}\right)}{G_1^{r_1+\frac{v_1}{2}}}\frac{\Gamma\left(r_2+\frac{v_2}{2}\right)}{G_2^{r_2+\frac{v_2}{2}}}-\left(\hat{w}_{SELF}\right)^2 \quad (17)$$

Bayes estimators and posterior risk for α_1,α_2 and w using PLF for inverted chi-square are as follows

$$\hat{\alpha}_{1,SELF}=\sqrt{\frac{1}{C_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^j\;B\left(r_1+m-l+1,r_2-m+1\right)\frac{\Gamma\left(r_1+\frac{v_1}{2}+3\right)}{G_1^{r_1+\frac{v_1}{2}+3}}\frac{\Gamma\left(r_2+\frac{v_2}{2}\right)}{G_2^{r_2+\frac{v_2}{2}}}} \quad (18)$$

$$\rho\left(\hat{\alpha}_{1,SELF}\right)=2*\left(\sqrt{\frac{1}{C_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^j\;B\left(r_1+m-l+1,r_2-m+1\right)\frac{\Gamma\left(r_1+\frac{v_1}{2}+3\right)}{G_1^{r_1+\frac{v_1}{2}+3}}\frac{\Gamma\left(r_2+\frac{v_2}{2}\right)}{G_2^{r_2+\frac{v_2}{2}}}}-\hat{\alpha}_{1,SELF}\right) \quad (19)$$

$$\hat{\alpha}_{2,SELF}=\sqrt{\frac{1}{C_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^j\;B\left(r_1+m-l+1,r_2-m+1\right)\frac{\Gamma\left(r_1+\frac{v_1}{2}\right)}{G_1^{r_1+\frac{v_1}{2}}}\frac{\Gamma\left(r_2+\frac{v_2}{2}+3\right)}{G_2^{r_2+\frac{v_2}{2}+3}}} \quad (20)$$

$$\rho\left(\hat{\alpha}_{2,SELF}\right)=2*\left(\sqrt{\frac{\frac{1}{C_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^j B\left(r_1+m-l+1,r_2-m+1\right)}{\frac{\Gamma\left(r_1+\frac{v_1}{2}\right)}{G_1^{r_1+\frac{v_1}{2}}}\frac{\Gamma\left(r_2+\frac{v_2}{2}+3\right)}{G_2^{r_2+\frac{v_2}{2}+3}}}}-\hat{\alpha}_{2,SELF}\right) \tag{21}$$

$$\hat{w}_{SELF}=\sqrt{\frac{1}{C_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^j B\left(r_1+m-l+3,r_2-m+1\right)\frac{\Gamma\left(r_1+\frac{v_1}{2}\right)}{G_1^{r_1+\frac{v_1}{2}}}\frac{\Gamma\left(r_2+\frac{v_2}{2}\right)}{G_2^{r_2+\frac{v_2}{2}}}} \tag{22}$$

$$\rho\left(w_{SELF}\right)=2*\left(\sqrt{\frac{\frac{1}{C_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^j B\left(r_1+m-l+3,r_2-m+1\right)}{\frac{\Gamma\left(r_1+\frac{v_1}{2}\right)}{G_1^{r_1+\frac{v_1}{2}}}\frac{\Gamma\left(r_2+\frac{v_2}{2}\right)}{G_2^{r_2+\frac{v_2}{2}}}}}-\hat{w}_{SELF}\right) \tag{23}$$

where $C_1=\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^j B\left(r_1+m-l+1,r_2-m+1\right)\frac{\Gamma\left(r_1+\frac{v_1}{2}\right)}{G_1^{r_1+\frac{v_1}{2}}}\frac{\Gamma\left(r_2+\frac{v_2}{2}\right)}{G_2^{r_2+\frac{v_2}{2}}}$.

Bayes estimators and posterior risk for α_1,α_2 and w using PLF for exponential prior are as follows

$$\hat{\alpha}_{1,PLF}=\frac{1}{R_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^j B\left(r_1+m-l,r_2-m\right)\frac{\Gamma\left(r_1+1\right)}{K_1^{r_1+1}}\frac{\Gamma\left(r_2\right)}{K_2^{r_2}} \tag{24}$$

$$\rho\left(\hat{\alpha}_{1,PLF}\right)=\frac{1}{R_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^j \\ B\left(r_1+m-l,r_2-m\right)\frac{\Gamma\left(r_1+2\right)}{K_1^{r_1+2}}\frac{\Gamma\left(r_2\right)}{K_2^{r_2}}-\left(\hat{\alpha}_{1,SELF}\right)^2 \tag{25}$$

$$\hat{\alpha}_{2,PLF}=\frac{1}{R_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^j B\left(r_1+m-l,r_2-m\right)\frac{\Gamma\left(r_1\right)}{K_1^{r_1}}\frac{\Gamma\left(r_2+1\right)}{K_2^{r_2+1}} \tag{26}$$

$$\rho\left(\hat{\alpha}_{2,PLF}\right)=\frac{1}{R_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^j \\ B\left(r_1+m-l,r_2-m\right)\frac{\Gamma\left(r_1\right)}{K_1^{r_1}}\frac{\Gamma\left(r_2+2\right)}{K_2^{r_2+2}}-\left(\hat{\alpha}_{2,SELF}\right)^2 \tag{27}$$

$$\hat{w}_{PLF}=\frac{1}{R_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^j B\left(r_1+m-l+1,r_2-m+1\right)\frac{\Gamma\left(r_1\right)}{K_1^{r_1}}\frac{\Gamma\left(r_2\right)}{K_2^{r_2}} \tag{28}$$

$$\rho\left(w_{PLF}\right)=\frac{1}{R_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^j \\ B\left(r_1+m-l+2,r_2-m+2\right)\frac{\Gamma\left(r_1\right)}{K_1^{r_1}}\frac{\Gamma\left(r_2\right)}{K_2^{r_2}}-\left(\hat{w}_{SELF}\right)^2 \tag{29}$$

Bayes estimators and posterior risk for α_1,α_2 and w using PLF for inverted chi-square are as follows

$$\hat{\alpha}_{1,SELF}=\sqrt{\frac{1}{R_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^j B\left(r_1+m-l,r_2-m\right)\frac{\Gamma\left(r_1+2\right)}{K_1^{r_1+2}}\frac{\Gamma\left(r_2\right)}{K_2^{r_2}}}$$

(30)

$$\rho\left(\hat{\alpha}_{1,PLF}\right)=2*\left(\sqrt{\frac{1}{R_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^j B\left(r_1+m-l,r_2-m\right)\frac{\Gamma\left(r_1\right)}{K_1^{r_1}}\frac{\Gamma\left(r_2+2\right)}{K_2^{r_2+2}}}-\hat{\alpha}_{1,SELF}\right)$$

(31)

$$\hat{\alpha}_{2,PLF}=\sqrt{\frac{1}{R_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^j B\left(r_1+m-l,r_2-m\right)\frac{\Gamma\left(r_1\right)}{K_1^{r_1}}\frac{\Gamma\left(r_2+2\right)}{K_2^{r_2+2}}} \tag{32}$$

$$\rho\left(\hat{\alpha}_{2,PLF}\right)=2*\left(\sqrt{\frac{1}{R_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^{j} B\left(r_1+m-l,r_2-m\right)\frac{\Gamma\left(r_1\right)}{K_1^{r_1}}\frac{\Gamma\left(r_2+2\right)}{K_2^{r_2+2}}}-\hat{\alpha}_{2,SELF}\right) \quad (33)$$

$$\hat{w}_{PLF}=\sqrt{\frac{1}{R_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^{j} B\left(r_1+m-l+1,r_2-m+1\right)\frac{\Gamma\left(r_1\right)}{G_1^{r_1}}\frac{\Gamma\left(r_2\right)}{G_2^{r_2}}} \quad (34)$$

$$\rho\left(w_{PLF}\right)=2*\left(\sqrt{\frac{1}{R_1}\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^{j} B\left(r_1+m-l+2,r_2-m+2\right)\frac{\Gamma\left(r_1\right)}{K_1^{r_1}}\frac{\Gamma\left(r_2\right)}{K_2^{r_2}}}-\hat{w}_{SELF}\right)$$

(35)

where $R_1=\sum_{m=0}^{n-r}\sum_{l=0}^{m}\binom{n-r}{m}\binom{m}{l}(-1)^{j} B\left(r_1+m-l+1,r_2-m+1\right)\frac{\Gamma\left(r_1\right)}{K_1^{r_1}}\frac{\Gamma\left(r_2\right)}{K_2^{r_2}}$.

4. SIMULATION STUDY

In this section, Simulation studied has been carried for evolution of prior distributions and loss function. For a sample size of $n\in\{50,100,200,500\}$ with $\alpha_1\in\{2,4\},\alpha_2=\{5,8\}$ and $p=\{0.20,0.50\}$ data has been generated by inverse transformation technique in such manner. Firstly, random numbers from $U(0,1)$ has been generated. The observation has been taken from first subpopulation if $u<p$ and the observation is selected from second subpopulation.

Table 1
Simulation for Exponential Prior when $p = 0.50, t = 40$

n	$\alpha_1 = 2, \alpha_2 = 5$		$\alpha_1 = 4, \alpha_2 = 5$		$\alpha_1 = 2, \alpha_2 = 7$		$\alpha_1 = 4, \alpha_2 = 7$	
	SELF	**PLF**	**SELF**	**PLF**	**SELF**	**PLF**	**SELF**	**PLF**
50	1.9475 (0.1574)	1.9872 (0.0772)	3.6578 (0.5532)	3.7388 (0.1452)	1.9556 (0.1586)	1.9872 (0.0772)	3.6677 (0.5560)	3.7320 (0.1449)
100	1.9770 (0.0797)	1.9922 (0.0393)	3.8358 (0.2996)	3.8678 (0.0762)	1.9750 (0.0795)	1.9941 (0.0393)	3.8279 (0.2984)	3.8582 (0.0760)
200	1.9885 (0.0399)	1.9937 (0.0197)	3.9127 (0.1546)	3.9364 (0.0390)	1.9875 (0.0398)	1.9973 (0.0197)	3.9126 (0.1545)	3.9324 (0.0390)
500	1.9940 (0.0160)	1.9972 (0.0080)	3.9651 (0.0631)	3.9678 (0.0158)	1.9943 (0.0160)	2.001 (0.0080)	3.9652 (0.0631)	3.9725 (0.0158)
	$\alpha_1 = 2, \alpha_2 = 5$		$\alpha_1 = 4, \alpha_2 = 5$		$\alpha_1 = 2, \alpha_2 = 7$		$\alpha_1 = 4, \alpha_2 = 7$	
50	4.4475 (0.8152)	4.5256 (0.1757)	4.4542 (0.8188)	4.5333 (0.1761)	5.8979 (1.4305)	6.0088 (0.2333)	5.8742 (1.4182)	5.9957 (0.2328)
100	4.7272 (0.4550)	4.7747 (0.0941)	4.8538 (0.2378)	4.7726 (0.0940)	6.4118 (0.8360)	6.4865 (0.1278)	6.4046 (0.8341)	6.4687 (0.1275)
200	4.8540 (0.2378)	4.8795 (0.0484)	4.8538 (0.2378)	4.8721 (0.0484)	6.7061 (0.4538)	6.7310 (0.0668)	6.6986 (0.4528)	6.7250 (0.0668)
0.	4.9431 (0.0981)	4.9540 (0.0198)	4.9454 (0.0982)	4.9557 (0.0198)	6.8712 (0.1895)	6.8878 (0.0274)	6.8762 (0.1898)	6.8772 (0.0274)
	$\alpha_1 = 2, \alpha_2 = 5$		$\alpha_1 = 4, \alpha_2 = 5$		$\alpha_1 = 2, \alpha_2 = 7$		$\alpha_1 = 5, \alpha_2 = 7$	
50	0.5048 (0.0098)	0.4327 (0.0049)	0.5048 (0.0098)	0.4327 (0.0049)	0.5048 (0.0098)	0.4327 (0.0049)	0.5048 (0.0098)	0.4327 (0.0049)
100	0.5025 (0.0049)	0.4648 (0.0024)	0.5025 (0.0049)	0.4648 (0.0024)	0.5025 (0.0049)	0.4648 (0.0024)	0.5025 (0.0049)	0.4648 (0.0024)
200	0.5012 (0.0024)	0.4848 (0.0012)	0.5012 (0.0024)	0.4848 (0.0012)	0.5012 (0.0024)	0.4848 (0.0012)	0.5012 (0.0024)	0.4848 (0.0012)
500	0.5004 (0.0009)	0.4948 (0.0005)	0.5004 (0.0009)	0.4948 (0.0005)	0.5004 (0.0009)	0.4948 (0.0005)	0.5004 (0.0009)	0.4948 (0.0005)

Table 2
Simulation for Inverted-Chi Square Prior when $p = 0.50, t = 40$

n	$\alpha_1 = 2, \alpha_2 = 5$		$\alpha_1 = 4, \alpha_2 = 5$		$\alpha_1 = 2, \alpha_2 = 7$		$\alpha_1 = 4, \alpha_2 = 7$	
	SELF	**PLF**	**SELF**	**PLF**	**SELF**	**PLF**	**SELF**	**PLF**
50	2.0304 (0.1689)	2.0610 (0.0788)	3.8918 (0.6179)	3.9735 (0.1520)	2.0237 (0.1676)	2.0681 (0.0791)	3.9065 (0.6230)	3.9757 (0.1521)
100	2.0142 (0.0821)	2.0324 (0.0397)	3.9515 (0.3157)	3.9801 (0.0778)	2.0183 (0.0824)	2.0315 (0.0397)	3.9517 (0.3157)	3.9831 (0.0778)
200	2.0044 (0.0404)	2.0184 (0.0199)	3.9742 (0.1589)	3.9961 (0.0395)	2.0077 (0.0405)	2.0158 (0.0199)	3.9752 (0.1589)	3.9987 (0.0395)
500	2.0016 (0.0160)	2.0070 (0.0079)	3.9901 (0.0639)	3.9961 (0.0159)	2.0054 (0.0161)	2.0061 (0.0079)	3.9956 (0.0640)	4.0021 (0.0159)
n	$\alpha_1 = 2, \alpha_2 = 5$		$\alpha_1 = 4, \alpha_2 = 5$		$\alpha_1 = 2, \alpha_2 = 7$		$\alpha_1 = 4, \alpha_2 = 7$	
50	4.7629 (0.9244)	4.8793 (0.1867)	4.7794 (0.9318)	4.8630 (0.1860)	6.4227 (1.6743)	6.5452 (0.2504)	6.4265 (1.6784)	6.5810 (0.2518)
100	4.8869 (0.4827)	4.9320 (0.0964)	4.8956 (0.4847)	4.9411 (0.0966)	6.7233 (0.9134)	6.7836 (0.1326)	6.7084 (0.9092)	6.7820 (0.1326)
200	4.9482 (0.2463)	4.9635 (0.0490)	4.9339 (0.2448)	4.9607 (0.0490)	6.8527 (0.4721)	6.8829 (0.0680)	6.8694 (0.4745)	6.8773 (0.0680)
500	4.9740 (0.0991)	4.9823 (0.0198)	4.9734 (0.0991)	4.9911 (0.0198)	6.9476 (0.1935)	6.9486 (0.0276)	6.9427 (0.1932)	6.9513 (0.0276)
n	$\alpha_1 = 2, \alpha_2 = 5$		$\alpha_1 = 4, \alpha_2 = 5$		$\alpha_1 = 2, \alpha_2 = 7$		$\alpha_1 = 5, \alpha_2 = 7$	
50	0.5046 (0.0093)	0.4650 (0.0047)	0.5046 (0.0093)	0.4650 (0.0047)	0.5046 (0.0093)	0.4650 (0.0047)	0.5046 (0.0093)	0.4650 (0.0047)
100	0.5024 (0.0048)	0.4720 (0.0024)	0.5024 (0.0048)	0.4720 (0.0024)	0.5024 (0.0048)	0.4720 (0.0024)	0.5024 (0.0048)	0.4720 (0.0024)
200	0.5012 (0.0024)	0.4817 (0.0012)	0.5012 (0.0024)	0.4817 (0.0012)	0.5012 (0.0024)	0.4817 (0.0012)	0.5012 (0.0024)	0.4817 (0.0012)
500	0.5004 (0.0009)	0.4915 (0.0005)	0.5004 (0.0009)	0.4915 (0.0005)	0.5004 (0.0009)	0.4915 (0.0005)	0.5004 (0.0009)	0.4915 (0.0005)

Table 3
Simulation for Exponential Prior when $p = 0.20, t = 30$

n	$\alpha_1 = 2, \alpha_2 = 5$		$\alpha_1 = 4, \alpha_2 = 5$		$\alpha_1 = 2, \alpha_2 = 7$		$\alpha_1 = 4, \alpha_2 = 7$	
	SELF	**PLF**	**SELF**	**PLF**	**SELF**	**PLF**	**SELF**	**PLF**
50	1.8659 (0.3769)	1.9545 (0.1819)	3.2189 (1.0979)	3.3820 (0.3148)	1.8751 (0.3816)	1.9516 (0.1816)	3.2237 (1.1031)	3.3881 (0.3153)
100	1.9287 (0.1945)	1.9898 (0.0959)	3.5764 (0.6641)	3.6778 (0.1772)	1.9327 (0.1955)	1.9849 (0.0956)	3.5772 (0.6645)	3.6901 (0.1778)
200	1.9657 (0.0989)	1.9983 (0.0491)	3.7840 (0.3658)	3.8408 (0.0942)	1.9678 (0.0991)	1.9902 (0.0488)	3.7917 (0.3675)	3.8335 (0.0941)
500	1.9930 (0.0401)	1.9989 (0.0198)	3.9185 (0.1550)	3.9325 (0.0390)	1.9859 (0.0398)	1.9970 (0.0198)	3.9196 (0.1551)	3.9348 (0.0391)
n	$\alpha_1 = 2, \alpha_2 = 5$		$\alpha_1 = 4, \alpha_2 = 5$		$\alpha_1 = 2, \alpha_2 = 7$		$\alpha_1 = 4, \alpha_2 = 7$	
50	4.6565 (0.5534)	4.7168 (0.1157)	4.6272 (0.5467)	4.7037 (0.1154)	6.6200 (1.0035)	6.3627 (0.1561)	6.2859 (1.0082)	6.3622 (0.1561)
100	4.8251 (0.2944)	4.8575 (0.0602)	4.8141 (0.2930)	4.8488 (0.0600)	6.7976 (0.5541)	6.6714 (0.0826)	6.6160 (0.5531)	6.6578 (0.0824)
200	4.9132 (0.1518)	4.9254 (0.0306)	4.9050 (0.1512)	4.9277 (0.0306)	6.7977 (0.2905)	6.8232 (0.0424)	6.8079 (0.2914)	6.8315 (0.0425)
500	4.9632 (0.0617)	4.9254 (0.0124)	4.9645 (0.0617)	4.9754 (0.0124)	6.9209 (0.1200)	6.9294 (0.0173)	6.9164 (0.1199)	6.9289 (0.0173)
n	$\alpha_1 = 2, \alpha_2 = 5$		$\alpha_1 = 4, \alpha_2 = 5$		$\alpha_1 = 2, \alpha_2 = 7$		$\alpha_1 = 5, \alpha_2 = 7$	
50	0.2076 (0.0153)	0.1815 (0.0031)	0.2076 (0.0153)	0.1815 (0.0031)	0.2076 (0.0153)	0.1815 (0.0031)	0.2076 (0.0153)	0.1815 (0.0031)
100	0.2039 (0.0078)	0.1875 (0.0015)	0.2039 (0.0078)	0.1875 (0.0015)	0.2039 (0.0078)	0.1875 (0.0015)	0.2039 (0.0078)	0.1875 (0.0015)
200	0.2019 (0.0039)	0.1918 (0.0008)	0.2019 (0.0039)	0.1918 (0.0008)	0.2019 (0.0039)	0.1918 (0.0008)	0.2019 (0.0039)	0.1918 (0.0008)
500	0.2007 (0.0015)	0.1989 (0.0003)	0.2007 (0.0015)	0.1989 (0.0003)	0.2007 (0.0015)	0.1989 (0.0003)	0.2007 (0.0015)	0.1989 (0.0003)

Table 4
Simulation for Inverted-Chi Square Prior when $p = 0.20, t = 30$

n	$\alpha_1 = 2, \alpha_2 = 5$		$\alpha_1 = 4, \alpha_2 = 5$		$\alpha_1 = 2, \alpha_2 = 7$		$\alpha_1 = 4, \alpha_2 = 7$	
	SELF	**PLF**	**SELF**	**PLF**	**SELF**	**PLF**	**SELF**	**PLF**
50	2.0553 (0.4450)	2.1473 (0.1928)	3.7032 (1.4180)	3.8875 (0.3490)	2.0511 (0.4428)	2.1529 (0.1932)	3.7139 (1.4297)	3.8815 (0.3488)
100	2.0285 (0.2118)	2.0804 (0.0984)	3.8667 (0.7661)	3.9775 (0.1882)	2.0329 (0.2127)	2.0952 (0.0991)	3.8768 (0.7692)	3.9688 (0.1877)
200	2.0184 (0.1034)	2.0394 (0.0495)	3.9388 (0.3932)	3.9834 (0.0968)	2.0128 (0.1027)	2.0449 (0.0497)	3.9420 (0.3940)	3.9916 (0.0970)
500	2.0080 (0.0406)	2.0198 (0.0199)	3.9733 (0.1588)	3.9990 (0.0395)	2.0092 (0.0406)	2.0207 (0.0199)	3.9759 (0.1590)	3.9969 (0.0395)
n	$\alpha_1 = 2, \alpha_2 = 5$		$\alpha_1 = 4, \alpha_2 = 5$		$\alpha_1 = 2, \alpha_2 = 7$		$\alpha_1 = 4, \alpha_2 = 7$	
50	4.8648 (0.5993)	4.9326 (0.1199)	4.8584 (0.5976)	4.9321 (0.1198)	6.6370 (1.1143)	6.7367 (0.1637)	6.6428 (1.1168)	6.7227 (0.1634)
100	4.9350 (0.3065)	4.9656 (0.0612)	4.9302 (0.3059)	4.9649 (0.0611)	6.8159 (0.5846)	6.8567 (0.0845)	6.8268 (0.5865)	6.8626 (0.0845)
200	4.9645 (0.1545)	4.9795 (0.0309)	4.9665 (0.1571)	4.9894 (0.0309)	6.9042 (0.2990)	6.9304 (0.0430)	6.9129 (0.2997)	6.9168 (0.0429)
500	4.9853 (0.0622)	4.9911 (0.0124)	4.9888 (0.0623)	4.9906 (0.0124)	6.9662 (0.1215)	6.9782 (0.0179)	6.9601 (0.1212)	6.9741 (0.0174)
n	$\alpha_1 = 2, \alpha_2 = 5$		$\alpha_1 = 4, \alpha_2 = 5$		$\alpha_1 = 2, \alpha_2 = 7$		$\alpha_1 = 5, \alpha_2 = 7$	
50	0.2188 (0.0158)	0.2115 (0.0035)	0.2188 (0.0158)	0.2115 (0.0035)	0.2188 (0.0158)	0.2115 (0.0035)	0.2188 (0.0158)	0.2115 (0.0035)
100	0.2097 (0.0081)	0.2058 (0.0018)	0.2097 (0.0081)	0.2058 (0.0018)	0.2097 (0.0081)	0.2058 (0.0018)	0.2097 (0.0081)	0.2058 (0.0018)
200	0.2049 (0.0045)	0.2030 (0.0010)	0.2049 (0.0045)	0.2030 (0.0010)	0.2049 (0.0045)	0.2030 (0.0010)	0.2049 (0.0045)	0.2030 (0.0010)
500	0.2019 (0.0018)	0.2011 (0.0005)	0.2019 (0.0018)	0.2011 (0.0005)	0.2019 (0.0018)	0.2011 (0.0005)	0.2019 (0.0018)	0.2011 (0.0005)

5. CONCLUSION

In this research article, we want to find suitable posterior distribution and loss function for the estimation of mixture model of reflected Weibull distribution. From the simulation study, we can conclude that by increasing sample size Bayes estimators approaches to their true value and also posterior risks decreases. On comparing both loss function precautionary loss function is performing better as its posterior risk is least as compare to square error loss function. Similarly exponential prior is also better for same above said reason. So, exponential prior with precautionary loss function can be used for further study. We have used only informative prior in our study, others can use non-informative priors and others priors with the aid of other loss functions can continue this study. We have also worked on two component mixture, other can work on three or more components for this distribution.

REFERENCES

1. Ng, H.K.T., Luo, L., Hu, Y. and Duan, F. (2012). Parameter estimation of three-parameter Weibull distribution based on progressively Type-II censored samples. *Journal of Statistical Computation and Simulation*, 82(11), 1661-1678.
2. Danish, M.Y. and Aslam, M. (2014). Bayesian inference for the randomly censored Weibull distribution. *Journal of Statistical Computation and Simulation*, 84(1), 215-230.
3. Cohen, A.C. (1973). The reflected Weibull distribution. *Technometrics*, 15(4), 867-873.
4. Almalki, S.J. and Nadarajah, S. (2014). Modifications of the Weibull distribution: A review. *Reliability Engineering & System Safety*, 124, 32-55.
5. Noor, F. and Aslam, M. (2013). Bayesian inference of the inverse Weibull mixture distribution using type-I censoring. *Journal of Applied Statistics*, 40(5), 1076-1089.
6. Singh, S.K., Singh, U. and Sharma, V.K. (2013). Bayesian prediction of future observations from inverse Weibull distribution based on Type-II hybrid censored sample. *International Journal of Advanced Statistics and Probability*, 1(2), 32-43.
7. Aslam, M. (2003). An application of prior predictive distribution to elicit the prior density. *Journal of Statistical Theory and Applications*, 2(1), 70-83.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 243-248

ON BAYESIAN ESTIMATION OF THE GENERALIZED INVERSE WEIBULL DISTRIBUTION USING NON-INFORMATIVE PRIORS

Muraffa Hanif[1], Wajiha Nasir[1], Mahnoor Shazad[1] and **Muhammad Zubair[2]**
[1] Department of Statistics, Govt. College Women University,
Sialkot, Pakistan. Email: murafahanif@gmail.com
wajiha.nasir@gcwus.edu.pk
mahnoors240@gmail.com
[2] Department of Statistics, University of Sargodha,
Sargodha, Pakistan. Email: zubair.bakhsh@uos.edu.pk

ABSTRACT

The generalized inverse Weibull distribution has the ability to model failure rates which are quite common in reliability and biological studies. A three-parameter generalized inverse Weibull distribution with decreasing and unimodal failure rate has been studied. We have worked in the case of scale parameter. Posterior distribution has been derived by using non-informative priors. Bayes estimators and their corresponding risks for posterior distributions has been derived by using different loss functions. Performance of Bayes estimator and their corresponding risk has been studied by Monte Carlo Simulation for different sample sizes.

1. INTRODUCTION

Estimates which are not only based on current information but also use prior information of sample data are said to have a Bayesian approach. Bayesian statistics uses rule of probability to make inference. Bayesian inference had been working as widely used and taught tool before the developments in statistics. The Inverse Weibull distribution is another life time probability distribution which can be used in the reliability engineering discipline. The Inverse Weibull distribution can be used to model a variety of failure characteristics such as infant mortality, useful life and wear-out periods. The Inverse Weibull distribution can also be used to determine the cost effectiveness and maintenance periods of reliability centered maintenance activities. The three parameters generalized inverse Weibull distribution has been estimated. Feroze and Aslam (2012) has studied Burr Type VII distribution using Bayesian analysis. They have utilized non-informative prior under different sampling schemes for posterior estimation. Feroze (2012) has studied inverse Gaussian distribution using Bayesian analysis. He used non-informative as well as informative priors to estimate the shape parameter of inverse Gaussian distribution.

Let G(t) be the cdf of the inverse Weibull distribution discussed by Drapella (1993), Mudholkar and Kollia (1994) and Jiang et al. (1999), among others. The cdf of the GIW distribution can be defined by elevating G(t) to the power of $\gamma > 0$, say

$F(t) = G(t)^{\gamma} = \exp-\left[\gamma\left(\frac{\alpha}{t}\right)^{\beta}\right]$. Hence the GIW density function with three parameters $\alpha, \beta, \gamma > 0$ is given by

$$f(t) = \gamma\beta\alpha^{\beta} t^{-(\beta+1)} \exp-\left[\gamma\left(\frac{\alpha}{t}\right)^{\beta}\right], \tag{1}$$

where γ is the scale parameter.

2. BAYESIAN ANALYSIS USING COMPLETE SAMPLES

In this section, Bayesian analysis under complete samples technique has been utilized. Posterior distribution has been derived under non- informative prior (uniform and Jeffery). The Bayes estimators and their corresponding risks are derived using Square error loss function (SELF), Precautionary Loss Function (PLF), and DeGroot Loss Function (DeLF).

Let the random sample $x_1, x_2, ..., x_n$ is assumed to be taken from GIW distribution with known shape parameter ϕ and assuming $\phi = 1$ and with unknown parameter. Then likelihood function is

$$L(t, \alpha, \beta, \gamma) = \gamma^{n}\beta^{n}\alpha^{n\beta}\prod_{i=t}^{n} t^{-(\beta+1)} \exp-\left[\gamma\left(\frac{\alpha}{t}\right)^{\beta}\right], \tag{2}$$

The uniform distribution is used as non- informative prior is

$$p(\gamma) = 1 \tag{3}$$

Now, the posterior distribution of γ using 2 and 3 is

$$p(\gamma / \underline{x}) = \frac{\left\{\left(\sum_{i=1}^{n}\frac{\alpha}{t}\right)^{\beta}\right\}^{n+1} \gamma^{(n+1)-1} e^{-\gamma\left(\sum_{i=1}^{n}\frac{\alpha}{t}\right)^{\beta}}}{\Gamma(n+1)} \tag{4}$$

so $\gamma | x \sim Gamma(\alpha_1, \beta_1)$, where $\alpha_1 = n+1$ and $\beta_1 = \sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}$

The Bayes estimator and posterior risks under SELF, PLF, and DeLF using uniform prior are as follows.

$$\hat{\gamma}_{(SELF)} = \frac{n+1}{\sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}} \quad \rho_{(SELF)} = \frac{n+1}{\left[\sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}\right]^{2}}$$

$$\hat{\gamma}_{(PLF)} = \frac{\sqrt{(n+1)(n+2)}}{\sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}} \qquad \rho_{(PLF)} = 2\frac{\sqrt{(n+1)(n+2)}-n+1}{\sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}}$$

$$\hat{\gamma}_{(DELF)} = \frac{n+2}{\sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}} \qquad \rho_{(DELF)} = \frac{1}{n+2}$$

Now the Jeffery prior used as non- informative prior is

$$p(\gamma) = \frac{1}{\gamma} \tag{5}$$

Now, the posterior distribution of γ using 2 and 5 is in this form

$$p(\gamma / \underline{x}) = \frac{\left\{\left(\sum_{i=1}^{n}\frac{\alpha}{t}\right)^{\beta}\right\}^{n}\gamma^{n-1}e^{-\gamma\left(\sum_{i=1}^{n}\frac{\alpha}{t}\right)^{\beta}}}{\Gamma(n)} \tag{6}$$

so $\gamma\big|x \sim Gamma(\alpha_2,\beta_2)$, where $\alpha_2 = n$, $\beta_2 = \sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}$

The Bayes estimator and posterior risks under SELF, PLF, and DeLF using Jeffery prior are as follows:

$$\hat{\gamma}_{(SELF)} = \frac{n}{\sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}} \quad \rho_{(SELF)} = \frac{n}{\left[\sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}\right]^{2}}$$

$$\hat{\gamma}_{(PLF)} = \frac{\sqrt{n(n+1)}}{\sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}} \quad \rho_{(PLF)} = 2\frac{\sqrt{n(n+1)}-n}{\sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}}$$

$$\hat{\gamma}_{(DeLF)} = \frac{n+1}{\sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}} \quad \rho_{(DeLF)} = \frac{1}{n+1}.$$

3. SIMULATION STUDY

Simulation studies are computer experiment that involve creating data by pseudorandom sampling. The key strength of simulation studies is the ability to understand the behavior. Simulation is used advantageously in a number of simulations. This includes providing the empirical estimation of sampling distributions, studying the misspecification of assumptions in statistical procedures, determining the power in hypothesis tests, etc. A simulation study is carried out to obtain the BEs and PRs under different loss functions using different priors. The simulation process is repeated 10,000 times considering generation of random samples of $n \in \{30, 70, 100, 200\}$ from inverse method assuming $\alpha \in \{2,5\}, \beta \in \{2,3\}$ and $\gamma \in \{2,3\}$ and the results have then been averaged. These results are shown in the following tables

Table 1
Expression of BEs and PRs using Uniform Prior

	$\alpha=2,\beta=2,\gamma=2$			$\alpha=2,\beta=2,\gamma=3$		
	SELF	PLF	DeLF	SELF	PLF	DeLF
30	2.1344 (0.1522)	2.1717 (0.0684)	2.1981 (0.0312)	3.2136 (0.3452)	3.2605 (0.1027)	3.3023 (0.0312)
70	2.0570 (0.0605)	2.0755 (0.0289)	2.0848 (0.0138)	3.0854 (0.1360)	3.1051 (0.0432)	3.1363 (0.0138)
100	2.0412 (0.0416)	2.0484 (0.0201)	2.0628 (0.0098)	3.0631 (0.0938)	3.0760 (0.0302)	3.0868 (0.0098)
200	2.0185 (0.0203)	2.0252 (0.0100)	2.0306 (0.0049)	3.0325 (0.0459)	3.0383 (0.0150)	3.0493 (0.0049)
	$\alpha=2,\beta=3,\gamma=2$			$\alpha=2,\beta=3,\gamma=3$		
30	2.1391 (0.1527)	2.1781 (0.0686)	2.2144 (0.0312)	3.2082 (0.3435)	3.2614 (0.1027)	3.3124 (0.0312)
70	2.0581 (0.0605)	2.0707 (0.0288)	2.0825 (0.0138)	3.0800 (0.1356)	3.1126 (0.0433)	3.1250 (0.0138)
100	2.0408 (0.0416)	2.0489 (0.0201)	2.0654 (0.0098)	3.0555 (0.0933)	3.0745 (0.0302)	3.0867 (0.0098)
200	2.0245 (0.0204)	2.0254 (0.0100)	2.0320 (0.0049)	3.0293 (0.0458)	3.0401 (0.0150)	3.0498 (0.0049)
	$\alpha=5,\beta=2,\gamma=2$			$\alpha=5,\beta=2,\gamma=3$		
30	2.1369 (0.1525)	2.1724 (0.0684)	2.2082 (0.0312)	3.1999 (0.3420)	3.2529 (0.1024)	3.3069 (0.0312)
70	2.0592 (0.0606)	2.0686 (0.0288)	2.0865 (0.0138)	3.0905 (0.1365)	3.1101 (0.0433)	3.1242 (0.0138)
100	2.0398 (0.0416)	2.0478 (0.0201)	2.0589 (0.0098)	3.0609 (0.0936)	3.0756 (0.0302)	3.0894 (0.0098)
200	2.0198 (0.0204)	2.0263 (0.0100)	2.0322 (0.0049)	3.0303 (0.0459)	3.0358 (0.0150)	3.0447 (0.0049)
	$\alpha=5,\beta=3,\gamma=2$			$\alpha=5,\beta=3,\gamma=3$		
30	2.1429 (0.1534)	2.1710 (0.0683)	2.1952 (0.0312)	3.2112 (0.3446)	3.2537 (0.1024)	3.3117 (0.0312)
70	2.0585 (0.0605)	2.0732 (0.0288)	2.0846 (0.0138)	3.0892 (0.1364)	3.1066 (0.0432)	3.1295 (0.0138)
100	2.0424 (0.0417)	2.0515 (0.0201)	2.0610 (0.0098)	3.0592 (0.0936)	3.0790 (0.0302)	3.0889 (0.0098)
200	2.0221 (0.0204)	2.0259 (0.0100)	2.0284 (0.0049)	3.0327 (0.0459)	3.0360 (0.0150)	3.0452 (0.0049)

Table 2
Expression of BEs and PRs using Jeffreys Prior

	$\alpha=2,\beta=2,\gamma=2$			$\alpha=2,\beta=2,\gamma=3$		
	SELF	PLF	DeLF	SELF	PLF	DeLF
30	2.0720 (0.1481)	2.1059 (0.0684)	2.1381 (0.0322)	3.1126 (0.3345)	3.1525 (0.1025)	3.2005 (0.0322)
70	2.0337 (0.0599)	2.0419 (0.0288)	2.0546 (0.0140)	3.0452 (0.1343)	3.0710 (0.0434)	3.0816 (0.0140)
100	2.0231 (0.0413)	2.0282 (0.0201)	2.0407 (0.0099)	3.0285 (0.0926)	3.0453 (0.0302)	3.0647 (0.0099)
200	2.0123 (0.0203)	2.0163 (0.0100)	2.0196 (0.0049)	3.0173 (0.0457)	3.0241 (0.0150)	3.0304 (0.0049)
	$\alpha=2,\beta=3,\gamma=2$			$\alpha=2,\beta=3,\gamma=3$		
30	2.0667 (0.1474)	2.1043 (0.0684)	2.1429 (0.0322)	3.0929 (0.3301)	3.1496 (0.1024)	3.2092 (0.0322)
70	2.0333 (0.0599)	2.0423 (0.0288)	2.0580 (0.0140)	3.0487 (0.1347)	3.0668 (0.0433)	3.0834 (0.0140)
100	2.0191 (0.0411)	2.0328 (0.0201)	2.0386 (0.0099)	3.0327 (0.0929)	3.0478 (0.0302)	3.0587 (0.0099)
200	2.0114 (0.0203)	2.0145 (0.0100)	2.0203 (0.0049)	3.0144 (0.0456)	3.0213 (0.0150)	3.0300 (0.0049)
	$\alpha=5,\beta=2,\gamma=2$			$\alpha=5,\beta=2,\gamma=3$		
30	2.0680 (0.1477)	2.0999 (0.0682)	2.1354 (0.0322)	3.1158 (0.3354)	3.1601 (0.1027)	3.2096 (0.0322)
70	2.0252 (0.0594)	2.0427 (0.0288)	2.0571 (0.0140)	3.0416 (0.1340)	3.0667 (0.0433)	3.0859 (0.0140)
100	2.0212 (0.0412)	2.0285 (0.0201)	2.0422 (0.0099)	3.0303 (0.0927)	3.0459 (0.0302)	3.0649 (0.0099)
200	2.0129 (0.0203)	2.0148 (0.0100)	2.0209 (0.0049)	3.0184 (0.0457)	3.0212 (0.0150)	3.0321 (0.0049)
	$\alpha=5,\beta=3,\gamma=2$			$\alpha=5,\beta=3,\gamma=3$		
30	2.0612 (0.1466)	2.1044 (0.0684)	2.1390 (0.0322)	3.1082 (0.3332)	3.1576 (0.1026)	3.2110 (0.0322)
70	2.0251 (0.0594)	2.0434 (0.0288)	2.0618 (0.0140)	3.0382 (0.1337)	3.0688 (0.0433)	3.0853 (0.0140)
100	2.0173 (0.0411)	2.0284 (0.0201)	2.0392 (0.0099)	3.0322 (0.0928)	3.0484 (0.0302)	3.0602 (0.0099)
200	2.0132 (0.0203)	2.0157 (0.010)	2.0193 (0.0049)	3.0138 (0.0456)	3.0214 (0.0150)	3.0244 (0.0049)

4. COMMENTS AND CONCLUSION

We concluded from this study, by increasing sample size Bayes posterior risk decreases and Bayes estimator approaches to its true value of the parameter. By increasing the value of parameter Bayes posterior risk also increases. After the simulation and comparing different loss functions, the results shows that DeLF is performing better because its Bayes posterior risk is smaller than other loss functions. Uniform Prior gives the best estimator as compare to Jeffery's. Uniform prior with DeLF is performing better among others. Uniform prior is found to be the best prior for the estimation. The work has been done by using only two non-informative priors with three loss functions others can extend their study by using other prior and loss functions. We have used only complete samples techniques others can used censored samples or other sampling schemes.

REFERENCES

1. De Gusmao, F.R., Ortega, E.M. and Cordeiro, G.M. (2011). The generalized inverse Weibull distribution. *Statistical Papers*, 52(3), 591-619.
2. Berger, J.O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.
3. Feroze, N. and Aslam, M. (2012). On Bayesian analysis of Burr type VII distribution under different censoring schemes. *International Journal of Quality, Statistics, and Reliability*, 1-5, doi: 10.1155/2012/248146.
4. Feroze, N. (2012). Estimation of scale parameter of inverse Gaussian distribution under a Bayesian framework using different loss functions. *Scientific Journal of Review*, 1(3), 39-52.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 249-257

FORECASTING OF GOLD PRICES IN PAKISTAN USING ARIMA MODEL

Sidra Razzaq§, M. Waqas and **Samina Satti**
Department of Statistics, University of Wah, Wah Cantt, Pakistan
§Email: sidrarazaaq@gmail.com

ABSTRACT

The data of Gold prices in Pakistan shows that there is variability in the mean and variance of the data. To check the stationarity of the data, ACF's and PCF's are found which shows that the data is not stationary. As the data is not stationary, so an attempt is made to make it stationary. For this purpose correlogram of first difference was applied which made the data stationary. After the fulfilment of the basic assumptions, seven models were estimated. The model having the less Akaike information criterion (AIC) and Bayesian information criterion (BIC) was selected. In this paper the (1, 1, 0) model has the least value of Akaike and Bayesian information criterion. So this model is the best fitted model. After plotting the normality curve, forecasting was performed.

KEYWORDS

Gold prices, ACF, PCF, Price Forecasting.

1. INTRODUCTION

Gold has been preferred by men and women since the beginning of time. The metal gold plays an important role in various sectors i.e. jewelry, computers, electronics, aerospace, glassmaking and medical.

Government holds gold as a standard for currency equivalents. The demand of gold will be increasing due to its uses in the near future. In today's world, like many commodities, the price of gold is driven by supply and demand. The prices of gold are prominently increasing in this era especially since the beginning of 2006. There is very little literature available in predicting the properties of gold and other financial assets in Pakistan. The present research study seeks to address this gap. Hence the forecasting of gold prices is of great importance. The aim of this research is to forecast the value of Pakistani Gold utilizing ARMA and ARIMA Model. The accurate forecasting of the gold price will help the policymakers and the investors to take better investment decisions in the market. Hence this paper gives an insight of forecasting of Gold price through time-series ARMA and ARIMA model.

2. LITERATURE REVIEW

The prediction of gold prices is receiving prominent attention among the researchers these days. Since gold is an important asset for policymakers and investors, hence people are much interested in keeping account of the forecasted prices of gold. There is no prominent research work that has been done about the gold market of Pakistan. Tufail and Batool (2013) assess the inflation-hedging properties of gold compared to other assets such as real estate, stock exchange securities, and foreign currency holdings. Shahbaz et al. (2014) investigated whether a gold investment is a hedge against inflation in case of Pakistan. Similarly, Nadeem et al. (2014) aimed to study the effects of inflation, stock price, international gold price, rupees per dollar exchange rate, international oil price and income on domestic gold price. Another research conducted by Khalid et al. (2014) studied the forecasting of Gold Price taking the evidence from Pakistan Market.

3. DATA AND METHODOLOGY

Keeping the interest of investor in mind, this research is aimed at predicting the gold prices of Pakistan's market. This study employs monthly gold prices from 2006 to 2018. The data has been taken from Historical Chart of gold prices in Pakistan.

As this research work has only one variable so it is a univariate time series data. For such type of data we will see weather our data is autoregressive or not. In time series analysis, the second model is the moving-average (MA model). It is also known as moving-average process, which is a common approach for modeling univariate time series. The moving-average model specifies that the output variable depends linearly on the current and various past values of a stochastic (imperfectly predictable) terms. Together with the autoregressive (AR) model, the moving-average model is a special case and the key component of the more general ARMA and ARIMA models. ARMA stands for autoregressive moving average while ARIMA stands for autoregressive integrated moving average. This method is also known as the Box-Jenkins method. Contrary to the AR model, the finite MA model is always stationary. A model is applied to a data according to its pattern. In order to test whether or not the series and their error terms are auto correlated, we usually use collelogram which gives us the values of ACF and PACF.

AR stands for autoregressive is denoted by p. When p =0, it means that there is no auto-correlation in the series. When p=1, it means that the series auto-correlation is till one lag. MA stands for moving average. It is denoted by q. q=1 means that it is an error term and it is auto-correlation with one lag.

In ARIMA time series analysis, integrated is denoted by d. Integration is the inverse of differencing. When d=0, it means that the series is stationary. There is no need to take the difference of it. When d=1, it means that the series is not stationary and to make it stationary, we need to take the first difference. When d=2, it means that the series has been differenced twice. More than two-time difference is not reliable, usually.

ARIMA model has basically four steps i.e. identification, estimation, diagnostic checking and forecasting. Before using ARIMA model, it is important to check the stationarity of the data. This means that the series are normally distributed and the mean and variance are constant over a long period of time. For this purpose we use a unit root test.

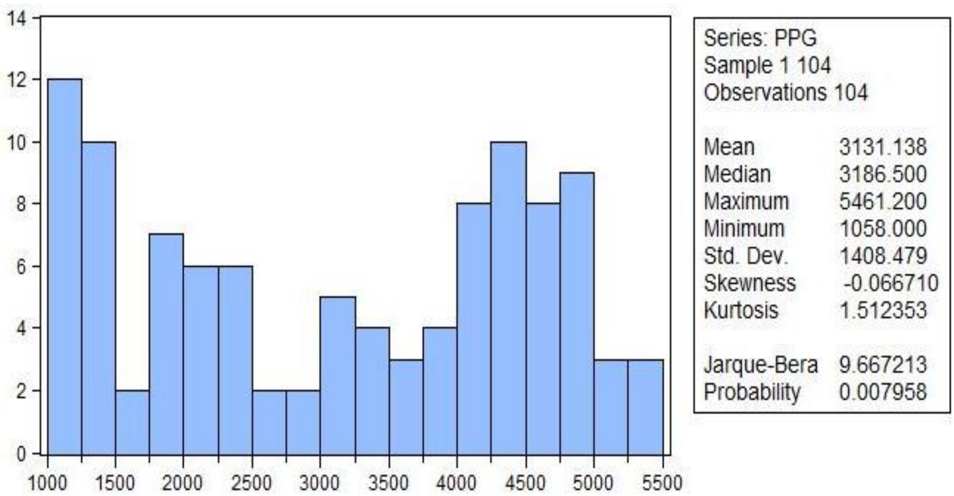
The selection of the regression model will be made by observing the autocorrelation function (ACF) and partial ACF (PACF). The ACF and PACF are used to find the evidence of its stationary condition. If the ACF dies off smoothly at a geometric rate, and the PACs become zero after one lag, then a first-order autoregressive model is appropriate. Alternatively, if the AC zero after one lag and the partial ACs declined geometrically, then a first-order moving average process is necessary.

Augmented Dickey-Fuller test is one of the widely used methods for testing if the variable is free of the trend component. If the series is non-stationary, the series needs to be transformed to first difference data and tested again. The data would be stationary. If it is not then the series is again difference till it becomes stationary. The next step would be to identify the model.

The Durbin-Watson test is the standard test for auto correlation. It is used to detect the presence of autocorrelation. Its value always lies between 0 and 4. If d = 2, it indicates the presence of no autocorrelation. If it is less than 2, there is an evidence of positive correlation and if it is greater than 2, then the errors terms are negatively correlated.

4. MODEL ESTIMATION

The first step that is taken before the model selection is to find out the nature of the data. For this purpose Histogram is constructed.



From the graph of Histogram it can be seen that the data is not perfectly normal but it is also not too much skewed. So there is no need to apply any kind of transformation on this data. This is also supported by Jarque-Bera which is a goodness of fit test.

5. ARMA Model

ARMA is a widely adopted model for predicting and forecasting using the time series data. As we know that Auto-Regressive models and Moving Average models are found to only partially explained the log-returns of stock prices. So we combine the Autoregressive model and Moving Average model to produce a more sophisticated model, Auto Regressive Moving Average (ARMA) model. In other words we can say that ARMA model is simply the merger between AR (p) and MA(q) models.

Date: 12/28/18 Time: 11:57
Sample: 1 104
Included observations: 104

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.968	0.968	100.30	0.000
		2	0.941	0.067	196.07	0.000
		3	0.916	0.015	287.66	0.000
		4	0.896	0.070	376.08	0.000
		5	0.872	-0.050	460.74	0.000
		6	0.843	-0.099	540.69	0.000
		7	0.819	0.056	616.97	0.000
		8	0.793	-0.045	689.27	0.000
		9	0.770	0.016	758.12	0.000
		10	0.746	-0.009	823.37	0.000
		11	0.722	-0.018	885.08	0.000
		12	0.696	-0.037	943.14	0.000
		13	0.667	-0.074	997.01	0.000
		14	0.644	0.070	1047.8	0.000
		15	0.621	-0.004	1095.6	0.000
		16	0.598	-0.016	1140.4	0.000
		17	0.572	-0.036	1182.0	0.000
		18	0.547	-0.026	1220.3	0.000
		19	0.521	-0.036	1255.5	0.000
		20	0.493	-0.057	1287.4	0.000
		21	0.466	-0.009	1316.2	0.000
		22	0.437	-0.039	1341.8	0.000
		23	0.409	-0.011	1364.6	0.000
		24	0.372	-0.172	1383.6	0.000
		25	0.341	0.058	1399.8	0.000
		26	0.310	-0.037	1413.4	0.000
		27	0.280	-0.009	1424.6	0.000
		28	0.251	0.020	1433.7	0.000
		29	0.220	-0.029	1440.8	0.000
		30	0.189	-0.076	1446.2	0.000
		31	0.156	-0.034	1449.9	0.000
		32	0.125	-0.035	1452.2	0.000
		33	0.095	0.008	1453.6	0.000
		34	0.062	-0.083	1454.2	0.000
		35	0.028	-0.035	1454.4	0.000
		36	-0.009	-0.076	1454.4	0.000

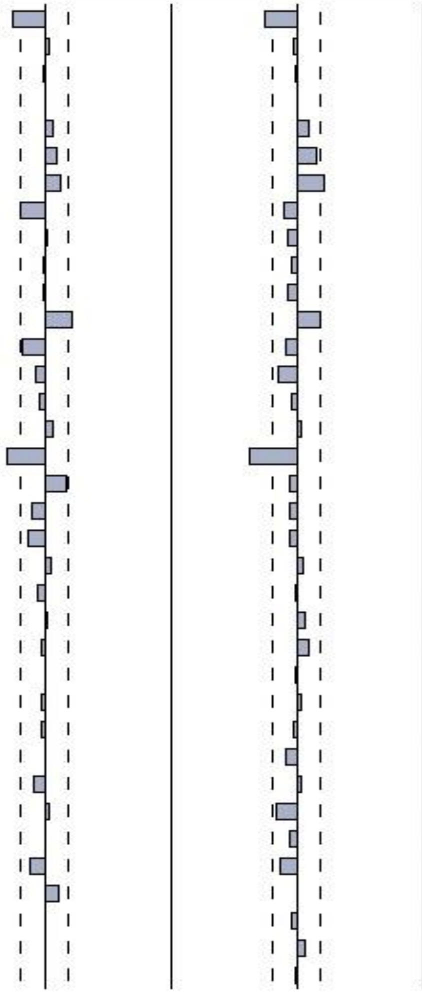
Correlogram is applied for checking the randomness in a data. As it can be seen from the values of ACF and PACF, there is too much randomness in the data. To remove this randomness we have to take the first or second difference. As ARMA model can only be applied at levels, so ARMA model is not fit for our data.

6. ARIMA Model

Box-Jenkins ARIMA (Auto Regressive Integrated Moving Average) model is one of the extensively used models for forecasting. In this model the prediction is made only from the historical series of a variable. To reduce the randomness of the data, first difference is taken and then correlogram is applied. It can been seen that the results have improved at first difference.

Date: 12/28/18 Time: 00:59
Sample: 1 104
Included observations: 103

Autocorrelation | Partial Correlation



	AC	PAC	Q-Stat	Prob
1	-0.268	-0.268	7.6427	0.006
2	0.044	-0.031	7.8473	0.020
3	-0.015	-0.012	7.8722	0.049
4	0.013	0.008	7.8915	0.096
5	0.079	0.091	8.5796	0.127
6	0.104	0.162	9.7828	0.134
7	0.133	0.229	11.770	0.108
8	-0.189	-0.099	15.840	0.045
9	0.026	-0.075	15.917	0.069
10	-0.004	-0.043	15.919	0.102
11	-0.007	-0.068	15.925	0.144
12	0.227	0.194	22.037	0.037
13	-0.185	-0.090	26.140	0.016
14	-0.070	-0.146	26.731	0.021
15	-0.033	-0.044	26.867	0.030
16	0.073	0.041	27.532	0.036
17	-0.306	-0.380	39.331	0.002
18	0.173	-0.050	43.130	0.001
19	-0.101	-0.060	44.432	0.001
20	-0.135	-0.061	46.809	0.001
21	0.065	0.064	47.365	0.001
22	-0.055	-0.001	47.766	0.001
23	0.018	0.074	47.809	0.002
24	-0.031	0.092	47.943	0.003
25	0.002	-0.006	47.943	0.004
26	-0.026	0.047	48.035	0.005
27	-0.022	-0.023	48.106	0.007
28	0.008	-0.096	48.116	0.010
29	-0.085	0.049	49.176	0.011
30	0.035	-0.156	49.359	0.014
31	0.008	-0.062	49.368	0.019
32	-0.115	-0.134	51.374	0.016
33	0.119	0.004	53.562	0.013
34	0.011	-0.049	53.580	0.018
35	0.006	0.077	53.586	0.023
36	0.010	-0.004	53.603	0.030

The next step is to identify the model. Before the identification process the presence of autocorrelation is checked using a Unit Root test.

Null Hypothesis: D(PPG) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-13.15405	0.0000
Test critical values:	1% level	-3.495677	
	5% level	-2.890037	
	10% level	-2.582041	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(PPG,2)
Method: Least Squares
Date: 12/28/18 Time: 00:55
Sample (adjusted): 3 104
Included observations: 102 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(PPG(-1))	-1.269806	0.096533	-13.15405	0.0000
C	54.26595	17.37268	3.123638	0.0023

R-squared	0.633739	Mean dependent var	1.530392
Adjusted R-squared	0.630076	S.D. dependent var	280.6902
S.E. of regression	170.7196	Akaike info criterion	13.13734
Sum squared resid	2914519.	Schwarz criterion	13.18881
Log likelihood	-668.0041	Hannan-Quinn criter.	13.15818
F-statistic	173.0290	Durbin-Watson stat	2.007171
Prob(F-statistic)	0.000000		

The Augmented Dickey-Fuller statistics is a negative number. The more negative it is, the stronger the chance is that there is no unit root present. It can also be seen that the value of Durbin-Watson statistic is 2.00, which means that there is no autocorrelation present in this data.

ARIMA model takes the historical data and decomposes it into an autoregressive process. In ARIMA (p, d, q) model, where p attributes the order of the auto regressive process, d is the order of the data stationary and q is the order of the moving average terms based on Box-Jenkins methodology. The study used the Akaike information criterion (AIC) and Schwartz Bayesian criterion (BIC) for choosing alternative ARIMA model. The models with the lowest AIC and BIC is chosen for the forecasting. The seven ARIMA models selected for this study are shown in Table 1.

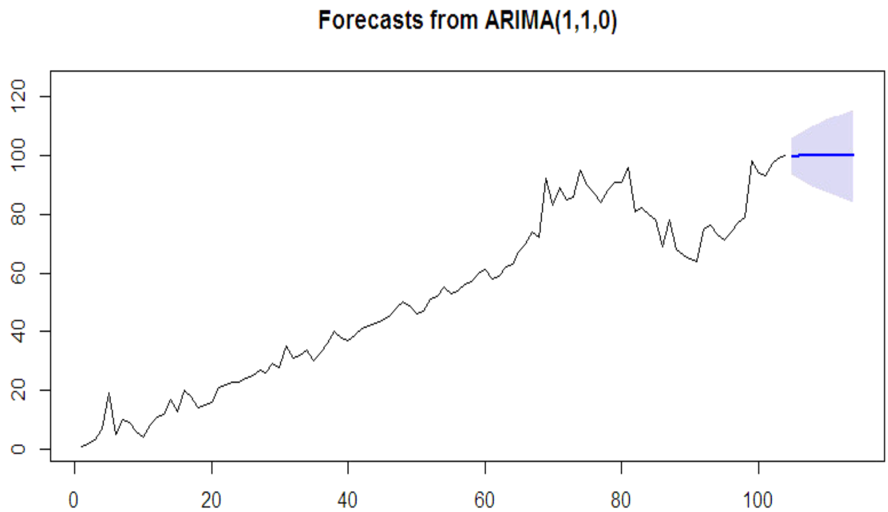
Table 1
ARIMA Models

ARIMA	Sigma^2	AIC	BIC
(0,1,1)	23.85	622.05	627.32
(1,1,0)	**23.47**	**620.44**	**625.71**
(1,1,1)	23.66	622.25	630.16
(1,1,7)	23.08	626.05	649.76
(1,1,8)	23.10	627.33	653.67
(8,1,1)	22.79	625.79	652.14
(8,1,8)	18.57	621.37	666.16

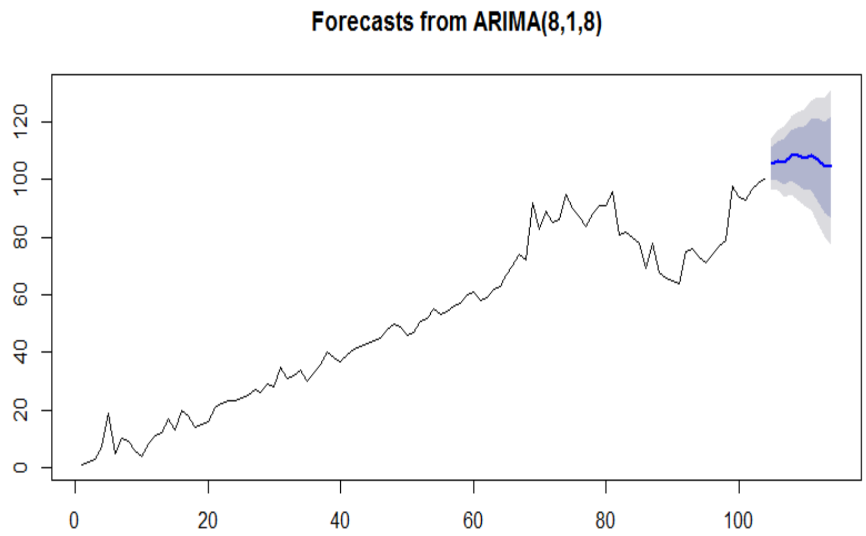
From these models, a model is selected whose values of AIC and BIC is the lowest of all. It can be seen that the best fitted model is (1,1,0).

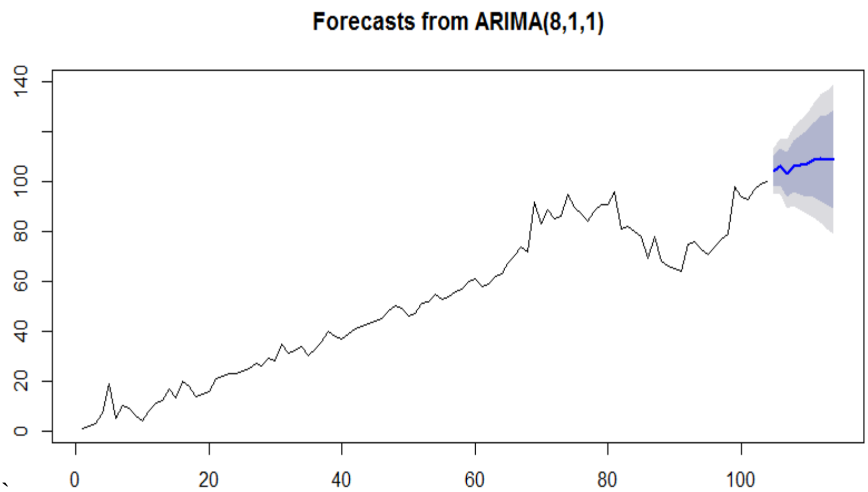
7. FORECASTING

The main objective of this study is to predict the future gold price in Pakistan. The forecasting of these models is shown with the help of the graphs.



As suggested by the AIC and BIC the best fitted model is (1, 1, 0) which can also be seen from the graph.





The graph of (8, 1, 8) and (8, 1, 1) are shown for comparison with (1, 1, 0). The smaller the forecasting error is, the better the prediction model.

The present research suggests that ARIMA model can be used for forecasting the gold prices in Pakistan. It also attempts to identify the sensitivity of gold prices. It is observed that the Pakistani gold prices with first difference are significant in explaining the variation in the price of gold.

8. CONCLUSION

The study aims to establish and corroborate the prediction of gold price in Pakistan. The paper uses the monthly time series data to discover the forecasting of gold price. The study used ARIMA model to predict the gold price. The study has further used seven different forecasting models such as (0,1,1), (1,1,0), (1,1,1), (1,1,7), (1,1,8), (8,1,1), (8,1,8) to determine the best model. The result shows that ARIMA (1,1,0) is the best model for gold price prediction since the AIC and BIC is lowest in this model.

9. ACKNOWLEDGEMENT

The authors wish to thank anonymous referees for their careful reading and constructive suggestions which leads to the improvement over an earlier version of the paper. Thanks to the Department of Statistics, University of Wah, Wah Cantt, Pakistan for their logistic and financial support for this research.

REFERENCES

1. Khalid, M., Sultana, M. and Zaidi, F. (2014). Forecasting Gold Price: Evidence from Pakistan Market. *Research Journal of Finance and Accounting*, 5(3), 70-74.
2. Nadeem, W. Zakaria, M., and Kayani, F. (2014). Impact of Macroeconomic Factors Upon Gold Prices in Pakistan. *Pakistan Journal of Social Sciences (PJSS)*, 34, 383-395.
3. Shahbaz, M., Tahir, M.I., Ali, I. and Rehman, I.U. (2014). Is gold investment a hedge against inflation in Pakistan? A co-integration and causality analysis in the presence of structural breaks. *The North American Journal of Economics and Finance*, 28, 190-205.
4. Tufail, S. and Batool, S. (2013). An Analysis of the Relationship between Inflation and Gold Prices Evidence from Pakistan. *The Lahore Journal of Economics*, 18(2), 1-35.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 259-264

USE OF MOTIVATIONAL TECHNIQUES BY PROSPECTIVE TEACHER IN ENHANCING STUDENT'S PARTICIPATION AND ACHIEVEMENT IN THE SUBJECT OF ENGLISH AT ELEMENTARY LEVEL

Numaila Hanif Chaudhry
University of Education, Lahore, Pakistan
Email: numailachaudhry381@gmail.com

ABSTRACT

The study describes the Use of motivational techniques by prospective teacher in enhancing student's participation and achievement in the subject of English at elementary level. The study indicate that the systems which are developed on the basis of ARCS Model(attention, relevance, confidence, satisfaction) raise the attention of the students during instruction, develop a relevance to the students' requirements, create a positive expectation for success and help having a satisfaction by reinforcing success.

Aims of study;

- To observe the effect of motivational techniques used by the prospective teacher on student's achievement in the subject of English at elementary level
- To observe the effect of motivational techniques used by the prospective teacher on student's participation in the subject of English at elementary level

The population of the study was all the section of 7th grade of elementary level students enrolled in Kinnaird Girls High School Empress Road Lahore. The sample of the research was conducted on section of 7th grade students enrolled in Kinnaird Girls High School Empress Road Lahore. English achievement test and observational sheet was developed by the researcher to check the achievement scores and participation of 7th grade students respectively. Researcher conducted a pre-test and pre-observation then after interventions of motivational techniques post-test and post-observation of the class. Researcher analyzed the data through SPSS. The conclusion of this study showed that there is significant difference between achievement and participation of the students in the subject of English at elementary level with and without using motivational techniques.

KEYWORDS

Motivational techniques, participation, achievement, elementary level.

INTRODUCTION

Motivation is a force that energizes, sustains and directs behavior towards a goal. When students pay attention to their work and study for tests, we say they are motivated (Eggen & Kauchak, 1997, P, 341). Weinert (1987) writes, the word motivation is derived from Latin word "mover" meaning "to move" motivation stimulates the movement of an individual. Motivation refers to those factors which increase or decrease the vigor of an individual's activity. In the field of education, motivation is often called effort. In term of motivation searches for those factors which increase the student effort to make desirable responses (Decco 1979, P, 13). Based on above definitions it is considered that "motivation is an internal process that activates, guides and maintain behavior"

Types of motivation

Motivation can be described in two broad categories

- Intrinsic motivation
- Extrinsic motivation

Intrinsic motivation is motivation to engage in an activity for its own sake. For example, intrinsically motivated learners study because studying is viewed as enjoyable or worthwhile in itself (Eggen & Kauchak, 1997, P, 342)

Extrinsic motivation refers to motivation to engage in an activity as a means to an end. For example, extrinsically motivated learners may study hard for a test because they believe studying will lead to a high test score, teacher compliments, a good grade in the class, or some other reward (Eggen & Kauchak, 1997, P, 342)

For a teacher student's, motivation is important because it can serve as both as an objective in itself and a means for further achievement of other educational objectives. As an objective motivation become interested in certain educational and intellectual activities (Gage & Berliner, 1984, PP, 373-374). Motivational Design Theory asserts that instructional material should be configured with the strategies which increase the attention, relevance, confidence and satisfaction of the students for an instructional design which ensures the continuity of learning motivation (Keller, 1983; Keller & Kopp, 1987). The ARCS model of motivational design arouse interest in the learners to complete this quality instructional material. There are numerous instructional design models out of which The ARCS Model of Motivational Design is the most effective models to overcome the challenge of high non completion rate. The ARCS Model of Motivational Design was created by John Keller while he was researching ways to supplement the learning process with motivation. The model consists of four main areas: Attention, Relevance, Confidence, and Satisfaction. According to Keller, the goal of the many instructional design theories that have been developed is providing an effective and efficient instruction. However, in these theories the aspect of motivation has been generally ignored whereas motivation has to be essence of learning. Attention and relevance according to John Keller's ARCS motivational theory are essential to learning.

The first 2 of 4 key components for motivating learners, attention and relevance can be considered the backbone of the ARCS theory, the latter components confidence and satisfaction relying upon the former. The study indicate that the systems which are developed on the basis of ARCS Model raise the attention of the students during instruction, develop a relevance to the students' requirements, create a positive expectation for success and help having a satisfaction by reinforcing success. Education is must for development of the country and for the prosperity of its people. Pakistan has realized this fact and it is trying to sweep out illiteracy and to raise the standard of education. Teacher is an important character in this respect, because a student's performance depends on the teacher's strategies, method and styles in the classroom. For effective teaching motivational techniques are very important. A good teacher uses different strategies to motivate students. Through this study we will be able to know the techniques which are used by teacher to motivate the students at elementary level.

OBJECTIVES

The main objectives of study are

- To observe the effect of motivational techniques used by the prospective teacher on student's achievement in the subject of English at elementary level
- To observe the effect of motivational techniques used by the prospective teacher on student's participation in the subject of English at elementary level

HYPOTHESIS

H01 There is no significant difference between the achievement scores of the student in the subject of English with and without using motivational techniques

H02 There is no significant difference between the participation of the student in the subject of English with and without using motivational techniques

RESEARCH METHODOLOGY

The purpose of the study was to determine the effect of motivational techniques used by prospective teacher on student's participation and achievements. The research design for the study was quantitative, experimental research designs because it is appropriate to the aim of the study. Students of 7^{th} grade in Government Kinnaired Girls High School were the participant of the study. Researcher selected one section as a sample. English achievement test was developed by the researcher to check the achievement scores of 7^{th} grade students. For this purpose researcher selected the three chapters of the subject of English. Observation sheet was developed by the researcher to check the participation of 7^{th} grade students. Researcher conducted a pre-test or post-test and pre-observation sheet or post-observation sheet of the class after studying the related issues and trends of prospective teacher motivational techniques on student participation and achievements.

DATA COLLECTION PROCEDURE

The researcher seeks to observe the effect of motivational techniques used by prospective teacher to enhancing student's participation and achievement through pre-post test and pre-post observation sheet. It is pertinent to mention that the researcher made the pre-post test to observe the student's achievement. The researcher guided the students how to attempt the test. It is also stated that the researcher administered and supervised the test also. Researcher made the pre-post observation sheet to observe the student's participation. Researcher has selected students then took pre-test and pre-observation. Researcher used intervention of motivational technique during teaching practice days which are consisted of only 24 days. After 20 days researcher took post-test and post-observations.

DATA ANALYSIS

Data was analyzed through SPSS (22.0) to observe the effect of motivational techniques used by prospective teacher to enhancing student's participation and achievement. Descriptive statistics, t-test, mean, standard deviation was applied. After collecting the data of pre-test, pre-observation and post-test, post-observation of students then find out the results.

FINDINGS AND CONCLUSIONS

The findings of the study were discussed

H01 There is no significant difference between the achievement scores of the student in the subject of English with and without using motivational techniques.

Table 4.1
Summary of paired sample t-test to find the difference between the achievement scores of the students in the subject of English with and without using motivational techniques

	Mean	SD	T	Df	Sig.(2-tailed)
Pre-post achievement	7.20000	3.02176	13.051	29	000

Paired sample t-test was applied to compare pre-test and post-test scores of students. So the null hypothesis stating that there is no significant difference between pre-test and post-test scores of students after giving them motivational techniques. And it is concluded that there is the significant difference between the achievement scores of the students in the subject of English with and without using motivational techniques

H02 There is no significant difference between the participation of the student in the subject of English with and without using motivational techniques.

Table 4.2
Summary of paired sample t-test to find the difference between the participation of the students in the subject of English with and without using motivational techniques

	Mean	SD	T	Df	Sig.(2-tailed)
Pre-post observation	5.30000	2.10336	13.801	29	000

Paired sample t-test was applied to compare pre-test and post-test scores of students. So the null hypothesis stating that there is no significant difference between pre-test and post-test scores of students after giving them motivational techniques. And it is concluded that there is the significant difference between the participation of the students in the subject of English with and without using motivational techniques.

CONCLUSION

Following conclusion was drawn from the findings of the study. The finding of results showed that there is significant difference achievement and participation of the students in the subject of English with and without using motivational techniques.

RECOMENDATIONS

Following recommendations are made based on the findings.

- Motivation is basically an essential ingredient for learning so teacher should motivate their students effectively.
- To motivate their student teacher should use ARCS Model of motivation which is also used in this study.
- As per research the motivational method has vital role for enhancing student’s participation and achievement.
- It is added more that student gains their targeted goals in future through this technique.
- It is pertinent to mention here that this study established the standards of learning.

REFERENCES

1. Anderson, E. and Weitz, B. (1992). The use of pledges to build and sustain commitment in distribution channels. *Journal of Marketing Research*, 29(1), 18-34.
2. Andualem, T. (1997). Attributions and Academic Achievement of Educatiori, Medicine and Polytechnic Freshman Students in Bahir Dar. *The Ethiopian Journal of Education*, 17(2), 63-78.
3. Barrett, G.V. and Depinet, R.L. (1991). A reconsideration of testing for competence rather than for intelligence. *American Psychologist*, 46(10), 1012.

4. Boyd, R.D. (1964). Analysis of the ego-stage development of school-age children. *The Journal of Experimental Education*, 32(3), 249-257.
5. De Raad, B. and Schouwenburg, H.C. (1996). Personality in learning and education: A review. *European Journal of Personality*, 10(5), 303-336.
6. Elliott, S.N. and Travers, J.F. (1996). *Educational psychology: Effective teaching, effective learning*. Madison, WI: Brown & Benchmark.
7. Freiberg, H.J. (2005). *School climate: Measuring, improving and sustaining healthy learning environments*. Routledge.
8. Griggs, R.A., Jackson, S.L., Christopher, A.N. and Marek, P. (1999). Introductory psychology textbooks: An objective analysis and update. *Teaching of Psychology*, 26(3), 182-189.
9. Keller, J.M. (1999). Using the ARCS motivational process in computer-based instruction and distance education. *New Directions for Teaching and Learning*, 1999(78), 37-47.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 265-270

SOME PREDICTIVE ESTIMATORS OF FINITE POPULATION MEAN IN DOUBLE SAMPLING

Huma Dastgir[1], **Asifa Kamal**[2] and **Mahnaz Makhdum**[3]
Lahore College for Women University, Lahore, Pakistan
Email: [1]huma.dastgir@gmail.com
[2]asifa.k53@gmail.com
[3]minz.mak@hotmail.com

ABSTRACT

This study is based on the development of two exponential type estimators of finite population mean in two-phase sampling scheme utilizing additional information under Predictive Estimation Approach. Following Singh and Majhi (2014) exponential estimator, some modification is made and by assuming different relations among study and auxiliary variables, two exponential type predictive estimators are proposed. The goodness of the pro-posed estimators is discussed by deriving the biases and mean square errors, and also the theoretical conditions for which the proposed estimators perform better than some competing estimators. The generalized forms of the proposed predictive estimators are also developed and the expressions for biases and optimum mean square errors of generalized estimators are derived. For comparison purpose, Absolute biases relative to populations mean and percentage efficiency relative to mean per unit estimator are also calculated. The results provide enough evidence that proposed estimators perform better than all competing estimators, hence it is concluded that proposed estimators perform better than competing estimators, when used under predictive approach in two-phase sampling in the presence of two supplementary variables.

1. INTRODUCTION AND METHODOLOGY

Let the finite population $U = U_1, U_2, \dots, U_N$ consists of N units. Let Y be the study variable whose mean is to be estimated, and X and Z be two auxiliary variables which are related with study variable. Let y_i, x_i and z_i $(1 \le i \le N)$ be the ith values of variables Y, X and Z respectively.

In two-phase sampling mechanism, the information about population is gathered in two phases such that a sample of size $\acute{n}$ (say) is selected from a population of size N at first phase, and then a sub-sample of size n $(n < \acute{n})$ is selected randomly from the first phase sample. Then $\bar{y}, \bar{x}$ and $\bar{z}$ be the means of second phase sample values, where $\acute{\bar{x}}$ and $\acute{\bar{z}}$ be the means of first phase sample values.

Under prediction approach of estimation, let $\acute{s}$ and s be the collection of units in the first and second phase samples of size n and $\acute{n}$ respectively. Let $\bar{s}$ be the collection of those units of U which are not included in s, such that $\bar{s}$ = U- s.

Let r_1 and r_2 be the decomposition of $\bar{s}$ such that r_1 = U - $\acute{s}$ and $r_2 = \bar{s} \cap \acute{s}$. i.e. unobserved units of population at first and second phase of sampling, respectively. Under this predictive approach, population mean $\bar{Y}$ can be written as following:

$$\bar{Y} = \frac{1}{N}\sum_{i\in s} Y_i + \frac{1}{N}\sum_{i\in r_2} Y_i + \frac{1}{N}\sum_{i\in r_1} Y_i \tag{1}$$

where s, r_1 and r_2 be the decomposition of U into three mutually exclusive sets. Thus, the equation (1) can be written as:

$$\bar{Y} = f\bar{y} + \left(\acute{f} - f\right)\bar{Y}_2 + (1 - \acute{f})\bar{Y}_1$$

where $f = \frac{n}{N}$ and $\acute{f} = \frac{\acute{n}}{N}$.

The value of $\bar{y}$ is exactly known as this is the observed part of the population. However, the problem is to predict $\overline{Y_1}$ and $\overline{Y_2}$ only. Let K_1 and K_2 be the predictors of $\overline{Y_1}$ and $\overline{Y_2}$, respectively. Then an estimator of $\bar{Y}$ in two-phase sampling, under predictive approach can be written as follows:

$$\bar{Y} = f\bar{y} + \left(\acute{f} - f\right)K_2 + \left(1 - \acute{f}\right)K_1 \tag{2}$$

For different choices of K_1 and K_2, equation(2) is used to develop predictive estimators of finite population mean in double sampling.

2. PROPOSED PREDICTIVE ESTIMATORS

Following Sahoo and Sahoo (2001) two-phase predictive approach and motivated by Singh and Majhi (2014), the following exponential estimators of finite population mean under predictive estimation approach in two-phase sampling scheme are proposed.

Estimator I:

Assuming X has an exponential relation with study variable Y in both phases, and both X and Z are positively correlated with study variable Y, the proposed predictive estimator with two auxiliary variables is as following:

$$\bar{\tilde{\tilde{Y}}}^{e}_{1(2)pred.} = f\bar{y} + \left(\acute{f} - f\right)\bar{y}\exp\left(\frac{\bar{X}_2 - \bar{x}}{\bar{X}_2 + \bar{x}}\right) + \left(1 - \acute{f}\right)\bar{y}\,\frac{\bar{Z}_1}{\bar{\acute{z}}}\exp\left(\frac{\bar{\acute{x}} - \bar{x}}{\bar{\acute{x}} + \bar{x}}\right) \tag{3}$$

Estimator II:

Assuming Z has an exponential relation with study variable Y only in first phase of sampling, while X has a positive correlation with Y in both phases of sampling, the proposed predictive estimator is as following:

$$\bar{\tilde{\tilde{Y}}}^{e}_{2(2)pred.} = f\bar{y} + \left(\acute{f} - f\right)\bar{y}\frac{\bar{X}_2}{\bar{x}} + (1-\acute{f})\bar{y}\,\frac{\bar{\acute{x}}}{\bar{\bar{x}}}\exp\left(\frac{\bar{Z}_1 - \bar{\acute{z}}}{\bar{\bar{Z}}_1 + \bar{\acute{z}}}\right) \tag{4}$$

where $\bar{X}_1$ and $\bar{Z}_1$ the predictive means of non-sampled population values at first phase of sampling and $\bar{X}_2$ is the predictive mean of non-sampled population values at second phase of sampling, are given as:

$$\bar{X}_1 = \frac{N\bar{X} - \acute{n}\bar{\acute{x}}}{N - \acute{n}}$$

$$\bar{Z}_1 = \frac{N\bar{Z} - \acute{n}\bar{\acute{z}}}{N - \acute{n}}$$

$$\bar{X}_2 = \frac{\acute{n}\bar{\acute{x}} - n\bar{x}}{\acute{n} - n}$$

2.1 Biases and Mean Square Errors of Proposed Estimators

The biases and mean square errors of proposed estimators to the first order of approximation are obtained as:

$$\text{Bias}\,(\tilde{\bar{\bar{Y}}}^{e}_{1(2)pred.}) = \bar{Y}\left[\o\, C_x^2\left\{\frac{\left(3\acute{f} - 3f - \acute{f}f\right)}{8\left(f\acute{-}f\right)} - \frac{D_{yx}}{2}\right\}\right.$$
$$\left. + \acute{\o}\, C_z^2\,\{\acute{f} - D_{yz}\} + \acute{\o}\, C_x^2\left\{\frac{D_{yx}}{2} - \frac{3[\acute{f}^2 + (1-\acute{f})(\acute{f}-f)]}{8(f\acute{-}f)}\right\}\right] \quad (5)$$

and

$$\text{MSE}(\tilde{\bar{\bar{Y}}}^{e}_{1(2)pred.}) = \bar{Y}^2[\o C_y^2 + (\emptyset - \acute{\emptyset})C_x^2\{\tfrac{1}{4} - -D_{yx}\} + \acute{\o}C_z^2\{1 - 2D_{yz}\}] \quad (6)$$

Similarly,

$$\text{Bias}(\tilde{\bar{\bar{Y}}}^{e}_{2(2)pred.}) = \bar{Y}\left[\o C_x^2\{f - D_{yx}\} + \acute{\o}\, C_x^2\,\{D_{yx} - 1\} + \acute{\o}\, C_z^2\left\{\frac{(3-4\acute{f})}{8\,(1-\acute{f})} - \frac{D_{yz}}{2}\right\}\right] \quad (7)$$

$$\text{MSE}(\tilde{\bar{\bar{Y}}}^{e}_{2(2)pred.}) = \bar{Y}^2\left[\o\, C_y^2 + (\emptyset - \acute{\emptyset})C_x^2\{1 - 2D_{yx}\} + \acute{\o}C_z^2\{\tfrac{1}{4} - D_{yz}\}\right] \quad (8)$$

where $\emptyset = \frac{1}{n} - \frac{1}{N}$ and $\acute{\emptyset} = \frac{1}{\acute{n}} - \frac{1}{N}$; also $D_{ij} = \rho_{ij}\frac{C_i}{C_j}$ for $i = x, y, z; j = x, y, z (i \neq j)$ and C_i is the coefficient of variation for $\dot{\imath}$th variable.

a. Efficiency Comparison

In this section, the efficiency of proposed estimators is theoretically compared with some already existing estimators of finite population mean with respect to their mean square errors. The proposed estimators are compared with usual mean per unit ($\bar{\boldsymbol{y}}$), Singh and Choudhury (2012) two-phase exponential ratio estimator, Singh and Majhi (2014) exponential type chain ratio estimators, and two-phase predictive estimators proposed by Sahoo and Sahoo (2001).

Table 1 gives the theoretical conditions under which the proposed estimators perform better than competing estimators.

Table 1
Theoretical Conditions

Competing Estimators	Proposed Estimator I	Proposed Estimator II
Mean Per Unit ($\bar{y}$)	$D_{yx} > \frac{1}{4}$; $D_{yz} > \frac{1}{2}$	$D_{yx} > \frac{1}{2}$; $D_{yz} > \frac{1}{4}$
Singh and Choudhary (2012) Exponential Ratio Estimator	$D_{yz} > \frac{3}{4}$	$D_{yx} > \frac{3}{4}$
Singh and Majhi (2014) Exponential Type Chain Ratio Estimator I	$D_{yx} < \frac{3}{4}$; $D_{yz} > \frac{3}{4}$	Equally Efficient
Singh and Majhi (2014) Exponential Type Chain Ratio Estimator II	Equally Efficient	$D_{yx} > \frac{3}{4}$; $D_{yz} < \frac{3}{4}$
Sahoo and Sahoo (2001) Two Phase Predictive Estimator I	$D_{yx} > \frac{1+2f'}{4}$	$D_{yx} > \frac{1+f'}{2}$; $D_{yz} < \frac{3}{4}$
Sahoo and Sahoo (2001) Two Phase Predictive Estimator II	$D_{yx} < \frac{3}{4}$; $D_{yz} > \frac{1-(1-f')^2}{2f'}$	$D_{yz} < \frac{4(1-f')^2 - 1}{4(1-2f')}$

b. Generalized Forms of Proposed Predictive Estimators.

In this section, the generalized forms of proposed predictive estimators are studied. The biases and optimum mean square errors of proposed generalized estimators are derived.

Generalized Estimator I:

The generalized form of the proposed predictive estimator I (3) is given as following. Here α and β are the constants of generalization.

$$\tilde{\tilde{\bar{Y}}}^{e}_{G1(2)pred.} = f\bar{y} + (\acute{f} - f)\bar{y}\exp\left(\alpha\left\{\frac{\bar{X}_2 - \bar{x}}{\bar{X}_2 + \bar{x}}\right\}\right) + (1 - \acute{f})\bar{y}\left(\frac{\bar{Z}_1}{\acute{\bar{z}}}\right)^{\beta}\exp\left(\frac{\acute{\bar{x}} - \bar{x}}{\acute{\bar{x}} + \bar{x}}\right) \tag{9}$$

With

$$Bias\left(\tilde{\tilde{\bar{Y}}}^{e}_{G1(2)pred.}\right) = \bar{Y}\left[\emptyset C_x^2\left\{\frac{\alpha^2\acute{f}^2 + 2\alpha\acute{f}(\acute{f} - 2f) + 3(1 - \acute{f})(\acute{f} - f)}{8(\acute{f} - f)} - \frac{(\alpha\acute{f} + 1 - \acute{f})D_{yx}}{2}\right\}\right.$$
$$\left. + \acute{\emptyset}C_x^2\left\{\frac{(\alpha\acute{f} + 1 - \acute{f})D_{yx}}{2} - \frac{[\acute{f}^2(\alpha^2 + 2\alpha) + 3(1 - \acute{f})(\acute{f} - f)]}{8(\acute{f} - f)}\right\} + \acute{\emptyset}C_x^2\{\beta^2\acute{f} - \beta D_{yz}\}\right] \tag{10}$$

and

$$MSE\left(\tilde{\tilde{\bar{Y}}}^{e}_{G1(2)pred.}\right) = \bar{Y}^2 C_y^2\left[\emptyset(1 - \rho_{yx}^2) + \acute{\emptyset}(\rho_{yx}^2 - \rho_{yz}^2)\right] \tag{11}$$

where

$\hat{\alpha} = \frac{2D_{yx} - (1 - \acute{f})}{\acute{f}}$ and $\hat{\beta} = D_{yz}$

Generalized Form Estimator II:

The generalized form of the proposed predictive estimator II(4) is given as following. Here α and β are the constants of generalization.

$$\tilde{\tilde{\bar{Y}}}^{e}_{G2(2)pred.} = f\bar{y} + (\acute{f} - f)\bar{y}\left(\frac{\bar{X}_2}{\bar{x}}\right)^{\alpha} + (1 - \acute{f})\bar{y}\frac{\acute{\bar{x}}}{\bar{x}}\exp\left\{\beta\left(\frac{\bar{Z}_1 - \acute{\bar{z}}}{\bar{Z}_1 + \acute{\bar{z}}}\right)\right\} \tag{12}$$

With

$$Bias\left(\tilde{\tilde{\bar{Y}}}^{e}_{G2(2)pred.}\right) = \bar{Y}\left[\emptyset C_x^2\left\{\alpha^2 f - (1 - \acute{f} + \alpha\acute{f})D_{yx}\right\}\right.$$
$$\left. + \acute{\emptyset}C_z^2\left\{\frac{\beta^2 + 2\beta(1 - 2\acute{f}))}{8(1 - \acute{f})} - \frac{\beta D_{yz}}{2}\right\} + \acute{\emptyset}C_x^2(1 - \acute{f} + \alpha\acute{f})(D_{yx} - 1)\}\right] \tag{13}$$

$$MSE\left(\tilde{\tilde{\bar{Y}}}^{e}_{G1(2)pred.}\right) = \bar{Y}^2 C_y^2\left[\emptyset(1 - \rho_{yx}^2) + \acute{\emptyset}(\rho_{yx}^2 - \rho_{yz}^2)\right] \tag{14}$$

$\hat{\alpha} = \frac{D_{yx} - (1 - \acute{f})}{\acute{f}}$ and $\hat{\beta} = 2D_{yz}$

For different values of α and β, many other predictive estimators can also be obtained.

4. NUMERICAL STUDY

To study the performance of the proposed predictive estimators with the above discussed estimators, numerical study is carried out using five real life data sets. The description of these data sets is given in Table 1.

The absolute relative biases and percentage relative efficiencies of proposed and competing estimators relative to mean per unit estimator are calculated in Table 3 and

Table 4 respectively. The results of Table 3 show that proposed estimator I has less amount of relative bias than all competing estimators for all populations except for population V but it has quite high relative efficiency than all estimators for same population. Similarly, proposed estimator II has high relative bias for population II, III and V but its efficiency is quite high than competing estimators when the theoretical conditions are fulfilled.

Table 2
Population Parameters

Parameters	N	$\acute{n}$	n	$\bar{Y}$	C_y	C_x	C_z	ρ_{yx}	ρ_{yz}	ρ_{xz}
Cochran(1977) Pop I	34	15	10	4.92	1.012	1.232	1.072	.7326	.6430	.6837
Fisher(1936) Pop II	50	18	8	2.03	.1342	.1073	.0984	.5377	.3221	.4010
Fisher(1936) Pop III	50	20	10	2.77	.1122	.0857	.1092	.5259	.5605	.754
Shukla(1966) Pop IV	50	15	8	2.58	.2943	.3410	.1304	.48	.37	.73
Anderson(1958) Pop V	25	10	7	183.84	.0546	.0526	.0488	.7108	.6932	.7346

Table 3
Absolute Relative Biases

Estimators	Pop I	Pop II	Pop III	Pop IV	Pop V
Mean per unit ($\bar{y}$)	0	0	0	0	0
Singh and Choudhary (2012) Exponential Ratio Estimator	.037258	.000622	.000295	.003915	.000864
Singh and Majhi (2014) Exponential Type Chain Ratio Estimator I	.023196	.000315	.000146	.003939	.0000293
Singh and Majhi (2014) Exponential Type Chain Ratio Estimator II	.020562	.000224	.000163	.00127	.0000328
Sahoo and Sahoo (2001) Two Phase Predictive Estimator I	.032546	.000242	.000169	.002187	.000118
Sahoo and Sahoo (2001) Two Phase Predictive Estimator II	.055018	.000505	.000287	.003891	.000208
Proposed Predictive Estimator I	.0033889	.000084	.000114	.000506	.0001442
Proposed Predictive Estimator II	.005666	.000724	.000354	.000636	.0001875

Table 4
Percentage Relative Efficiencies

Estimators	Pop I	Pop II	Pop III	Pop IV	Pop V
Mean per unit ($\bar{y}$)	100	100	100	100	100
Singh and Choudhary (2012) Exponential Ratio Estimator	99.87	100.5	119.88	89.51	104.72
Singh and Majhi (2014) Exponential Type Chain Ratio Estimator I	154.81	122.0	133.91	*	175.09
Singh and Majhi (2014) Exponential Type Chain Ratio Estimator II	159.50	118.56	127.19	122.09	180.35
Sahoo and Sahoo (2001) Two Phase Predictive Estimator I	156.80	114.66	124.44	121.46	173.38
Sahoo and Sahoo (2001) Two Phase Predictive Estimator II	156.12	121.02	134.24	*	*
Proposed Predictive Estimator I	159.50	*	*	122.09	180.35
Proposed Predictive Estimator II	*	122.0	133.91	*	*

*Conditions are not fulfilled.

5. CONCLUSION

It has been shown from the results that proposed estimators outrun the competing estimators and are more efficient when the theoretical conditions are satisfied.

REFERENCES

1. Anderson, T.W. (1958). *An introduction to multivariate statistical analysis*, Volume 2. Wiley New York.
2. Cochran, W.G. (1977). *Sampling techniques-3*. New York, John Wiley and Sons.
3. Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Human Genetics*, 7(2), 179-188.
4. Sahoo, L. and Sahoo, R. (2001). Predictive estimation of finite population mean in two-phase sampling using two auxiliary variables. *Indian Society of Agricultural Statistics (Inida)*, 54(2), 258-264.
5. Shukla, G. (1966). An alternative multivariate ratio estimate for finite population. *Calcutta Statistical Association Bulletin*, 15(2-3), 127-134.
6. Singh, B. and Choudhury, S. (2012). Exponential chain ratio and product type estimators for finite population mean under double sampling scheme. *Journal of Science Frontier Research in Mathematics and Design Sciences*, 12(6), 12-24.
7. Singh, G. and Majhi, D. (2014). Some chain-type exponential estimators of population mean in two-phase sampling. *Stat. Trans.*, 15(2), 221-230.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 271-278

CHARACTERIZATIONS OF SOME KUMARASWAMY GENERATED PROBABILITY DISTRIBUTIONS

Sharqa Hashmi and **Zarmina Gull**
Department of Statistics, Lahore College for Women University
Lahore, Pakistan
Email: sharqa.hashmi@lcwu.edu.pk

ABSTRACT

The characterizations of few Kumaraswamy-G distributions are presented in paper. The characterizations are based on two methods:

(i) by the ratio of two truncated moments
(ii) by hazard function.

The special cases for these Kumaraswamy-G families of distribution are also given with their characterization expressions based on ratio of two truncated moments.

1. INTRODUCTION

During the last few decades, several ways of extending well-known distributions have been introduced. The one of them is the class of distributions generated by Kumaraswamy (1980) distribution defined by Cordeiro and de Castro (2011).

The characterization of a distribution is an important aspect which can help the researcher to see if proposed model is the correct one. This research work is based on two types of characterizations of probability distribution:

(i) by the ratio of two truncated moments and (ii) by hazard rate function. These characterizations are presented for Kumaraswamy exponential Weibull distribution (Cordeiro, et al. 2016), Kumaraswamy pareto distribution (Pereira, et al. 2012), Kumaraswamy inverse Weibull distribution (Shahbaz et al. 2012).

Cordeiro, et al. (2016) derived Kumaraswamy exponential Weibull distribution with probability density function given as,

$$f(x) = \alpha\gamma[\lambda + k\beta x^{k-1}]e^{-\lambda x-\beta x^k}\left(1-e^{-\lambda x-\beta x^k}\right)^{\alpha-1} \left\{1-\left(1-e^{-\lambda x-\beta x^k}\right)^{\alpha}\right\}^{\gamma-1}, x>0 \tag{1}$$

its corresponding cumulative distribution function is

$$F(x) = 1-\left\{1-\left(1-e^{-\lambda x-\beta x^k}\right)^{\alpha}\right\}^{\gamma}. \tag{2}$$

where α,β,γ and k are positive parameters.

Bourguignon, et al. (2012) derived Kumaraswamy pareto distribution with probability density function given by,

$$f(x) = \frac{abk\beta^k}{x^{k+1}}\left\{1-\left(\frac{\beta}{x}\right)^k\right\}^{a-1}\left[1-\left\{1-\left(\frac{\beta}{x}\right)^k\right\}^a\right]^{b-1}, x > 0. \tag{3}$$

and corresponding cumulative distribution function is

$$F(x) = 1-\left[1-\left\{1-\left(\frac{\beta}{x}\right)^k\right\}^a\right]^b, \tag{4}$$

where β,k, a and b are positive parameters.

Shahbaz et al. (2012) introduced Kumaraswamy inverse Weibull distribution with probability density function given as,

$$f(x) = \frac{ab\alpha\beta}{x^{\beta+1}} e^{\left(-\frac{a\alpha}{x^\beta}\right)}\left\{1-e^{\left(-\frac{a\alpha}{x^\beta}\right)}\right\}^{b-1}, x > 0. \tag{5}$$

And corresponding cumulative distribution function is

$$F(x) = 1-\left\{1-e^{\left(-\frac{a\alpha}{x^\beta}\right)}\right\}^b \tag{6}$$

where α,a,b and β are positive parameters.

2. CHARACTERIZATIONS

The characterization is presented in subsections. This section deals with the characterization of probability distributions based on ratio of two truncated moments.

2.1 Characterization based on Two Truncated Moments

This subsection includes the characterization of Kumaraswamy-G probability distributions including Kumaraswamy exponential Weibull distribution (KEW), Kumaraswamy pareto distribution (KP), Kumaraswamy inverse Weibull distribution (KIW). The first characterization employs the theorem given in Glänzel (1987, 1990). Hamedani (2016, 2017) also presented characterizations of some distributions.

Proposition 2.1.1

Assume X: Ω → ℝ be the continuous r. v. and let

$$h(x) = \left(1-e^{-\lambda x-\beta x^k}\right)^{1-\alpha}\left\{1-\left(1-e^{-\lambda x-\beta x^k}\right)^{\alpha}\right\}^{1-\gamma}$$

and $\quad g(x) = h(x)\, e^{-\lambda x-\beta x^k}$

for $x > 0$. Then, X has pdf of Kumaraswamy Exponential Weibull Distribution (1) and if the function ξ which is defined in theorem (Glänzel, 1987,1990) is of form

$$\xi(x)=\frac{1}{2}\{e^{-\lambda x-\beta x^k}\}$$

Proof:

Assume X be random variable with (1),

$$\mathrm{E}[\mathrm{h(x)}|\mathrm{X} \geq \mathrm{x}] = \int_x^\infty h(x) f(x|\mathrm{X} \geq \mathrm{x}) dx$$

$$(1\text{-}\mathrm{F(x)})\ \mathrm{E}[\mathrm{h(x)}|\mathrm{X} \geq \mathrm{x}]$$

$$= \int_x^\infty \alpha\gamma[\lambda + k\beta x^{k-1}]e^{-\lambda x-\beta x^k}\left(1 - e^{-\lambda x-\beta x^k}\right)^{\alpha-1+1-\alpha}$$

$$\left\{1 - \left(1 - e^{-\lambda x-\beta x^k}\right)^{\alpha}\right\}^{\gamma-1+1-\gamma} dx$$

$$(1\text{-}\mathrm{F(x)})\ \mathrm{E}[\mathrm{h(x)}|\mathrm{X} \geq \mathrm{x}] = \alpha\gamma e^{-\lambda x-\beta x^k}$$

and

$$(1\text{-}\mathrm{F(x)})\ \mathrm{E}[\mathrm{g(x)}|\mathrm{X} \geq \mathrm{x}] = \frac{\alpha\gamma}{2}\left\{e^{-2\lambda x-2\beta x^k}\right\}$$

Further,

$$\xi(x)h(x) - g(x) = -\frac{h(x)}{2}\left[e^{-\lambda x-\beta x^k}\right]$$

Contrarily, if ξ is of form given above, then

$$s'(x) = \frac{\xi'(x)h(x)}{\xi(x)h(x)-g(x)} = \lambda + k\beta x^{k-1}$$

By integrating above we have

$$s(x) = \lambda x + \beta x^k$$

Alternatively

Let X: $\Omega \rightarrow (0, \infty)$ be a continuous random variable and assume $h(x)$ given in Proposition (1). The continuous random variable X has pdf (1) if and only if there exist functions ξ and g defined in Proposition satisfying the following differential equation

$$\frac{\xi'(x)h(x)}{\xi(x)h(x) - g(x)} = \lambda x + \beta x^k$$

The general solution of above differential equations

$$\xi(x) = \left\{e^{-\lambda x-\beta x^k}\right\}^{-1}\int -g(x)\,(h(x))^{-1}[\lambda x + \beta x^k]e^{-\lambda x-\beta x^k}dx + D$$

where D is constant of integration.

Proposition 2.1.2

Assume X: $\Omega \rightarrow \mathbb{R}$ be the continuous random variable and let

$$h(x) = \left[1 - \left\{1 - \left(\frac{\beta}{x}\right)^k\right\}^a\right]^{1-b}$$

and

$$g(x) = h(x)\left\{1 - \left(\frac{\beta}{x}\right)^k\right\}^a$$

for $x > 0$. Then, X has pdf of Kumaraswamy Pareto Distribution (3) and if the function ξ which is defined in theorem (Glänzel, 1987) is of form

$$\xi(x)=\frac{1}{2}\left\{1+\left\{1-\left(\frac{\beta}{x}\right)^{k}\right\}^{a}\right\}.$$

Alternatively

Let X: Ω → (0, ∞) be a continuous random variable and assume $h(x)$ given in Proposition (2). The continuous random variable X has pdf (3) if and only if there exist functions ξ and g defined in Proposition (2) satisfying the following differential equation

$$\frac{\xi'(x)h(x)}{\xi(x)h(x)-g(x)}=\frac{ak\left(\beta^{k}/x^{k+1}\right)\left\{1-\left(\frac{\beta}{x}\right)^{k}\right\}^{a-1}}{\left[1-\left\{1-\left(\frac{\beta}{x}\right)^{k}\right\}^{a}\right]}$$

The general solution of above differential equations

$$\xi(x)=\left\{\left[1-\left\{1-\left(\frac{\beta}{x}\right)^{k}\right\}^{a}\right]\right\}^{-1}$$
$$\int -g(x)\,(h(x))^{-1}\,ak\left(\beta^{k}/x^{k+1}\right)\left\{1-\left(\frac{\beta}{x}\right)^{k}\right\}^{a-1}dx+D$$

where D is constant of integration.

Proposition 2.1.3

Assume X: Ω → ℝ be the continuous random variable and let

$$h(x)=\left\{1-e^{\left(-\frac{a\alpha}{x^{\beta}}\right)}\right\}^{1-b}$$

and

$$g(x)=h(x)e^{\left(-\frac{a\alpha}{x^{\beta}}\right)}$$

for $x > 0$. Then, X has pdf of Kumaraswamy Inverse Weibull Distribution (5) and if the function ξ is of form

$$\xi(x)=\frac{1}{2}\left\{e^{\left(-\frac{a\alpha}{x^{\beta}}\right)}\right\}$$

Alternatively:

Let X: Ω → (0, ∞) be a continuous random variable and assume $h(x)$ given in Proposition (3). The continuous random variable X has pdf (5) if and only if there exist functions ξ and g defined in Proposition (3) satisfying the following differential equation

$$\frac{\xi'(x)h(x)}{\xi(x)h(x)-g(x)}=\frac{-a\alpha\beta}{x^{\beta+1}}$$

The general solution of above differential equations

$$\xi(x) = \left\{e^{\left(-\frac{a\alpha}{x^{\beta}}\right)}\right\}^{-1} \int g(x)\,(h(x))^{-1} \frac{a\alpha\beta}{x^{\beta+1}} e^{\left(-\frac{a\alpha}{x^{\beta}}\right)} dx + D$$

where D is constant of integration.

2.2 Special Cases of Kumaraswamy-G Distributions

2.2.1 Special Cases of Kumaraswamy Exponential Distribution

We provide here the expressions of $g(x)$, $h(x)$ and $\xi(x)$ to characterize of the following distributions.

1. **Kumaraswamy Modified Weibull Distribution:** By putting $\lambda = 0$ in pdf (1), following expressions are obtained:

   $$h(x) = \left(1 - e^{-\beta x^k}\right)^{1-\alpha} \left\{1 - \left(1 - e^{-\beta x^k}\right)^{\alpha}\right\}^{1-\gamma}, \text{ for } x > 0$$

   and

   $$g(x) = h(x) e^{-\beta x^k},$$

   The random variable X follows Kumaraswamy Modified Weibull Distribution if and only if the function ξ defined in Theorem (see Appendix) is of the form

   $$\xi(x) = \frac{1}{2}\{e^{-\beta x^k}\}.$$

2. **Exponential Weibull Distribution:** By putting α=1 and γ=1in pdf (1), following expressions are obtained:

   $$h(x) = 1, \quad \text{for } x > 0$$

   and

   $$g(x) = h(x) e^{-\lambda x - \beta x^k},$$

   The random variable X follows Exponential Weibull Distribution if and only if the function ξ defined in Theorem (see Appendix) is of the form $\xi(x) = \frac{1}{2}\{e^{-\lambda x - \beta x^k}\}$.

3. **Modified Exponential Distribution:** By putting α=1, γ=1 and k=1in pdf (1), following expressions are obtained:

   $$h(x) = 1, \text{ for } x > 0$$

   and

   $$g(x) = h(x) e^{-\lambda x - \beta x},$$

   The random variable X follows Modified Exponential Distribution if and only if the function ξ defined in Theorem (see Appendix) is of the form $\xi(x) = \frac{1}{2}\{e^{-\lambda x - \beta x}\}$.

4. **Modified Rayleigh Distribution:** By putting α=1, γ=1 and k=2 in pdf (1), following expressions are obtained:

$h(x) = 1$, for $x > 0$

and

$g(x) = h(x)e^{-\lambda x-\beta x^2}$,

The random variable X follows Modified Rayleigh Distribution if and only if the function ξ defined in Theorem (see Appendix) is of the form $\xi(x) = \frac{1}{2}\{e^{-\lambda x-\beta x^2}\}$.

2.2.2 Special Cases of Kumaraswamy Pareto Distribution

We provide here the expressions of $g(x)$, $h(x)$ and $\xi(x)$ to characterize of the following distributions.

Pareto Distribution: By putting a=b=1 in pdf (3), following expressions are obtained:

$h(x) = 1$, for $x > 0$

and

$g(x) = h(x)\left\{1-\left(\frac{\beta}{x}\right)^k\right\}$,

The random variable X follows Pareto Distribution if and only if the function ξ defined in Theorem (see Appendix) is of the form $\xi(x) = \frac{1}{2}\left\{1+\left\{1-\left(\frac{\beta}{x}\right)^k\right\}^1\right\}$.

2.2.3 Special cases of Kumaraswamy Inverse Weibull Distribution

We provide here the expressions of $g(x)$, $h(x)$ and $\xi(x)$ to characterize of the following distributions.

Inverse Weibull Distribution: By putting a=b=1 in pdf (5), following expressions are obtained:

$h(x) = 1$, for $x > 0$

and

$g(x) = h(x)e^{\left(-\frac{\alpha}{x\beta}\right)}$,

The random variable X follows Inverse Weibull Distribution if and only if the function ξ defined in Theorem (see Appendix) is of the form $\xi(x) = \frac{1}{2}\left\{e^{\left(-\frac{\alpha}{x\beta}\right)}\right\}$.

2.3 Characterization based on Hazard Function

The characterization based on hazard function is presented by Hamedani (2016, 2017) and it involves the hazard function and the derivative of hazard function $h(x)$ and $h'(x)$ respectively. The hazard function is as follows

$h_F(x)=\frac{f(x)}{1-F(x)}$, x ϵ support of F.

Proposition 2.3.1

Assume X: $\Omega \rightarrow \mathbb{R}$ be the continuous random variable with pdf of Kumaraswamy Exponential (1). The X has pdf (1) if and only if its hazard function $h_F(x)$ which satisfies the differential equation

$$h_F'(x) - \frac{\beta k(k-1)x^{-1}}{[\lambda + k\beta x^{k-1}]} h_F(x) = \alpha\gamma[\lambda + k\beta x^{k-1}] \frac{d}{dx} \left\{ e^{-\lambda x - \beta x^k} \left(1 - e^{-\lambda x - \beta x^k}\right)^{\alpha-1} \left\{1 - \left(1 - e^{-\lambda x - \beta x^k}\right)^{\alpha}\right\}^{-1} \right\}, x > 0.$$

Proposition 2.3.2

Assume X: $\Omega \rightarrow \mathbb{R}$ be the continuous random variable with pdf of Kumaraswamy Pareto (3). The X has pdf (3) if and only if its hazard function $h_F(x)$ which satisfies the differential equation

$$h_F'(x) - (-k-1)x^{-1} h_F(x) = \frac{abk\beta^k}{x^{k+1}} \frac{d}{dx} \left\{ \left\{1 - \left(\frac{\beta}{x}\right)^k\right\}^{a-1} \left[1 - \left\{1 - \left(\frac{\beta}{x}\right)^k\right\}^{a}\right]^{-1} \right\}, x > 0.$$

Proposition 2.3.3

Assume X: $\Omega \rightarrow \mathbb{R}$ be the continuous random variable with pdf of Kumaraswamy Inverse Weibull (5). The X has pdf (5) if and only if its hazard function $h_F(x)$ which satisfies the differential equation

$$h_F'(x) - (-\beta-1)x^{-1} h_F(x) = \frac{ab\alpha\beta}{x^{\beta+1}} \frac{d}{dx} \left\{ e^{\left(-\frac{a\alpha}{x^{\beta}}\right)} \left\{1 - e^{\left(-\frac{a\alpha}{x^{\beta}}\right)}\right\}^{-1} \right\}, x > 0.$$

3. CONCLUSION

The main objective of study is to help researcher in choosing the specific model whether it fits the underlying requirement or not. An investigator can rely on characterizations of chosen distribution. The recent work deals with the characterization of three distributions in two different directions. The specials cases with their characterization expressions for ratio of truncated moments are also mentioned.

REFERENCES

1. Cordeiro, G.M., Saboor, A., Khan, M.N., Ozel, G. and Pascoa, M.A. (2016). The Kumaraswamy Exponential–Weibull Distribution: Theory and Applications. *Hacettepe Journal of Mathematics and Statistics*, 45(4), 1203-1229.
2. Glänzel, W. (1987). A characterization theorem based on truncated moments and its application to some distribution families. *In Mathematical statistics and probability theory* (pp. 75-84). Springer, Dordrecht.
3. Glänzel, W. (1990). Some consequences of a characterization theorem based on truncated moments. *Statistics*, 21(4), 613-618.

4. Hamedani, G.G. and Najibi, S.M. (2016). Certain Characterizations of Recently Introduced Distributions. *Pakistan Journal of Statistics and Operation Research*, 12(4), 551-577.
5. Hamedani, G.G. (2017). Characterizations of Kumaraswamy-Laplace, McDonald Inverse Weibull and New Generalized Exponential Distributions. *Pakistan Journal of Statistics and Operation Research*, 13(4), 727-737.
6. Hamedani, G.G., Cordeiro, G.M., Lima, M.C.S. and Nascimento, A.D.C. (2017). Some Extended Classes of Distributions: Characterizations and Properties. *Pakistan Journal of Statistics and Operation Research*, 13(4), 893-908.
7. Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46(1-2), 79-88.
8. Shahbaz, M.Q., Shahbaz, S. and Butt, N.S. (2012). The kumaraswamy–inverse weibull distribution. *Pakistan Journal of Statistics and Operation Research*, 8(3), 479-489.
9. Pereira, M.B., Silva, R.B., Zea, L.M. and Cordeiro, G.M. (2012). *The kumaraswamy Pareto distribution*. arXiv preprint arXiv:1204.1389.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 279-300

LAD-LASSO AND WLAD-LASSO FOR VARIABLE SELECTION IN THE PRESENCE OF OUTLIERS

Anam Zakir[1] and **Sohail Chand**[2]

[1] Department of Mathematics and Statistics
Virtual University of Pakistan, Lahore, Pakistan.
Email: anam.zakir@vu.edu.pk

[2] College of Statistical and Actuarial Sciences
University of the Punjab, Lahore, Pakistan
Email: sohail.stat@gmail.com

ABSTRACT

LAD (least absolute deviation) and WLAD (weighted least absolute deviation) are used to deal with the outliers. LASSO is an attractive shrinkage method for simultaneous variable selection. Different methods like AIC, BIC, LAD, LAD-LASSO, WLAD and WLAD-LASSO have been compared through simulation studies. In this paper, model accuracy and its prediction accuracy of these shrinkage methods have been discussed. This comparison based on different settings of signal to noise ratio, sample sizes and contamination rate in the dataset. Selection of tuning parameter has been made using BIC-type criterion.

The results of simulation study depict that LAD-LASSO performs better results when there are outliers in response variables only. If there is contamination in the data set along with the outliers in response variables, WLAD-LASSO shows comparatively better results. In general, the pattern shows that as the sample size increases the methods show an improved results and same happens when the noise ratio is decreased. Real-life applications of these methods are studied to illustrate the practical procedure.

INTRODUCTION

Outliers in the data may cause many problems. Most importantly, they tend to produce biased results in the process of parameter estimation. The process of parameter estimation is usually carried out with simple methods of estimating least squares regression. However, it is examined in the literature that in a situation where there are outliers (for details see section 1.2), ordinary least squares method failed to produce unbiased estimators (Hastie et al., 2009).

In a regression analysis, it is not necessary to ignore the key explanatory variable because it may produce biased estimates and can be very sensitive to the prediction results. The robust regression technique is used as an alternative to the OLS (Ordinary least squares). An interesting approach called as LAD (least absolute deviation) regression is used to minimize the sum of the absolute value of the error. There is abundant literature on the theoretical properties and calculation methods (LAD) for estimators. (See for example Hastie et al., 2009 and references there in). Numerous studies have focused on "Little p, large n" Settings.

The LAD method is among one of the special case of M-regression method which is commonly used as a heavy-tailed distribution of error. However, if the but if the explanatory variables contain outliers (for details see section 1.3.2), then it effects the overall performance of LAD estimators and these estimators are very sensitive to the influence points, such as an individual leverage point possibly may seriously disturb LAD regression estimator. To cope up with the extreme values of explanatory variables in parameter estimation, weighted LAD was proposed and discussed by Ellis and Morgenthaler, (1992). In weighted LAD process, the weights totally depend on the explanatory variables and are convenient to down the weights of the leverage points and eventually it will lessen the overall effect of the outliers in the process of estimation.

In the regression analysis, the Variable selection is most important. During initial phase, many variables are generally introduced to the regression model to mitigate potential bias modeling. Meanwhile, including such useless predictors, can eventually reduce the effectiveness of the estimation technique and produces less accurate forecasts. Meanwhile, choosing important predictors is another essential task. Several methods, such as, the Akaike information criterion (AIC), the Bayesian Information Criterion (BIC) and Mallows Cp, were proposed and established in this regard.

For the model selection and estimation of regression models, the least absolute shrinkage and selection operator (LASSO) was proposed by Tibshirani (1996). The LASSO method stands to utilize variable selection and parameter estimation at the same time and both methods are incorporated in the problem of minimization. Though, the LASSO is a special case of the least squares regression with the L_1-penalized function which suffers from heavy outliers or outliers in the data queue. Hence, LASSO is not impervious to outliers or heavy-tailed error distributions.

Recently, LAD and LASSO regression methods both were combined (the regression method LAD-LASSO) to conduct a robust parameter estimation and selection of variables concurrently. However, it is a fact that the parameter estimation method LAD is only resistant to outliers in the response variable, but cannot resist the leverage points. Weighted LAD, a form of the LASSO method is to find robust regression estimators and select appropriate predictors. The WLAD LASSO weights criterion on leverage points and can be impervious to extreme values of the response variable for regression so resulting estimators are less impressionable to leverage points and outliers.

1.1 Regression Analysis

Regression analysis is a statistical process which is used to find relationship between two or more variables. It is extensively used in various disciplines such as biological, social and behavioral sciences. For forecasting purposes, it is also broadly used in the business sector. It may be utilized for some other applications such as description and control. The regression model structure follows an approach which involves data preparation and collection, removal of the redundant predictors from the data, improvement of the model, finally selection and validation of the model (Kutner et al., 2004).

A regression model that deals with more than one independent variable is referred to as a multiple regression model.

There might be "p" independent variables for one response (y) variable. The model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \varepsilon_i \quad i = 1, 2, ..., n \tag{1.1}$$

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i \ , \ i = 1, 2, ..., n \tag{1.2}$$

is defined as multiple regression model with p predictors. And n is the sample size The parameter βo is called slope and βj, j=1, 2,…,p are called partial regression coefficients because it represents the expected change in dependent variable (y) when there is per unit change in x_{ij} and all other independent variables held constant. Furthermore, ε_i is known as random error term which follows the assumptions as the error term ε_i has mean zero i.e., $E(\varepsilon_i) = 0$. It has a constant variance σ^2. All the error terms are uncorrelated. There is another very important assumption regarding the distribution of error term that it follows normal distribution with mean zero and variance σ^2 i.e., $\varepsilon_{\mathrm{i}} \sim N\left(0 \ , \ \sigma^2\right)$. If any of these assumptions are violated, the predictions and inferences produced by such regression model will be highly misleading and inefficient (Montgomery et al., 2001).

Up till now, multiple linear regression model has been thoroughly discussed. The simple linear regression model deals with only one parameter. In this case, a response variable takes the effect of single predictor. It can be defined as

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ , \ i = 1, 2, ..., n \tag{1.3}$$

The Model (1.3) follows the same assumptions as followed by the above Model (1.1). The Model contains only one predictor (1.3) and all parameters are not in exponent form. As well as parameters are not multiplied or divided by other parameter. It shows that the independent variable is in linear form. As the Model (1.3) is simple and linear in independent variables and parameters so it is named as simple linear regression model. (Kutner et al., 2004).

Furthermore, it is assumed that data are centered without loss of generality, β_0 an intercept term which can be excluded $(\beta_0 = 0)$ from the developed model (Tibshirani, 1996). So, with the exclusion of intercept, the matrix form of the model (1.3) is given as:

$$y = X\beta + \varepsilon \ , \tag{1.4}$$

where

$$\mathrm{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} \quad \mathrm{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}^T = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n\times p} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}_{p\times 1} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}_{n\times 1}$$

The least squares method helps to evaluate unknown parameters which is the most popular method. This method helps in minimizing the sum of squares of residual. This technique estimates the complete model even when the model has some exactly zero pentameters i.e., there is a sparse representation in the model. When the true model is sparse, the least square method generates low bias estimation and large prediction variance. The main reason behind these large prediction variance is because it always generates non-zero estimates (Hastie et al., 2009). In the regression, when the error term doesn't follow the normal pattern or there exist outliers in the data then ordinary least square estimates provide highly incorrect and biased results. The ordinary least square estimates cannot handle prediction accuracy which let the model difficult in interpretation. (1.2)

1.2 Outliers

The Outliers are those observations in the data which usually have very large or small values as compared to other observations. The regression outlier is determined by Rousseeuw and Leroy (1987), according to them the point which deviates from the linear relationship can be driven subtracting that point from all other observations, or at least most of these points. There are for two aspects of outliers. One aspect is that outliers can adversely affect data analysis as they generally increase the error variance and eventually affect the statistical test results, they can decrease regularity, and they can produce seriously influenced results. Secondly, there is the possibility of providing useful information about the data by these outliers (Seo, 2002).

1.3 Types of Outliers

There are different types of outliers. Some of them are discussed ahead.

1.3.1 Residual Outlier

This is a point having a large standardized residual or when used in the n observations to suit a model. It can be distinguishing between regression of residual aberrant and aberrant values based on a point may be atypical regression residual without atypical and vice versa (Montgomery et al., 2001).

1.3.2 X- Space outliers (also called leverage points)

This is a point which is only peripheral in relation to the X-coordinates. The x-space outlier can be a regression outlier and/or residual outlier. an influential point, as compare to other observations, has very large impact on the calculated estimates (Chatterjee et al., 1986).

1.3.3 Y-Space outliers (also called vertical outliers)

This is a point that is a device only because it has an unusual detail. The effect on the regression model observation bases on their y coordinates. Therefore, point could be a regression and / or a residual outlier. It is known as a typical vertical outlier (Zaman et al., 2001).

1.3.4 X and Y Space outliers

A point which is peripheral in both coordinates (X-coordinate and Y-coordinate) may be a regression outlier or a residual outlier or maybe both, can have a very small effect or no effect on overall equation. The determining factor is the general configuration of other points.

1.4 Detection of Outliers

To detect outlier, mainly two methods are used (formal and informal) tests. The formal tests are also known as tests of discordancy while the informal tests are known as labeling methods. Tests of discordancy include Grubbs test, quartile method, Dixon’s test, Hampel method and Generalized ESD (extreme studentized variate). Labeling methods include Standard deviation (SD) method, Z-score, Modified Z-score, Tukey’s method (Boxplot), MAD_E method and Median rule (Manoj et al., 2013).

1.5 Robust Regression

Sensitivity of least squares on one or more observations compelled Box and Anderson (1955) to use the word "robust" as part of a study on the effects of deviations from the standard model assumptions. A robust statistical method has come to mean that the procedure always has desirable properties when the model assumptions are not well satisfied. In applications, typically found in data sets which contain heavy tailed errors and leverage points, these outlying points can have a strong impact on the method of least squares, which means they will "Wrench" a regression equation in its direction. A robust regression method is the one that absorbs the effect that would be very influential by using the least squares. In other words, a robust regression tends to leave residuals connected with large outliers and identifying influence points much easier (Stapleton et al., 1995).

1.5.1 Least Absolute Deviation

The LAD (least absolute deviation) is the simplest technique alternative to ordinary least square (OLS) was presented by Roger Joseph in 1757. The Huber (1987) discussed pros and cons of this technique. This technique is mostly well-suited to long tailed distributions e.g., Laplace or Cauchay. Least absolute deviation is also defined as L1-norm criterion which is defined as

$$\min \sum_{i=1}^{n} |r_i|$$

where

$$r_i = y_i - \hat{y}_i$$

i.e., the residual is measured by the difference between the observed and estimated value (Dodge, 1997).

1.5.2 Weighted Least Absolute Deviation

The LAD method is only robust to outliers in y-space and this method is extremely sensitive to the leverage points in the dataset, even a single leverage point can be badly influenced on the estimation process. To overcome this issue another method was proposed which is known as WLAD. It was firstly considered by Ellis and Mergenthaler (1992). It is obtained as

$$\min \sum_{i=1}^{n} w_i |r_i|, \; i = 1, 2, 3, \ldots, n$$

where

$$r_i = y_i - \widehat{y}_i$$

and w_i are the weights used for press down the leverage points and calculated as a robust measure of x_i. Later, Giloni et al., (2006) suggested an updated system of the WLAD regression. The simulation results showed that there is consistent performance when there are leverage points.

1.6 Variable Selection

The selection of variables is one way of selecting the best subset of variables (predictors) in the model development.

In many situations, it is important to select large subset of predictors (Miller, 1990). The basic force behind the choice of variables is possibly to develop a parsimonious model, that can be achieved by removing irrelevant predictors and sustain a simpler model that can be easily interpreted. It is important to do not measure unnecessary predictors to produce an economical model in a sense that it can save time. Other purpose is to eliminate noise because irrelevant predictors are instigated to add noise to the estimation of the parameters of concern and these results in over fitting of the model. Furthermore, to avoid multicollinearity, as it is very useful in avoiding all those predictors which express almost indistinguishable patterns across the sample.

Sometimes regressors may be a linear combination of other independent variable so it provides no information. This approach is highly useful to avoid infinite many solutions if the total observations are fewer than the number of predictors.

For study purpose, usually one or more than one model is considered. We usually consider subset of many variables in the model to enhance the accuracy of the prediction system which is of great concern. It is best to include more predictors in the model for the sake of interpretation purposes (Tibshirani, 1996). Due to collinearity problems, prediction and variable, selection attributes cannot be gratified. The regression coefficients have great impact for the contribution of several correlated variables caused as large variance. The addition of small subsections of variables may produce prejudiced models (Miller, 1990). So, the main determination under consideration is to obtain a balance bias-variance adjustment.

The OLS is known as the approach of full model as it does not embrace the subset of predictors. Meanwhile, as the bias-variance tradeoff is concerned the OLS estimators usually contain considerable low bias and usually have large variance. Reducing the variance and bias let the explanation becomes more tough (Tibshirani, 1996). There are several discrete methods quested in the literature for the improvement of the overall performance of OLS estimates e.g., subset selection.

1.7 Shrinkage Methods

Relevant variable selection and prediction accuracy is usually achieved by shrinkage methods. Focus of developing shrinkage methods is attaining oracle procedures to achieve the fundamental properties of statistical learning. Oracle procedures are intended to enjoy prediction and consistent variable selection simultaneously.

Shrinkage methods are used for estimation also known as penalization methods or regularization methods. These methods are more useful when the number of predictors is larger than the number of observations (p>n) i.e., in high dimensional scenarios. But these may also be beneficial in small dimensional situations when the sparse representation is anticipated of the fundamental model (Sundberg, 2002).

1.7.1 The Lasso

The LASSO (least absolute shrinkage and selection operator) was proposed by Tibshirani (1996). It is a method of regularization which does not simply perform shrinkage but also perform variable selection. It sets the coefficients exactly zero or it shrinks towards zero. OLS estimates are not explicitly used in this method. The idea of the lasso was derived from the shrinkage method, a non-negative garrote proposed by Breiman (1995).

The lasso owns the two oracle properties which are as follows.

1. If the true regression coefficient has about zero components then they are projected as zero with probability approaches to 1.
2. The non-zero components are assessed efficiently and also when the correct sub-model is already identified.
3. The lasso estimator can be explained as

$$\hat{\beta} = \arg\min\left\{\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right\},\ \lambda \geq 0. \tag{1.6}$$

In the equation (1.6) Tibshirani assumed that the solution for intercept (β_0) is $\widehat{\beta}_0 = \bar{y}$ and $\bar{y} = 0$ for all values of tuning parameter. Hence intercept is absent. The second term in (1.6) is called L_1 and λ is a non-negative tuning parameter. Larger the value of λ, more will be the shrinkage.

1.7.2 LAD-LASSO

It is the mixture of L_1 criterion (LAD) and the traditional shrinkage operator (LASSO). It is the useful technique when there are outliers explicitly in response variable. Wang et al., (2007) proposed LAD-LASSO technique which can function as shrinkage and selection operator and also resilient to outliers or heavy tailed errors. It is defined as

$$LAD - lasso = \sum_{i=1}^{n}\left|y_i - x_i'\beta\right| + n\sum_{j=1}^{p}\lambda_j\left|\beta_j\right| \tag{1.7}$$

It can be observed that it is the mixture of two methods. First part presents in LAD and the second part presents LASSO. Xu et al., (2010) also discussed the same idea. They both use Bayesian information type objective function for the estimation of tuning parameter.

1.7.3 WLAD-LASSO

According to Arslan (2011) that LAD-LASSO does not perform well when there are outliers in the explanatory variable and even a solitary leverage point may seriously affect its results. To overcome this problem, he proposed a Weighted LAD-LASSO technique. It can be defined as

$$WLAD-lasso = \sum_{i=1}^{n} w_i \left|y_i - x_i'\beta\right| + n\sum_{j=1}^{p} \lambda_j \left|\beta_j\right| \tag{1.8}$$

The weighted LAD-LASSO is more resistant as compare to LAD-LASSO it’s because it performs well even in the presence of heavy tailed distributions and outliers. Jung (2011) also discussed the same idea by considering some different simulation settings and weighting scheme.

1.8 Research Objectives

The main objective of this research is to study and compare different ways of dealing with outliers, influential variables and making an appropriate variable selection at a time. We will study model selection methods like AIC, AICc BIC and robust techniques like LAD and WLAD. We will also study the robust shrinkage techniques like LAD-LASSO and WLAD-LASSO. We will also study the model accuracy and its prediction accuracy. We will also make this comparison using real life data applications.

LITERATURE REVIEW

Breiman (1995) introduced a new method for better subset selection called non-negative garrote. He discussed drawbacks of subset selection. He discussed that there exists instability and lack of prediction accuracy.

Tibshirani (1996) argued over the drawbacks of OLS technique. He discussed the two techniques used to improve OLS estimates. One of them is subset selection and other one is ridge regression.

Fan and Li (2001) suggested the penalized likelihood approaches to overcome the problem of computational complexity of already existed variable selection methods and ignorance of stochastic errors.

Efron et al., (2004) examined LASSO and Forward Stepwise which are most commonly used model building methods. Their work mainly focused the degrees of freedom of the LARS algorithm. A Cp type statistic was driven by them to make a comparison among all the estimates. It requires computational efforts as much required by the OLS.

Giloni et al., (2005) discussed that LAD is carried out by minimizing the absolute values of residuals. As, LAD regression is highly sensitive to leverage points as it has breakdown of point $1/n$ and n is the sample size.

For variable selection and regularization method Zou and Hastie (2005) proposed a new method named as elastic net. At first they discussed naïve elastic net but it has a drawback of over shrinkage in regression problems. So, they introduced a new modified technique. They discussed that in high dimensional situations it is very important to be concerned about grouped selection problems.

Zou (2006) proposed Adaptive LASSO and investigated that LASSO can perform more efficiently by a little modification in it. It can even accomplish the oracle properties. Adaptive LASSO can be calculated by imposing certain weights on L_1-penality for penalizing the coefficients.

Wang et al., (2007) discussed about robust regression shrinkage and consistent variable selection. The least absolute deviation (LAD) is widely used for robust regression and least absolute selection.

Wang et al., (2009) after detailed study of shrinkage methods, they suggested a new and modified BIC-type criterion for estimation of the tuning parameter in shrinkage methods. They showed that under certain condition of fixed dimension it can consistently detect the correct model.

Arslan (2011) proposed weighted LAD regression method. He suggested that by combining weighted least absolute deviation (WLAD), least absolute shrinkage and selection operator (LASSO) we can attain a robust parameter estimation and variable selection both at the same time.

Jung (2011) discussed the drawbacks of LAD-LASSO. He discussed that LAD-LASSO does not perform well when there are outliers in x-space or there are leverage points. He proposed a more robust version of LASSO which is not sensitive to leverage points and heavy tailed distributions.

Tibshirani (2011) gave a brief detail about LASSO in his novel paper which is accepted globally (Tibshirani, 1996). He also pointed out the generalization of the LASSO. He forced on studying existing techniques rather developing new ones.

The penalized regression estimation and its most important aspect have been discussed by Sun et al., (2013). These methods mainly depend upon the tuning parameter that helps in balancing the trade-off between model sparsity and model fitting. They studied that properties of already existed criteria like cross validation, GCV, Mallow’s C_p, AIC and BIC and discussed that these follow the maximization of the posterior model probability and minimization of estimated prediction error.

Jung (2014) proposed a combined method of robust regression estimation and Smoothly Clipped Absolute Deviation (SCAD) function which enjoys the properties of unbiasedness for large coefficients and sparsity at the same time.

METHODOLOGY

3.1 Model Selection Methods

According to Shao’s definition, model selection methods can be classified into two major groups or classes. One of those classes consists of all efficient model selection criteria and the other consists of consistent model selection criteria.

3.1.1 Bayesian Information Criterion

Bayesian information criterion, also known as Schwarz criterion as it was established by Schwarz (1978). As we have discussed in section 3.1.1 that BIC belongs to the consistent model selection criteria. It is penalized by the complexity of the model and is a likelihood criterion. It selects the model which minimizes the following criterion

$$BIC = \ln \hat{\sigma}^2 + \frac{\ln n}{n} r$$

where ln is the natural log, $\hat{\sigma}^2$ is the residual sum of squares, r is the number of parameters and n is the number of observation.

3.1.2 Akaike Information Criterion

The Information Criterion by Akaike was developed in 1974 and called as Akaike Information Criterion. It is closely unbiased if the model is either over fitted or correct. It selects the best model as the one which minimizes the criterion defined by

$$AIC = n \log \hat{\sigma}^2 + 2\left(p+1\right)$$

This criterion is from the class of efficient model selection criteria.

3.1.3 Corrected Akaike Information Criterion

Hurvich and Tsai (1989) revealed that the AIC tends to give biased result with small sample size. They suggested a corrected version of Akaike information criteria which is less biased. This criterion provides better result as compared to another efficient criterion when the true model has finite dimension. AICc can be defined as

$$AICc = n \log \hat{\sigma}^2 + n \frac{1 + {p}/{n}}{1 - {(p+2)}/{n}}$$

3.2 Robust Regression Shrinkage Methods

These methods are the combination of robust regression techniques and the shrinkage methods mainly LASSO. These are helpful for consistent variable selection especially when there are outliers in the data.

3.2.1 Computation of LAD-LASSO

As we have discussed the criterion of LAD-LASSO in section 1.7.4. Here are the steps for the computation of LAD-LASSO.

- We need to calculate robust estimates $\hat{\mu}$ and $\hat{\Sigma}$ for the computation of LAD-LASSO.
- Estimate β by minimizing the unpenalized LAD objective function 8
- Compute tuning parameter using estimated β

- In the computation of LAD-LASSO we use true an augmented dataset. This dataset consists of $\left\{\left(y_i^*, \mathrm{x}_i^*\right)\right\}$ with $i = 1, \ldots. n, \mathrm{n}+1, \ldots, n+p$, where $\left(y_i^*, \mathrm{x}_i^*\right) = \left(y_i, \mathrm{x}_i\right)$ for $1 \le i \le \mathrm{n}$, $\left(y_{n+j}^*, \mathrm{x}_{n+j}^*\right) = \left(0, n\lambda_j e_j\right)$ for $1 \le j \le p$, where e_j to be a p-p-dimensional vector, where the j^{th} component equal to 1 and all others equal to be zero.
- Calculate LAD Regression estimators using LAD regression estimators for the dataset $\left(y_i^*, \mathrm{x}_i^*\right)$ for $i = 1, 2, 3, \ldots, n, n+1, \ldots, n+p$

3.2.2 Computation of WLAD-LASSO

It involves following steps:

- We need to calculate robust estimates $\hat{\mu}$ and $\hat{\Sigma}$ for the computation of WLAD-LASSO. Weights w_i can be calculated using this information as we have discussed in section.
- Estimate β by minimizing the unpenalized WLAD objective function
- Compute tuning parameter using estimated β
- Using the dataset $\left(y_i^*, \mathrm{x}_i^*\right)$ for $i = 1, 2, 3, \ldots, n, n+1, \ldots, n+p$ where $\left(y_i^*, \mathrm{x}_i^*\right) = \left(w_i y_i, \mathrm{w}_i\, \mathrm{x}_i^*\right)$ for $i = 1, 2, \ldots, n$ and $\left(y_i^*, \mathrm{x}_i^*\right) = \left(0, n\hat{\lambda}_j e_j\right)$ $\mathit{for}\ j = 1, 2, \ldots, p$
- Calculate LAD Regression estimators using LAD regression estimators for the dataset $\left(y_i^*, \mathrm{x}_i^*\right)$ for $i = 1, 2, 3, \ldots, n, n+1, \ldots, n+p$

3.3 Selection of Tuning Parameter

Tuning parameter is one of the most important parts of shrinkage methods so selection of tuning parameter has very vital role. It is very important to perform consistent variable selection as well as for shrinkage. It becomes difficult to satisfy oracle properties with unsuitable selection of tuning parameter. Many methods are available in literature to perform the selection of tuning parameter. In this study we used BIC-type criterion for this purpose.

3.3.1 BIC-type Criterion

Wang et al., (2007) measured the following BIC-type objective function

$$\sum_{i=1}^{n}\left|y_i - \mathrm{x}_i'\beta\right| + n\sum_{i=1}^{p}\lambda_j\left|\beta_j\right| - \log\left(.5n\lambda_j\right)\log\left(n\right)$$

This can penalize the complexity of the model same as the BIC using the factor log(n). It generates the tuning parameter estimate as

$$\hat{\lambda}_j = \frac{\log\left(n\right)}{n\left|\tilde{\beta}_j\right|}$$

where $\tilde{\beta}_j$ is the estimated unpenalized LAD estimator. This tuning parameter estimate satisfies the properties (1) $\sqrt{n}\hat{\lambda}_j \to 0$ for $j \le p_0$ (2) $\sqrt{n}\hat{\lambda}_j \to \infty$ for $j > p_0$.

3.4 The Over Fitting Effect

Wang et al., (2007) illustrated the concept of under fitted, over fitted and correctly fitted in detail. Significant variables or non-zero components are basically used to determine any of the above mentioned effect. Let p_o is the size of significant variables in the true known sub model represented as $S_T = \{j_1, ..., j_p\}$ and $S_\lambda = \{j : \hat{\beta}_{j,\lambda} \ne 0\}$ where $\hat{\beta}_{j,\lambda}$ are the estimates. Generally, $\Omega = [0, \lambda_{\max}]$ is the bounded interval where the optimal value of λ (tuning parameter) lies. Ω_-, Ω_o and Ω_+ are the subsets of Ω which represent the under fitted, over fitted and correctly fitted models. Moreover, an under fitted model is represented as $\Omega_- = \{\lambda \in \Omega : S_\lambda \not\supset S_T\}$ in which at least one significant variable is overlooked in the estimated model, over fitted model is the one which consists of at least one insignificant variable and all the significant variables and it can be illustrated as $\Omega_+ = \{\lambda \in \Omega : S_\lambda \supset S_T \text{ and } S_\lambda \ne S_T\}$ and when the estimated model contains all significant variables as specified by the true model so, the correctly fitted model can be specified as $\Omega_0 = \{\lambda \in \Omega : S_\lambda = S_T\}$.

3.5 Root Mean Square Error (RMSE)

It is defined by Draxler et al., that RMSE rebukes the variance. This measure is widely used in model evaluation. It is a very useful as it depicts the complete photo of error distribution. It can be defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \quad i = 1, 2, ..., n$$

It is stated that it is comparatively better at revealing the best model with respect to its performance.

3.6 Mean Absolute Prediction Error (MAPE)

It is a measure to check the forecast accuracy of the model. Swanson et al., (1999) discussed it in detail the methodology of MAPE and as well as its advantages over other criteria. Wang et al., (2007) also used it in their paper to illustrate the forecasting accuracy of proposed method. It can be calculated as

$$MAPE = \left(\left(\sum |F_i - O_i| \big/ O_i \right) \Big/ N \right) \times 100.$$

RESULTS AND DISCUSSION

We used different contamination rates (ε) with the values 0,0.1,0.2 and 0.3 to exhibit the influential values in x dataset. Let $m=[\varepsilon n]$ be the number of contaminated data, where n is sample size. The first $n-m$ x-data points are generated as $x_{1i} \sim Np(0,V)$ where $V=(v_{ij})$ with $v_{ij}=0.5^{|i-j|}$ and next contaminated data points are generated as $x_{2i} \sim Np(u,I)$ with $u \neq 0$. These two datasets are combined to make one dataset of p-dimensional.

4.1 Scenario 1

We set $\beta=(0.5,1.0,1.5,2.0,0,0,0,0)$ i.e., $p=8$. In other words, first four regression variables are significant and rests are not and $v_{ij}=0.5^{|i-j|}$ $i,j=1,2,3,...,p$ which shows the moderate correlation between predictors. This is also studied by Wang et al., (2007).

4.2 Model Evaluation Measures when there is No Contamination

Scenarios are evaluated based on correct fit, RMSE and MAPE. First, we will discuss the results when there are no leverage points or no outliers in x-space.

4.2.1 Comparison based on Model Fitting Criteria

Table 4.1 shows that LAD-LASSO 93.4% correctly fits the model which is comparable to the BIC and WLAD-LASSO with the values 77.5% and 90.8% respectively for sample size 50 and noise ratio 0.5. It can be observed that the overall performance of all methods with respect to correct fit improve with the increasing sample size except the AICc. It can also be observed that the performance of these methods goes down with the increase of noise ratio from 0.5 to 1. All the methods 0% under fits the model when sample size is large and noise ratio is small. When there is small sample size and small noise ratio and even when there is small sample size and weak signal to noise ratio and even when there is any sample size with strong noise ratio, all the methods under fits the model except LAD and WLAD. LAD and WALAD 100% over fit the data in all scenarios. Identification of correct number of zeros in the model improves with the increase in sample size. LAD-LASSO and WLAD-LASSO are close in identification of no. of zeros in the model correctly. When the noise ratio is increased identification of correct number of zeros in the model reduces e.g., for LAD-LASSO it reduces 3.945 to 3.667, 3.969 to 3.757 and 3.982 to 3.827 for sample sizes 50, 100 and 200 respectively.

Table 4.1
Simulation results of Scenario 1 where $\beta = (0.5, 1.0, 1.5, 2.0, 0, 0, 0, 0)$.
Random error has double exponential distribution. Contamination rate is 0%, signal to noise ratio is 0.5 with different sample sizes (n = 50, 100, 200)

n	Method	UF	CF	OF	OF_1	OF_2	OF_3	OF_4	Incorrect	Correct
50	AIC	0.004	0.676	0.32	0.245	0.064	0.011	0	0.408	3.592
	AICc	0.001	0.208	0.791	0.385	0.28	0.11	0.016	1.34	2.66
	BIC	0.009	0.775	0.216	0.18	0.03	0.006	0	0.261	3.739
	LAD	0	0	1	0	0	0	1	4	0
	LASSO	0.016	0.934	0.05	0.048	0.002	0	0	0.055	3.945
	WLAD	0	0	1	0	0	0	1	4	0
	WLAD-LASSO	0.044	0.908	0.048	0.048	0	0	0	0.053	3.947
	ORACLE	0	1	0	0	0	0	0	0	4
100	AIC	0	0.704	0.296	0.242	0.052	0.001	0.001	0.353	3.647
	AICc	0	0.242	0.758	0.354	0.301	0.083	0.02	1.285	2.715
	BIC	0	0.861	0.139	0.131	0.008	0	0	0.147	3.853
	LAD	0	0	1	0	0	0	1	4	0
	LASSO	0	0.964	0.036	0.035	0.001	0	0	0.037	3.963
	WLAD	0	0	1	0	0	0	1	4	0
	WLAD-LASSO	0.001	0.97	0.029	0.028	0.001	0	0	0.031	3.969
	ORACLE	0	1	0	0	0	0	0	0	4
200	AIC	0	0.705	0.295	0.249	0.043	0.001	0.002	0.346	3.654
	AICc	0	0.236	0.764	0.347	0.295	0.108	0.014	1.317	2.683
	BIC	0	0.914	0.086	0.085	0.001	0	0	0.087	3.913
	LAD	0	0	1	0	0	0	1	4	0
	LASSO	0	0.982	0.018	0.018	0	0	0	0.018	3.982
	WLAD	0	0	1	0	0	0	1	4	0
	WLAD-LASSO	0	0.978	0.022	0.022	0	0	0	0.022	3.978
	ORACLE	0	1	0	0	0	0	0	0	4

4.2.2 Comparison based on RMSE

This table 4.2 contains RMSE (root mean square error) which is a measure of model evaluation. We can see the performance of one model by comparing its RMSE value by the RMSE value of oracle. We computed average RMSE and median RMSE. RMSE is small for AIC and AlCc as compared to other methods when signal to noise ratio is weak and this value increases as the sample size increases. RMSE of oracle decreases with the increase of sample size and these phenomena exist for both weak and strong noise ratios. The behavior of BIC, LAD and WLAD with respect to RMSE is same as AIC and AICc except the difference that their RMSE are slightly higher. RMSE of LAD-LASSO and WLAD-LASSO are close to oracle and does not increase when sample size increases. It is interesting to note that overall RMSE increases for all methods when them is strong signal to noise ratio. It is also notable that Median values of RMSE are slightly smaller than the values of Average RMSE.

Table 4.2
Simulation Results of Scenario 1 where $\beta = (0.5, 1.0, 1.5, 2.0, 0, 0, 0, 0)$.
Random Error has Double Exponential Distribution. Contamination Rate is 0%, Signal to Noise Ratio is 0.5 with Different Sample Sizes (n= 50,100,200)

n	Method	Scenario 1	
		Avg RMSE	Med RMSE
50	AIC	0.657	0.649
	AICc	0.645	0.638
	BIC	0.66	0.652
	LAD	0.668	0.66
	LASSO	0.694	0.683
	WLAD	0.669	0.658
	WLAD-LASSO	0.697	0.687
	ORACLE	0.72	0.71
100	AIC	0.682	0.676
	AICc	0.676	0.671
	BIC	0.684	0.678
	LAD	0.687	0.681
	LASSO	0.696	0.69
	WLAD	0.687	0.681
	WLAD-LASSO	0.697	0.692
	ORACLE	0.707	0.701
200	AIC	0.694	0.693
	AICc	0.691	0.69
	BIC	0.696	0.695
	LAD	0.697	0.695
	LASSO	0.701	0.7
	WLAD	0.697	0.695
	WLAD-LASSO	0.701	0.7
	ORACLE	0.704	0.702

4.3.3 Comparison based on MAPE

Table 4.3 includes MAPE (mean absolute prediction error) which measures the prediction accuracy of the model based on another set of 1000 simulated data. We can observe the overall pattern of all methods with weak signal to noise ratio show that MAPE (both average and median) of small sample is small and it gets lower for moderate sample size and tend to increase for large sample size. Average MAPE of AIC, AlCc, BIC and LAD am close to each other when sample size is 50 and noise to signal ratio is weak and in this scenario highest value of MAPE are of WLAD and WLAD-LASSO. Average MAPE of WLAD and WLAD-LASSO come closer to the values of MAPE of remaining methods as sample size is increased. Median MAPE of AIC, AlCc, BIC, LAD and WLAD-LASSO are close to each other when there is moderate sample size and low noise. WLAD has larger median MAPE in this scenario. MAPE of WLAD and WLAD-LASSO increases as sample size increases. Even when there is strong ratio of noise MAPE of all methods are close to each other except the WLAD which has highest MAPE among all. LAD-LASSO shows least MAPE (average and median) as compared to other methods and it is more close to the MAPE (average and median) of oracle and even for large sample size and with strong signal to noise ratio average MAPE of LAD-LASSO is smaller than oracle.

Table 4.3
Simulation Results of Scenario 1 where $\beta = (0.5, 1.0, 1.5, 2.0, 0, 0, 0, 0)$.
Random Error has Double Exponential Distribution. Contamination Rate is 0%, Signal to Noise Ratio is 0.5 with Different Sample Sizes (n= 50,100,200)

n	Method	Scenario 1	
		Avg RMSE	Med RMSE
50	AIC	1.893	0.641
	AICc	1.935	0.651
	BIC	1.884	0.627
	LAD	1.875	0.66
	LASSO	1.833	0.598
	WLAD	2.512	0.697
	WLAD-LASSO	2.383	0.643
	ORACLE	1.735	0.551
100	AIC	1.229	0.617
	AICc	1.278	0.614
	BIC	1.23	0.624
	LAD	1.237	0.617
	LASSO	1.22	0.598
	WLAD	1.418	0.642
	WLAD-LASSO	1.38	0.619
	ORACLE	1.184	0.586
200	AIC	1.35	0.683
	AICc	1.382	0.685
	BIC	1.341	0.679
	LAD	1.35	0.685
	LASSO	1.323	0.667
	WLAD	1.353	0.735
	WLAD-LASSO	1.399	0.722
	ORACLE	1.246	0.661

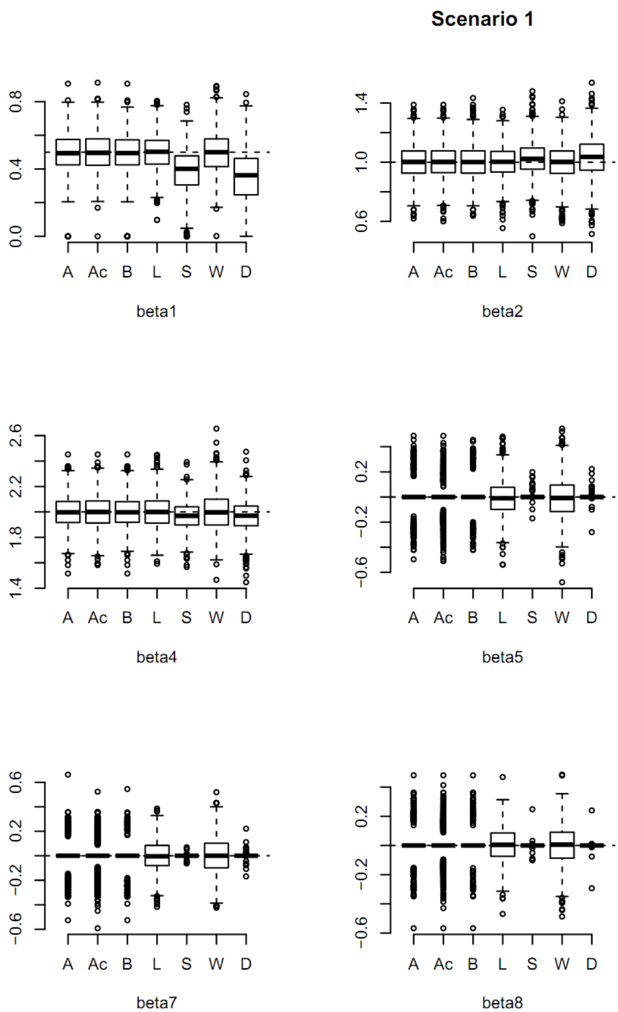


Figure 4.1: Boxplots of the estimates of Scenario 1 where $\beta=(0.5,1.0,1.5,2.0,0,0,0,0)$. Random error has double exponential distribution. Contamination rate is 0%, signal to noise ratio is 0.5 and n=100. A represents AIC, Ac represents AICc, B represents BIC, L represents LAD, S represents LAD-LASSO, W represents WLAD and D represents WLAD-LASSO.

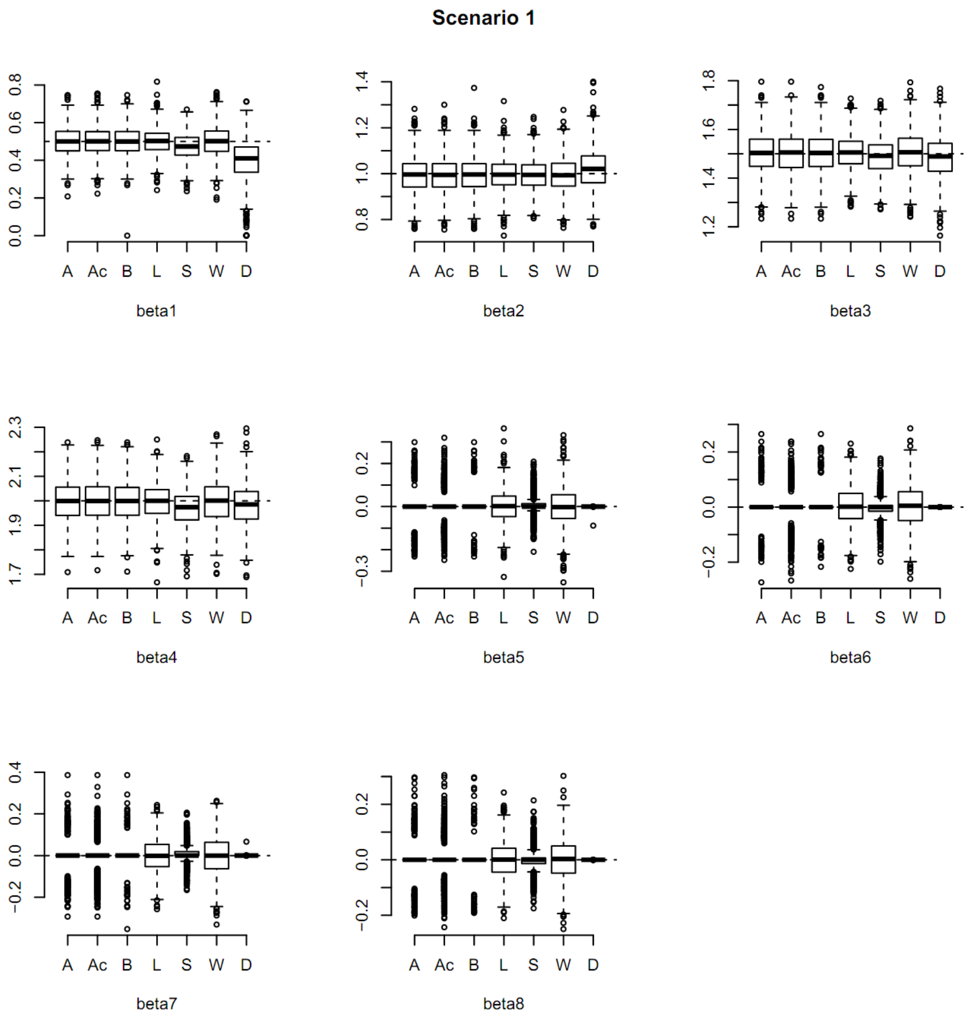


Figure 4.2: Boxplots of the estimates of Scenario 1 where $\beta = (0.5, 1.0, 1.5, 2.0, 0, 0, 0, 0)$. Random error has double exponential distribution. Contamination rate is 20%, signal to noise ratio is 0.5 and n=100. A represents AIC, Ac represents AICc, B represents BIC, L represents LAD, S represents LAD-LASSO, W represents WLAD and D represents WLAD-LASSO.

4.5 Real Data Examples:

In this section of the thesis, we applied various methods to estimate the models of different real datasets.

4.5.1 Example 1 (Prostate Cancer Data)

This dataset was primarily used by Stamey et al., (1989) to inspect the relationship between several clinical measures and the level of prostate specific antigen (lpsa). This data set has also been studied by Tibshirani (1996). Later on, by Fu (1998) , Yuan and Lin (2007) and Sun et al., (2013)

The prostate cancer data contains eight variables of clinical measure which are as log cancer volume (lcavol), log prostate weight (lweight), log benign prostate hyperplasia amount (lbph), age, seminal vesicle invasion (svi), log capsular penetration (lcp), gleason score (gleason) and percentage score 4 or 5 (pgg45). This data is from 97 patients who are aged from 41 to 97 and they got the radical prostatectomy. We applied different techniques on this dataset to estimate that which clinical measures are truly relevant in predicting the response variable (lpsa). For such purpose, we used the methods of AIC, AICc, BIC, LAD, LAD-LASSO, WLAD and WLAD-LASSO.

Table 4.9
Coefficients for the Prostate Cancer Dataset

	Methods							
Predictors	**AIC**	**AICc**	**BIC**	**LAD**	**LAD-LASSO**	**WLAD**	**WLAD-LASSO**	**OLS**
lcavol	0.6198	0.6713	0.6198	0.5651	0.3229	0.5757	0.5294	0.6651
lweight	0.2835	0.2632	0.2835	0.3042		0.3015		0.2665
age		-0.1557		-0.1891		-0.1791		-0.1582
lbph		0.1412		0.2015		0.1834		0.1403
svi	0.2756	0.3115	0.2756	0.3659		0.3692		0.3153
lcp		-0.1468		-0.1679		-0.1848		-0.1483
gleason				0.1235		0.1219		0.0355
pgg45		0.1502		0.1446		0.1635		0.1257

This table shows that how many predictors are important for the prediction of the level of prostate specific antigen (lpsa). AIC and BIC somehow shows similar results whereas LAD-LASSO and WLAD-LASSO estimates only one variable and shrinks all others to zero.

SUMMARY AND CONCLUSION

It can be observed that the overall performance of all methods with respect to correct fit improve with the increasing sample size except the AICc. It can also be observed that the performance of these methods goes down with the increase of noise ratio from 0.5 to 1. All the methods 0% under fits the model when sample size is large and noise ratio is small. When there is small sample size and small noise ratio and even when there is small sample size and weak signal to noise ratio and even when there is any sample size with strong noise ratio, all the methods under fits the model except LAD and WLAD. LAD and WALAD 100% over fit the data in all scenarios. Identification of correct number of zeros in the model improves with the increase in sample size. LAD-LASSO and WLAD-LASSO are close in identification of no. of zeros in the model correctly.

We can see the performance of one model by comparing its RMSE value by the RMSE value of oracle. We computed average RMSE and median RMSE. RMSE is small for AIC and AlCc as compared to other methods when signal to noise ratio is weak and this value increases as the sample size increases. RMSE of oracle decreases with the increase of sample size and these phenomena exist for both weak and strong noise ratios. The behavior of BIC, LAD and WLAD with respect to RMSE is same as AIC and AICc except the difference that their RMSE are slightly higher. RMSE of LAD-LASSO and WLAD-LASSO are close to oracle and does not increase when sample size increases. It is interesting to note that overall RMSE increases for all methods when them is strong signal to noise ratio. It is also notable that Median values of RMSE are slightly smaller than the values of Average RMSE.

MAPE (mean absolute prediction error) which measures the prediction accuracy of the model based on another set of 1000 simulated data. We can observe the overall pattern of all methods with weak signal to noise ratio show that MAPE (both average and median) of small sample is small and it gets lower for moderate sample size and tend to increase for large sample size. Average MAPE of AIC, AlCc, BIC and LAD are close to each other when sample size is 50 and noise to signal ratio is weak and in this scenario highest value of MAPE are of WLAD and WLAD-LASSO. Average MAPE of WLAD and WLAD-LASSO come closer to the values of MAPE of remaining methods as sample size is increased. Median MAPE of AIC, AlCc, BIC, LAD and WLAD-LASSO are close to each other when there is moderate sample size and low noise. WLAD has larger median MAPE in this scenario. MAPE of WLAD and WLAD-LASSO increases as sample size increases. Even when there is strong ratio of noise MAPE of all methods are close to each other except the WLAD which has highest MAPE among all. LAD-LASSO shows least MAPE (average and median) as compared to other methods and it is more close to the MAPE (average and median) of oracle and even for large sample size and with strong signal to noise ratio average MAPE of LAD-LASSO is smaller than oracle.

Overall, the performance of LAD-LASSO is highly affected by the contamination in the data as compared to its performance when there were no outliers in the x-space. AIC and BIC performs well. WLAD-LASSO is best method as compared to others when there are outliers in x-space as it can even 100% correctly fits the model. RMSE (average and median) of AlCc is smallest when there is any contamination rate with every sample size and there is weak signal to noise ratio. At the same time it is also interesting to note that with all above describe situation the RMSE of WLAD-LASSO is closest to oracle but it is not as smallest as AICc. The results show that when there is 20% contamination in the data, signals to noise ratio is weak and sample size is small then AIC has the smallest average MAPE whereas BIC has smallest Median MAPE. Average MAPE of AICc remains smallest as sample size is increased but hem Median MAPE of LAD-LASSO is smallest in this scenario. In high contamination, WLAD-LASSO has smallest Average MAPE for small sample size but as the sample size is increased this property moves from LAD-LASSO to BIC and this situation also holds for Median MAPE. Average MAPE of WLAD-LASSO is smallest for strong signal to noise ratio for moderate contamination with each sample size. For moderate and high contamination rate, LAD-LASSO shows smallest Median MAPE.

REFERENCES

1. Arslan, O. (2012). Weighted LAD-LASSO method for robust parameter estimation and variable selection in regression. *Computational Statistics & Data Analysis*, 56(6), 1952-1965.
2. Breiman, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics*, 37(4), 373-384.
3. Candes, E. and Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, 35(6), 2313-2351.
4. Chand, S. (2012). On tuning parameter selection of lasso-type methods-a monte carlo study. In *Applied Sciences and Technology (IBCAST), 2012 9th International Bhurban Conference on* (120-129). IEEE.
5. Chatterjee, S. and Hadi, A.S. (1986). Influential observations, high leverage points, and outliers in linear regression. *Statistical Science*, 1(3), 379-393.
6. Dodge, Y. (1997). LAD regression for detecting outliers in response and explanatory variables. *Journal of Multivariate Analysis*, 61(1), 144-158.
7. Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456), 1348-1360.
8. Frank, L.E. and Friedman, J.H. (1993). A statistical view of some chemometrics regression tools. *Technimetrics*, 35(2), 109-135.
9. Fu, S., Zou, X., Wang, X. and Liu, X. (1998). Determination of polyamines in human prostate by high-performance liquid chromatography with fluorescence detection. *Journal of Chromatography B: Biomedical Sciences and Applications*, 709(2), 297-300.
10. Giloni, A., Simonoff, J.S. and Sengupta, B. (2006). Robust weighted LAD regression. *Computational Statistics & Data Analysis*, 50(11), 3124-3140.
11. Hastie, T., Tibshirani, R. and Friedman, J.H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second edition. Springer, New York, NY.
12. Huber, P.J. (1987). The place of the L1-norm in robust estimation. *Computational Statistics & Data Analysis*, 5(4), 255-262.
13. Jung, K.M. (2011). Weighted least absolute deviation lasso estimator. *Communications for Statistical Applications and Methods*, 18(6), 733-739.
14. Jung, K.M. (2014). Robust estimator with the SCAD function in penalized linear regression. *The Standard International Journals*, 2(14), 156-160.
15. Jung, Y., Kim, J.K., Shiozawa, Y., Wang, J., Mishra, A., Joseph, J., Berry, J.E., McGee, S., Lee, E., Sun, H. and Wang, J. (2013). Recruitment of mesenchymal stem cells into prostate tumors promotes metastasis. *Nature communications*, 4, Article 1795. DOI: 10.1038/ncomms2766.
16. Kutner, M., Nachtsheim, C. and Neter, J. (2004). *Applied Linear Regression Models*. The McGraw-Hill.
17. Manoj, K. and Senthamarai Kannan, K. (2013). Comparison of methods for detecting outliers. *International Journal of Scientific and Engineering Research*, 4(9), 709-714.
18. Miller, A. (1990). *Subset selection in regression.* Monographs on Statistics and Applied Probability. Chapman and Hall.

19. Nguyen, P.L. (2006). Effects of social class and school conditions on educational enrollment and achievement of boys and girls in rural Viet Nam. *International Journal of Educational Research*, 45(3), 153-175.
20. Radchenko, P. and James, G.M. (2008). Variable inclusion and shrinkage algorithms. *Journal of the American Statistical Association*, 103(483), 1304-1315.
21. Radchenko, P. and James, G.M. (2011). Improved variable selection with forward-lasso adaptive shrinkage. *The Annals of Applied Statistics*, 5(1), 427-448.
22. Rousseeuw, J.P. and Leroy, M.A. (1987). *Robust regression and outlier detection*. John Wiley and Sons.
23. Seo, S. (2006). *A review and comparison of methods for detecting outliers in univariate data sets*, Doctoral dissertation, University of Pittsburgh.
24. Stapleton, H.J. (1995). *Linear Statistical Models*. John Wiley & Sons.
25. Sun, W., Wang, J. and Fang, Y. (2013). Consistent selection of tuning parameters via variable selection stability. *The Journal of Machine Learning Research*, 14(1), 3419-3440.
26. Sundberg, R. (2002). Shrinkage regression. *Encyclopedia of Environmetrics*, 4, 1994-1998.
27. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267-288.
28. Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3), 273-282.
29. Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67, 91-108
30. Wang, H., Li, G. and Jiang, G. (2007). Robust regression shrinkage and consistent variable selection through the LAD-Lasso. *Journal of Business & Economic Statistics*, 25(3), 347-355.
31. Xu, J. and Ying, Z. (2010). Simultaneous estimation and variable selection in median regression using Lasso-type penalty. *Annals of the Institute of Statistical Mathematics*, 62(3), 487-514.
32. Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1), 49-67.
33. Yuan, M. and Lin, Y. (2007). On the non-negative garrotte estimator. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2), 143-161.
34. Zaman, A., Rousseeuw, P.J. and Orhan, M. (2001). Econometric applications of high-breakdown robust regression techniques. *Economics Letters*, 71(1), 1-8.
35. Zhao, P. and Yu, B. (2006). On model selection consistency of Lasso. *Journal of Machine Learning Research*, 7(Nov), 2541-2563.
36. Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476), 1418-1429.
37. Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301-320.
38. Zou, H. and Zhang, H.H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Annals of Statistics*, 37(4), 1733-1751.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 301-313

FORECASTING OF PETROL OIL PRICES IN PAKISTAN: EVIDENCE FROM BOX-JENKINS MODELING

Muhammad Waqas[1], Sidra Razzaq[2] and **Samina Satti[3]**
Department of Statistics, University of Wah, Wah Cantt, Pakistan
Email: [1]muhammad.waqas@uow.edu.pk
[2]sidra.razzaq@uow.edu.pk
[3]samina.satti@uow.edu.pk

ABSTRACT

Graph of the derived data from 1920 to 2018 shows that there is variation in mean as well as variance. For removing variation in variance and making it constant we first apply LN transformation on original data, but it doesn't show stationary or constant mean and variance so in 2nd step we applied 1st difference transformation and 2nd difference transformation. That depicted our series perfectly stationary on 1st level of difference of LN of the series. The correlogram also become stationary at 1st difference. The value of Augmented Dickey Fuller (9.82), endorsed the non-existence of the unit root. The value of the Durbin Watson test is (2.02) by which concluded that our free of residuals autocorrelation. After fulfillment of basic assumptions, seven models were estimated. The models were taken, and their results were examined. Based on that the model having less value of Akaike Information Criterion and Bayesian Information Criterion was selected. In present study the 2nd, 3rd and 7th model has least value of Akaike Information Criterion and Bayesian Information. The 7th model is also recommended by the automatic ARIMA. So, after plotting the normality curve, forecasting is also displayed on the basis of most fit models for future concerns.

KEYWORDS

ADF, Durban Watson, ARIMA, Forecasting.

1. INTRODUCTION

Gasoline oil is one of the most used and renown oil in the world. Pakistan is using 450,000 barrels per day which ranks on the 34th biggest user in the world. China, USA, Japan, India South Korea and Netherland are largest oil importers in the world. Kuwait is importing 55.4%, Oman 15.6%, Netherlands 8.5% and Malaysia 5.4% the gasoline prices its self a big deal and challenging due the political and economic instability of oil producers and relationship with the oil purchasing countries. As well as it depends on the geo strategical location of the importers and experts of oil. There almost 10 oil refineries working in Pakistan. Most popular of them are Pak-Arab Refinery, National Refinery, Attock Refinery, Byco Petroleum Pakistan Ltd., Pakistan Refinery Ltd.

Pakistan is importing oil because it has shortest oil reserves that is only 350.6 million barrels of oil. There are over 1000 individual oil fields worldwide which have more oil in them than Pakistan's total oil reserves. Pakistan has 85,500 bbl/day oil production but it imports 166,000 bbl/day, and it refines 517,000 bbl/day of products while purchasing 381,200 bbl/day of refined products (Report by cia, Facebook) [4].

Pakistan’s most of the imports are fuel, petroleum commodities, edible oil, chemicals substances, and fertilizers. Petroliam products having largest share in all of the imports. Now it’s depending upon the geological strategic and political situation of exporting and importing countries and international relations, the prices of oil remains unstable and fluctuate more. As little unrest or political situation among countries variate the prices of oil also variate. The other factors behind the oil prices changes is the demand and supply factor that the population of Pakistan is increasing day by day so its oil demand is increasing. 25th-largest crude oil imports (372,800 barrels / day) in the world (by latest facts) [3].

Forecasting of oil prices is much challenging because of the uncertain and unpredictable patterns of the prices. Including the factor of world politics also affecting it a lot, which convert prediction into much tougher job. Multiple methodologies are adopted by the predictor, like cause and effect scenario by considering the covariates in analysis, but this method has limitations like when you have multiple variables contributing to a cause, then it’s difficult to specify a single or unique variable as the cause to change in prices. As a consequence, the calculations made on the estimated values of the explanatory variables exceptionally enlarged the forecast errors. This technique also called fundamental analysis. But on 2nd famous methodology is times series modeling, specifically called Box-Jenkins Methodology [1]. In this technique we forecast on the past behavior of the series by not much focusing on the other variables. Pakistan is most of time facing energy crises, especially increased after 2008 and continue so on. So, it’s too important to perform analysis on the data of energy crises and predict the results.

According to a magazine by The Express Tribune (2010), “The oil & gas sector outturned a handsome amount of 45%, including dividends, during the outgoing year. The sector outperformed the benchmark 100-share index at the KSE by 19% ton the back of rising foreign interest and volume-led growth in earnings”. Further details about the KSE and the oil & gas companies have been given in the next section [2].

2. LITERATURE REVIEW

Forecasting and Modelling of the Oil prices is not an easy task because of the drastically changing geo-political situation and world market fluctuations. Due to the instabilities in oil prices specifically throughout the Iraq war (1990), afghan war (2002) and over the globe financial crisis (2008). Throughout these periods crude oil prices have some momentous impact and volatile period looks considerable for the larger variations in the retrospective data of prices of the crude oil [1].

The ARIMA (p, d ,q)-Box-Jenkins Model given by Box and Jenkins is most famous methodology for construction of univariate time series forecasting model. The ARIMA (p, d, q) Model is estimated after making a series stationary by taking its difference at 1st level or 2nd level. The ARMA (p,q) is the general form of the ARIMA, where p and q denote to the order of the autoregressive (AR) terms and moving average(MA). [2] The process of identifications started from calculating autocorrelation function and partial-autocorrelation function through the correlogram. Extensive material is available on such modelling, e.g., see Brockwell and Davis (2003) and Weisang and Awazu (2008) for recent survey [3]. Theoretically, in oil manufacturing countries the effect on stock returns of oil prices is positive. Bjørnland (2009) wrote that higher oil prices displayed an instantaneous exchange of money from oil importers to exporters' countries. The Box-Jenkins methodology holds four phases identification on series, estimation of the models, diagnostic checking on the models and their efficiency on the basis of differ criteria's and forecasting in the basis of choose model.

The identification step contains choosing one or more ARIMA model(s) constructed on the estimated ACF and PACF plots. The ACF depiction of the AR (Auto Regressive) process falloffs exponentially to zero and its PACF plot displays cut-off to zero after the lag p. The ACF illustration of the MA process displays cut-off to zero after lag q while its PACF declines exponentially to zero. The combination of the AR (P) and the MA (q) processes is known as the ARMA (p,q) process. When the difference applied to make a series stationary that i(th) term indicate to the integrated of ARIMA process [9]. Diagnostic checking comprises on to evaluating the validity of the already considered and fitted models through possible statistically significant tests like Akaike information criterion, Bayesian criterion, log likelihood test statistics or calculation mean square error. Lastly, the best fitting ARIMA model would be used to make forecast.

3. METHODOLOGY

3.1 Identification

The data of the Gasoline sale prices was depicted to check the overall trend of the data, and to catch if any outlier exists in the series. Figure 1 illustrates the time series plot of data. As the data shown non-normality by its histogram and the value of Jarque Bera Test statistics 161.87 in Figure 2. Concluded that there exist skewness and high kurtosis. Stationarity in data tested by using the correlogram, Augmented Dicky Fuller statistics -2.73 which is too low and the Durbin Watson test 1.97. The data shown non-stationary, that is depicted below in Figures 3 and 4.

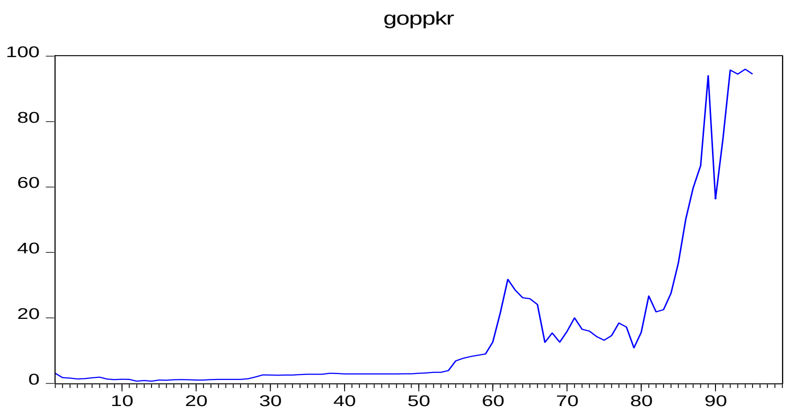


Figure 1: Graph of the Gasoline Prices

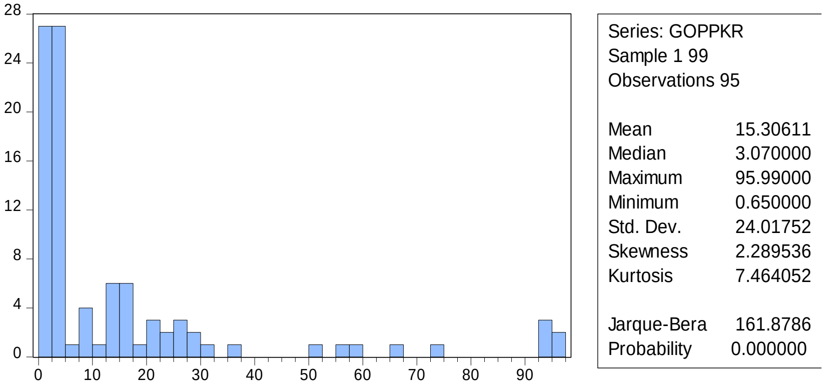


Figure 2: Descriptive Statistics of Gasoline Prices

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.903	0.903	80.021	0.000
		2	0.819	0.013	146.42	0.000
		3	0.739	-0.016	201.08	0.000
		4	0.640	-0.144	242.59	0.000
		5	0.572	0.096	276.05	0.000
		6	0.523	0.080	304.41	0.000
		7	0.415	-0.346	322.45	0.000
		8	0.344	0.103	334.98	0.000
		9	0.272	-0.047	342.93	0.000
		10	0.221	0.160	348.22	0.000
		11	0.192	-0.024	352.27	0.000
		12	0.168	-0.026	355.40	0.000
		13	0.150	0.142	357.92	0.000
		14	0.139	-0.079	360.12	0.000
		15	0.125	0.061	361.92	0.000
		16	0.123	-0.060	363.68	0.000
		17	0.125	0.061	365.53	0.000
		18	0.122	-0.013	367.32	0.000
		19	0.116	-0.085	368.96	0.000
		20	0.119	0.138	370.71	0.000
		21	0.129	0.020	372.78	0.000
		22	0.133	0.010	375.03	0.000
		23	0.143	-0.005	377.67	0.000
		24	0.144	-0.031	380.36	0.000
		25	0.138	0.061	382.88	0.000
		26	0.140	-0.038	385.49	0.000
		27	0.142	0.027	388.23	0.000
		28	0.133	-0.061	390.66	0.000
		29	0.131	0.034	393.06	0.000
		30	0.114	-0.037	394.89	0.000
		31	0.088	-0.075	396.01	0.000
		32	0.057	-0.045	396.49	0.000
		33	0.020	-0.071	396.54	0.000
		34	-0.023	-0.022	396.63	0.000
		35	-0.053	0.004	397.06	0.000
		36	-0.068	0.117	397.78	0.000

Figure 3: Correlogram of the Original Data

Null Hypothesis: GOPPKR has a unit root
Trend Specification: Intercept only
Break Specification: Intercept only
Break Type: Innovational outlier

Break Date: 85
Break Selection: Minimize Dickey-Fuller t-statistic
Lag Length: 2 (Automatic - based on Schwarz information criterion, maxlag=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.739635	0.8148
Test critical values:	1% level	-4.949133	
	5% level	-4.443649	
	10% level	-4.193627	

*Vogelsang (1993) asymptotic one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: GOPPKR
Method: Least Squares
Date: 12/27/18 Time: 08:31
Sample (adjusted): 4 95
Included observations: 92 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GOPPKR(-1)	0.840411	0.058252	14.42723	0.0000
D(GOPPKR(-1))	-0.383963	0.103553	-3.707905	0.0004
D(GOPPKR(-2))	-0.287462	0.101217	-2.840052	0.0056
C	1.672786	0.747239	2.238623	0.0278
INCPTBREAK	20.36147	4.107818	4.956761	0.0000
BREAKDUM	-6.294384	6.348631	-0.991455	0.3242

R-squared	0.950915	Mean dependent var	15.73554
Adjusted R-squared	0.948062	S.D. dependent var	24.28867
S.E. of regression	5.535381	Akaike info criterion	6.323191
Sum squared resid	2635.078	Schwarz criterion	6.487656
Log likelihood	-284.8668	Hannan-Quinn criter.	6.389571
F-statistic	333.2159	Durbin-Watson stat	1.978196
Prob(F-statistic)	0.000000		

Figure 4: ADF and DW Test for Unit Root

To achieve the stationarity the first difference was applied the data and checked by using the correlogram and Augmented Dickey Fuller test 13.36 ($p<0.01$). The test statistics of the ADF test and p-value were satisfactory, that means the data is stationary, that is depicted in the 5, 6. But the test statistics of the Durban Watson is 1.96 that caution about the existence of the auto-correlation among the residual terms of the data. To achieve the highest level of accuracy in the results the LN transformation on the data is applied and checked the results.

Lag	AC	PAC	Q-Stat	Prob
1	-0.217	-0.217	4.5827	0.032
2	-0.062	-0.114	4.9562	0.084
3	0.206	0.176	9.1440	0.027
4	-0.009	0.078	9.1516	0.057
5	0.076	0.128	9.7341	0.083
6	0.046	0.064	9.9537	0.127
7	0.020	0.046	9.9968	0.189
8	0.127	0.122	11.695	0.165
9	-0.103	-0.077	12.812	0.171
10	-0.053	-0.124	13.119	0.217
11	0.116	0.000	14.589	0.202
12	0.003	0.038	14.590	0.265
13	-0.089	-0.065	15.477	0.279
14	-0.064	-0.134	15.943	0.317
15	0.051	-0.007	16.238	0.366
16	0.033	0.071	16.363	0.428
17	-0.071	0.021	16.950	0.458
18	0.011	0.014	16.963	0.526
19	-0.050	-0.084	17.268	0.572
20	-0.075	-0.090	17.953	0.590
21	0.045	0.039	18.204	0.636
22	-0.121	-0.106	20.035	0.581
23	0.108	0.058	21.505	0.550
24	0.013	0.058	21.527	0.607
25	-0.033	0.120	21.671	0.655
26	0.047	0.067	21.962	0.691
27	0.101	0.140	23.328	0.667
28	-0.056	-0.011	23.755	0.694
29	-0.032	-0.098	23.898	0.734
30	0.064	-0.020	24.469	0.750
31	0.061	0.023	24.999	0.768
32	0.022	0.019	25.067	0.803
33	-0.002	-0.008	25.068	0.837
34	0.006	-0.026	25.073	0.867
35	-0.030	-0.076	25.209	0.889
36	0.004	-0.005	25.211	0.911

Figure 5: Correlogram of LN(Series)

Null Hypothesis: D(GOPPKR) has a unit root
Tr~~end~~ Specification: Intercept only
Break Specification: Intercept only
Break Type: Innovational outlier

Break Date: 92
Break Selection: Minimize Dickey-Fuller t-statistic
Lag Length: 0 (Automatic - based on Schwarz information criterion, maxlag=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-13.36103	< 0.01
Test critical values:	1% level	-4.949133	
	5% level	-4.443649	
	10% level	-4.193627	

*Vogelsang (1993) asymptotic one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(GOPPKR)
Method: Least Squares
Date: 12/27/18 Time: 08:37
Sample (adjusted): 3 95
Included observations: 93 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(GOPPKR(-1))	-0.337195	0.100082	-3.369194	0.0011
C	1.021862	0.632592	1.615357	0.1098
INCPTBREAK	0.959973	3.547447	0.270609	0.7873
BREAKDUM	25.22906	6.952169	3.628948	0.0005

R-squared	0.205056	Mean dependent var	0.997527
Adjusted R-squared	0.178260	S.D. dependent var	6.553826
S.E. of regression	5.941035	Akaike info criterion	6.443702
Sum squared resid	3141.335	Schwarz criterion	6.552631
Log likelihood	-295.6322	Hannan-Quinn criter.	6.487685
F-statistic	7.652523	Durbin-Watson stat	1.965372
Prob(F-statistic)	0.000131		

Figure 6: Test for Unit Root of LN(Data)

After taking the LN of the series the results are tested. Log transformations are used on time series data especially when the data series is not normal and showing high level of skewness. Log also harmonize the data and useful to make it stationary. The correlogram in Figure 7 displaying the results of the series with log and first difference. Data become stationary on this level and the results of Augmented Dickey Fuller test 9.82 ($p<0.000$) outturned that there is no unit root in the data. Autocorrelation among the residual terms checked by the Durban Watson statistics 2.02, shown that to accept the null hypothesis that there is no autocorrelation.

Null Hypothesis: D(LNGOPPKR) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.826765	0.0000
Test critical values:	1% level	-3.502238	
	5% level	-2.892879	
	10% level	-2.583553	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LNGOPPKR,2)
Method: Least Squares
Date: 12/27/18 Time: 18:34
Sample (adjusted): 3 95
Included observations: 93 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(LNGOPPKR(-1))	-0.986872	0.100427	-9.826765	0.0000
C	0.042530	0.022525	1.888133	0.0622

R-squared	0.514836	Mean dependent var	0.005999
Adjusted R-squared	0.509504	S.D. dependent var	0.305907
S.E. of regression	0.214243	Akaike info criterion	-0.222139
Sum squared resid	4.176913	Schwarz criterion	-0.167675
Log likelihood	12.32948	Hannan-Quinn criter.	-0.200148
F-statistic	96.56530	Durbin-Watson stat	2.027852
Prob(F-statistic)	0.000000		

Figure 7: Unit Root Test at LN (Data) at 1st Diff

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.013	0.013	0.0169	0.897
		2	-0.005	-0.005	0.0194	0.990
		3	0.001	0.001	0.0195	0.999
		4	0.058	0.058	0.3618	0.985
		5	0.012	0.011	0.3770	0.996
		6	0.067	0.068	0.8425	0.991
		7	0.008	0.007	0.8498	0.997
		8	0.043	0.041	1.0436	0.998
		9	-0.060	-0.063	1.4286	0.998
		10	0.069	0.064	1.9353	0.997
		11	-0.005	-0.011	1.9382	0.999
		12	-0.029	-0.037	2.0277	0.999
		13	-0.164	-0.162	5.0346	0.974
		14	-0.148	-0.163	7.5008	0.914
		15	0.035	0.040	7.6369	0.938
		16	0.039	0.033	7.8091	0.954
		17	-0.159	-0.142	10.780	0.868
		18	-0.065	-0.056	11.284	0.882
		19	-0.059	-0.033	11.701	0.898
		20	-0.069	-0.055	12.283	0.906
		21	0.039	0.068	12.475	0.926
		22	-0.057	-0.062	12.878	0.937
		23	0.108	0.147	14.368	0.916
		24	-0.033	0.006	14.508	0.934
		25	0.018	0.031	14.553	0.951
		26	0.100	0.076	15.875	0.939
		27	0.059	0.024	16.340	0.946
		28	0.039	0.058	16.545	0.957
		29	0.001	-0.003	16.545	0.969
		30	0.048	0.021	16.866	0.974
		31	0.096	-0.002	18.188	0.967
		32	0.093	0.086	19.454	0.960
		33	0.090	0.050	20.662	0.953
		34	0.055	0.030	21.118	0.959
		35	-0.069	-0.093	21.854	0.959
		36	-0.009	-0.034	21.866	0.969

Figure 8: Correlogram of LN(Data) at 1st Diff

3.2 Model Estimation

Analyzing the auto correlation function and partial auto correlation function we determine different models to fit ARIMA (p, d, q). But this approach is considered to be the subjective approach and it requires a lot of experience to determine p, d, q. Pictorial examination of the ACF and PACF can be quite unsatisfying and disingenuous. The correlograms do not depicts as anticipated. More generally it's a matter of trial and error to select a model from a variety of indicative models. Programmed model selection measures like the AIC and BIC delivers better approach to model selection as the selection is based on their minimization. Here different models are estimated a compared with each other by using the Akaike Information Criterion and Bayesian Information Criterion. Although comparison by using both criterion sometimes can sum up at varying models.

ARIMA (1,1,1), ARIMA (0,1,1), ARIMA (1,1,0), ARIMA (4,1,4), ARIMA (0,1,0), ARIMA (13,1,13), ARIMA (14,1,14) are estimated and there coefficients, log-likelihood, Akaike Information Criterion and Bayesian Information Criterions were computed and will be used to compare the models and for the selection of best mode for forecast.

```
ARIMA(1,1,1)

Coefficients:
         ar1      ma1
      0.9674  -0.9354
s.e.  0.0974   0.1406

sigma^2 estimated as 0.05008:  log likelihood=8.29
AIC=-10.59   AICc=-10.32   BIC=-2.96
```

```
ARIMA(0,1,1)

Coefficients:
         ma1
      0.0449
s.e.  0.1036

sigma^2 estimated as 0.05021:  log likelihood=7.73
AIC=-11.45   AICc=-11.32   BIC=-6.37
```

```
ARIMA(1,1,0)

Coefficients:
         ar1
      0.0472
s.e.  0.1062

sigma^2 estimated as 0.0502:  log likelihood=7.73
AIC=-11.46   AICc=-11.33   BIC=-6.38
```

```
ARIMA(4,1,4)

Coefficients:
          ar1     ar2      ar3      ar4     ma1      ma2     ma3     ma4
      -0.0529  1.6281  -0.0314  -0.8319  0.0581  -1.7096  0.0581  1.0000
s.e.   0.0890  0.1045   0.0730   0.0851  0.0831   0.0826  0.0844  0.0841

sigma^2 estimated as 0.04614:  log likelihood=12.3
AIC=-6.59   AICc=-4.45   BIC=16.29
```

```
ARIMA(13,1,13)

sigma^2 estimated as 0.04448:  log likelihood=19.13
AIC=15.74   AICc=38.65   BIC=84.41


ARIMA(14,1,14)

sigma^2 estimated as 0.04107:  log likelihood=21.25
AIC=15.5   AICc=42.68   BIC=89.25
```

3.3 Appropriate Model Selection and Forecasting

After comparison of the values of AIC and BIC we are selecting the model which has least value of both. That is our recommended model by R-package is ARIMA(0,1,0) while on the basis of least AIC (-13.81) and BIC (-8.72) criteria the most appropriated models on number 2nd and 3rd are ARIMA(1,1,0) having AIC (-11.46) AND BIC(-6.38) and ARIMA(0,1,1) having AIC (-11.45) and BIC (-11.37). So, the selected autoregressive moving averages models were utilized for forecasting of the future values.

```
ARIMA(0,1,0) with drift

Coefficients:
       drift
      0.0365
s.e.  0.0227

sigma^2 estimated as 0.04897:  log likelihood=8.9
AIC=-13.81   AICc=-13.68   BIC=-8.72
```

Forecasting is prediction based on historical values or activities. The events those impacted in past will continue to impact in future. Based on the best model selected ARIMA (0, 1, 0), ARIMA (0, 1, 1) and ARIMA (1, 1, 0) the forecasting is done for the future applications. Figures 9, 10 and 11.

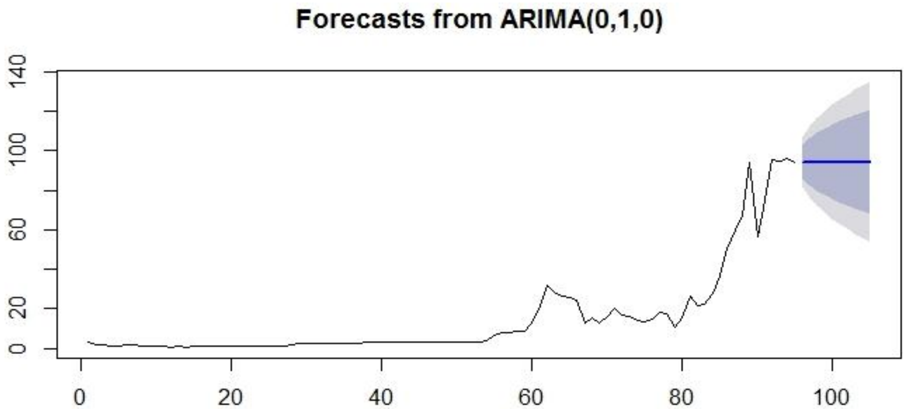


Figure 9: Forecasting by using Best Model-1

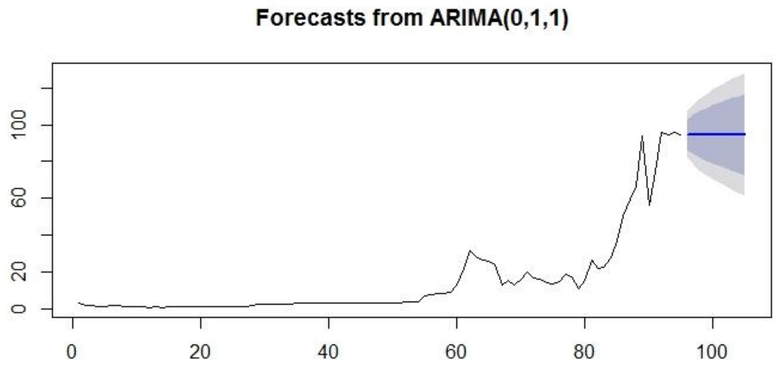


Figure 10: Forecasting by using Best Model-2

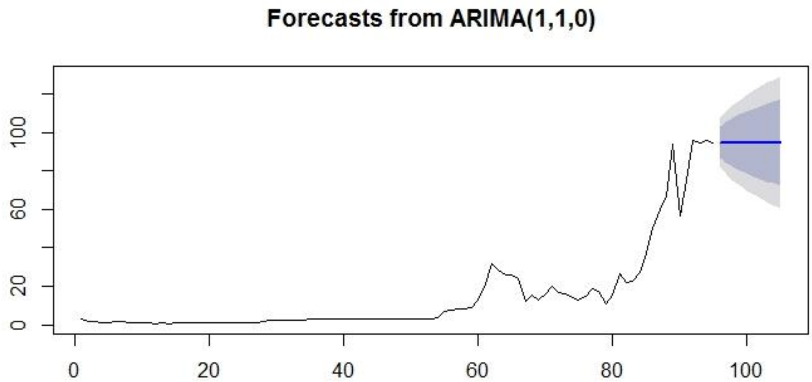


Figure 9: Forecasting by using Best Model-3

4. CONCLUSION

The data of average yearly gasoline oil prices ranging from 1920-2018 utilized for univariate time series analysis. The data is Pakistani rupees prices. First, by the time series plot and histogram, and Jarque Bera test the trend of overall series and the normality was tested. Stationarity was also checked on level and first difference. Data shown non-stationarity, so to make data normal and stationary LN transformation applied on the data. After the at 1^{st} difference the series depicted stationarity while the value of ADF(9.82) test and Durban Watson test (2.02) shown by statistics that the data at 1^{st} difference with LN is stationary and has no unit root as well as no auto correlation among the residual of the series. After the series become stationary, so by using the subjective approach on different level of Auto Regressive and Moving Averages ARIMA modeling was done. Among seven models of ARIMA, ARIMA (0, 1, 0), (1, 1, 0) and (0, 1, 1) outperformed as the best models based on the AIC and BIC criteria. As well as the selection of the best model was also done by the auto ARIMA model selection procedure that was also among the best illustrated models. Lastly based on the selected best models forecasting is performed for future applications.

5. ACKNOWLEDGEMENT

The authors wish to thank anonymous referees for their careful reading and suggestions which leads to the improvement over an earlier version of the paper. Thanks to the Department of Statistics, University of Wah, Wah cantt, Pakistan for their logistic and financial support for this research.

REFERENCES

1. Aamir, M. and Shabri, A.B. (2015). Modelling and Forecasting Monthly Crude Oil Prices of Pakistan: A Comparative Study of ARIMA, GARCH and ARIMA-GARCH Models. *Science International*, 27(3), 2365-2371.
2. Menash, E.K. (2015). Box-Jenkins Modelling and Forecasting of Brent Crude Oil Price. *MPRA Paper* No. 67748. https://mpra.ub.uni-muenchen.de/id/eprint/67748
3. Ali, A., Kashif, M. and Aslam, M. (2011). Estimation and Forecast of the Models for Stock Market Performance of the Oil & Gas Companies in Pakistan. *Pakistan Journal of Social Sciences (PJSS)*, 31(2), 345-363.
4. Zhao, G., Che, K. and Zhao, Y. (2012). A study on predicting coal market price in china based on time sequence models. International Journal of Business and Social Science, 3(4), 31-36.
5. Ye, M., Zyren, J. and Shore, J. (2005). A monthly crude oil spot price forecasting model using relative inventories. *International Journal of Forecasting*, 21(3), 491-501.
6. Ahmad, M.I. (2012) Modelling and forecasting Oman crude oil prices using Box–Jenkins techniques. *International Journal of Trade and Global Markets*, 5(1), 24-30.
7. Dickey, D.A. and Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*. 74(366a), 427-431.
8. Dawood, L.M. and Kareem, A.A. (2017). Statistical Methods for Modelling Petroleum Products Consumption in Iraq. *Journal of University of Thi-Qar*, 12(3), 197-213.
9. Akpanta, A.C. and Okorie, I.E. (2014). Application of Box-Jenkins Techniques in Modelling and Forecasting Nigeria Crude Oil Prices. *International Journal of Statistics and Applications*, 4(6), 283-291.
10. Khan, F. and Asghar, Z. (2010). *Modelling and Volatility analysis of oil companies in Pakistan*. InterStat, ISSN 1941-689X, Statistics on Internet.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 315-325

A NEW GENERALIZED ODD GUMBEL TYPE 2 EXPONENTIAL DISTRIBUTION

Ruqaya Naseer[1] and **Sharqa Hashmi**[2]
Department of Statistics, Lahore College for Women University
Email: [1]ruqaya.naseer@gmail.com
[2]sharqa.hashmi@gmail.com

ABSTRACT

We propose a new class of continuous distributions with four parameters named the generalized odd gumbel type 2 exponential distribution. For this distribution we derived reliability analysis including hazard rate function, survival function, reverse hazard rate function and cumulative hazard rate function. Also, some of its statistical properties are discussed including quantile function, moments and moment generating function. Furthermore, we derived order statistic of r^{th} smallest and n^{th} order statistic. We estimated the parameters of this distribution using ML estimation method. Application study is conducted for this distribution.

1. INTRODUCTION

Recently, many distributions have been made for modeling data. In statistical literature addition of one or more parameters in several classes is used to propose new distributions such as the Marshall-Olkin generated (MO-G) by Marshall and Olkin (1997), beta-G by Eugene et al. (2002). Kumaraswamy-G (Kw-G) by Cordeiro and de Castro (2011), transformer (T-X) by Alzaatreh et al (2013), Weibull-G by Bourguignon et al. (2014), Exponentiated half-logistic by Cordeiro et al. (2014a), logistic-X by Tahir et al. (2016), and Lomax generator by Cordeiro et al. (2014b).

Generalized odd distribution is basically originate from T-X family of distributions, defined by Alzaatreh et al. (2013) and its cdf is defined as

$$F(x) = \int_0^{W[G(x)]} r(t)\, dt \tag{1}$$

where $W[G(x)] = \left[\frac{G(x,\delta)^{\alpha}}{1-G(x,\delta)^{\alpha}}\right]$.

Let r (t) be the pdf of a random variable T $\in$[e, f], for $-\infty \leq e < f \leq \infty$.

The Gumbel Type-2 distribution is introduced by German mathematician Emil Gumbel (1958). The cumulative distribution (cdf) of Gumbel type-2 distribution is

$$G(x) = e^{-\theta x^{-\emptyset}} \quad 0 < x < \infty, \emptyset, \theta > 0 \tag{2}$$

and the corresponding probability density function (pdf) is given as:

$$g(x) = \emptyset\theta x^{-\emptyset-1} e^{-\theta x^{-\emptyset}} \quad 0 < x < \infty, \emptyset, \theta > 0 \tag{3}$$

where, θ is the scale parameter and $\emptyset$ is the shape parameter e shape parameter.

We proposed a new wider class of continuous distribution called the generalized odd Gumbel type 2 family of distributions. The cdf of GOG2 using T-X family of distribution Eq. (1) is given below

$$F(x;\alpha,\theta,\emptyset,\delta)=\int_0^{\left[\frac{G(x,\delta)^{\alpha}}{1-G(x,\delta)^{\alpha}}\right]}\emptyset\theta t^{-\emptyset-1}e^{-\theta t^{-\emptyset}}\,dt=e^{-\theta\left[\frac{G(x,\delta)^{\alpha}}{1-G(x,\delta)^{\alpha}}\right]^{-\emptyset}} \tag{4}$$

The probability density function corresponding to (4) is

$$\mathrm{f}(\mathrm{x};\alpha,\theta,\emptyset,\delta)=e^{-\theta\left[\frac{G(x,\delta)^{\alpha}}{1-G(x,\delta)^{\alpha}}\right]^{-\emptyset}}\left\{\alpha\theta\emptyset\left[\frac{G(x,\delta)^{\alpha}}{1-G(x,\delta)^{\alpha}}\right]^{-\emptyset-1}\left[\frac{G(x,\delta)^{\alpha-1}}{(1-G(x,\delta)^{\alpha})^{2}}\right]g(x,\delta)\right\} \tag{5}$$

$0< x<\infty\,,\alpha,\theta,\emptyset>0$

where g(x) and G(x) are the baseline pdf and cdf.

θ is the scale parameter, $\emptyset$ is one and α is another shape parameter $(\alpha>0)$ and δ is the parameter for baseline G.

The aim of this paper is to consider exponential distribution with four parameters called the generalized odd Gumbel type 2 exponential distribution. The cdf and pdf of exponential distribution are given as

$$G(x,\delta)=1-e^{-\lambda x}$$

and

$$g(x,\lambda)=\lambda e^{-\lambda x}\; x>0,\lambda>0 \tag{6}$$

In this article we consider a new way of exponential distribution with four parameters by replacing Eq. (6) in Eq. (5), called the generalized odd Gumbel type 2 exponential distribution.

Definition 1. A random variable X is said to have the generalized odd Gumbel type 2 Exponential distribution if it has the density:

$$f(x;\alpha,\theta,\varphi,\lambda)=e^{-\theta\left[\frac{(1-e^{-\lambda x})^{\alpha}}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{-\emptyset}}\left\{\alpha\lambda\theta\emptyset\left[\frac{(1-e^{-\lambda x})^{\alpha}}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{-\emptyset-1}\frac{(1-e^{-\lambda x})^{\alpha-1}}{\left(1-(1-e^{-\lambda x})^{\alpha}\right)^{2}}e^{-\lambda x}\right\} \tag{7}$$

The cumulative distribution function associated with (4) is given by

$$F(x;\alpha,\theta,\varphi,\lambda)=e^{-\theta\left[\frac{(1-e^{-\lambda x})^{\alpha}}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{-\emptyset}} \tag{8}$$

Figure 1, 2 and 3 illustrates some of the possible shapes of the pdf, cdf and hazard function of the generalized odd gumbel type 2 exponential distribution for selected values of the parameters $\emptyset$ and α respectively.

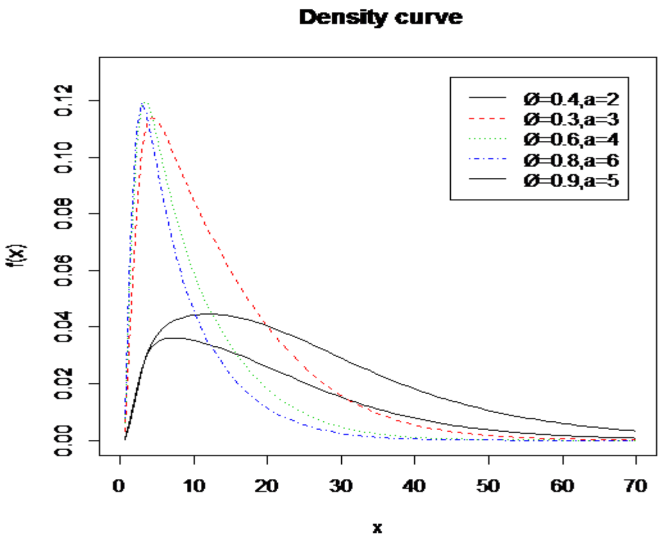


Fig. 1: The pdf of the Generalized odd Gumbel type 2 Exponential Distribution for Different values of Parameters ∅ and α.

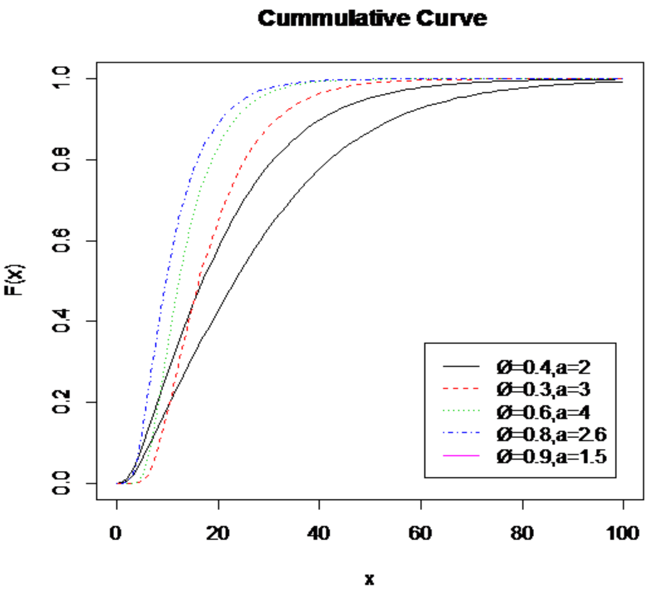


Fig. 2: The cdf of the Generalized odd Gumbel type 2 Exponential Distribution for Different values of Parameters ∅ and α.

2. RELIABILITY ANALYSIS

2.1 Survival Function

The reliability function (survival function) of the generalized odd Gumbel type 2 Exponential distribution is given by

$$S(x;\alpha,\theta,\varphi,\lambda) = 1 - F(x;\alpha,\theta,\varphi,\lambda) = 1 - e^{-\theta\left[\frac{(1-e^{-\lambda x})^{\alpha}}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{-\emptyset}} \quad (9)$$

2.2 Hazard Rate Function

The hazard rate function (failure rate) of a life- time random variable X the generalized odd Gumbel type 2 exponential distribution with four parameters is given by

$$h(x;\alpha,\theta,\emptyset,\lambda) = \frac{f(x;\alpha,\theta,\emptyset,\lambda)}{1-F(x;\alpha,\theta,\emptyset,\lambda)}$$

$$= \frac{e^{-\theta\left[\frac{(1-e^{-\lambda x})^{\alpha}}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{-\emptyset}}\left\{\alpha\theta\emptyset\lambda\left[\frac{(1-e^{-\lambda x})^{\alpha}}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{-\emptyset-1}\left[\frac{(1-e^{-\lambda x})^{\alpha-1}}{\left(1-(1-e^{-\lambda x})^{\alpha}\right)^{2}}\right]e^{-\lambda x}\right\}}{1-e^{-\theta\left[\frac{(1-e^{-\lambda x})^{\alpha}}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{-\emptyset}}} \quad (10)$$

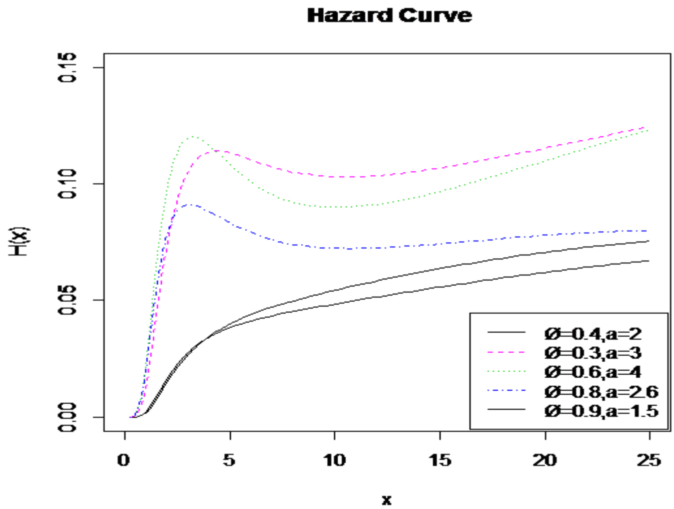


Fig. 3: The hazard rate Function of the Generalized odd Gumbel type 2 Exponential Distribution for Different values of Parameters $\emptyset$ and α

Also, the cumulative hazard rate function and reverse hazard rate function are as follows respectively.

$$H(x) = -\ln\left[1 - e^{-\theta\left[\frac{(1-e^{-\lambda x})^{\alpha}}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{-\emptyset}}\right] \quad (11)$$

$$R(x) = \left\{\alpha\theta\emptyset\lambda\left[\frac{(1-e^{-\lambda x})^{\alpha}}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{-\emptyset-1}\left[\frac{(1-e^{-\lambda x})^{\alpha-1}}{\left(1-(1-e^{-\lambda x})^{\alpha}\right)^{2}}\right]e^{-\lambda x}\right\} \quad (12)$$

2.3 Quantile Function

The quantile of any distribution is given by solving the equation

$$F\left(x_q\right) = q, 0 < q < 1$$

$$\mathrm{F}(x, \delta) = e^{-\theta\left[\frac{G(x,\delta)^{\alpha}}{1-G(x,\delta)^{\alpha}}\right]^{-\emptyset}} = \mathrm{q}$$

The quantile of the generalized odd gumbel type 2 exponential distribution is given by

$$\mathrm{x} = \frac{1}{\lambda} ln\left[\left(\frac{-\theta}{lnq}\right)^{-1/\emptyset} + 1\right]^{-1/\alpha} \tag{13}$$

By substituting q = 0.5 in (13) we will obtain the median of the generalized odd gumbel type 2 exponential distribution which is given by

$$\mathrm{x} = \frac{1}{\lambda} ln\left[\left(\frac{-\theta}{\ln(0.5)}\right)^{-1/\emptyset} + 1\right]^{-1/\alpha} \tag{14}$$

3. ORDER STATISTICS

Order statistics has an important role in quality control, reliability analysis and as well as in hydrological and extreme value analysis. Here we assume that $X_1, X_2, X_3, \ldots, X_{n-1}, X_n$ is a random sample from exponential distribution. Let, $X_{(1)}, X_{(2)}, \ldots, X_{(n-1)}, X_{(n)}$ be the ordered values of the preceding sample in non-decreasing order of magnitude.

3.1 n^th Order Statistic

The n^{th} order statistic of the generalized odd gumbel type 2 exponential distribution $X_{(n)} = \max(X_1, X_2, X_3, \ldots, X_{n-1}, X_n)$ is given by

$$f_{X_{(n)}}(x) = n[F(x)]^{n-1} f(x) = n\left[e^{-\theta\left[\frac{(1-e^{-\lambda x})^{\alpha}}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{-\emptyset}}\right]^{n}$$

$$\left\{\alpha\theta\emptyset\lambda\left[\frac{(1-e^{-\lambda x})^{\alpha}}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{-\emptyset-1}\left[\frac{(1-e^{-\lambda x})^{\alpha-1}}{(1-(1-e^{-\lambda x})^{\alpha})^{2}}\right]e^{-\lambda x}\right\}$$

3.2 Smallest Order Statistic

The smallest order statistic $X_{(1)} = \min(X_1, X_2, X_3, \ldots, X_{n-1}, X_n)$ is given by

$$f_{X_{(1)}}(x) = n[1-F(x)]^{n-1} f(x)$$

$$= n\left[1 - e^{-\theta\left[\frac{(1-e^{-\lambda x})^{\alpha}}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{-\emptyset}}\right]^{n-1} \times e^{-\theta\left[\frac{(1-e^{-\lambda x})}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{-\emptyset}}$$

$$\left\{\alpha\theta\emptyset\lambda\left[\frac{(1-e^{-\lambda x})^{\alpha}}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{-\emptyset-1}\left[\frac{(1-e^{-\lambda x})^{\alpha-1}}{(1-(1-e^{-\lambda x})^{\alpha})^{2}}\right]e^{-\lambda x}\right\}$$

3.3 rth Order Statistic

Generally, the distribution of the r^{th} order statistics with the generalized odd gumbel type 2 exponential distribution is as follows

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!\,(n-r)!}[F(x)]^{r-1}[1-F(x)]^{n-r}f(x)$$

$$= \frac{n!}{(r-1)!\,(n-r)!}\left[e^{-\theta\left[\frac{(1-e^{-\lambda x})^{\alpha}}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{-\emptyset}}\right]^{r} \times \left[1-e^{-\theta\left[\frac{(1-e^{-\lambda x})^{\alpha}}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{-\emptyset}}\right]^{n-r}$$

$$\times\left\{\alpha\theta\emptyset\lambda\left[\frac{(1-e^{-\lambda x})^{\alpha}}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{-\emptyset-1}\left[\frac{(1-e^{-\lambda x})^{\alpha-1}}{(1-(1-e^{-\lambda x})^{\alpha})^{2}}\right]e^{-\lambda x}\right\}$$

3.4 Useful EXPANSIONS

Based on general binomial and taylor expansion the pdf (7) can be expressed as:

$$f(x) = \sum_{i=k=0}^{\infty} w_{i,j} h_{\alpha(j-\emptyset i-\emptyset)}(x) \tag{15}$$

where

$$w_{i,j} = \frac{(-1)^{i}\binom{\emptyset i+\emptyset-1}{j}}{i!\,[j-\emptyset i-\emptyset]}$$

and

$$h_{\alpha(j-\emptyset i-\emptyset)}(x) = \theta^{2}\emptyset\alpha g(x)G(x)^{\alpha[j-\emptyset i-\emptyset]-1}$$

and

$$h_{\alpha}(x) = \alpha g(x)G(x)^{\alpha-1}$$

$G(x)^{\alpha-1}$ denote the pdf of exp-G distribution with power parameter α.

By integrating (15)

$$F(x) = \sum_{i=k=0}^{\infty} w_{i,j} H_{\alpha(j-\emptyset i-\emptyset)}(x) \tag{16}$$

where $H_{\alpha}(x) = G(x)^{\alpha}$.

Let $G(x)$ = 1-$[1-G(x)]$ and using binomial expansion, we have

$$G(x)^{\alpha[j-\emptyset i-\emptyset]} = \sum_{l=0}^{\infty}(-1)^{l}\binom{\alpha[j-\emptyset i-\emptyset]}{l}[1-G(x)]^{l}$$

After simplification

$$G(x)^{\alpha[j-\emptyset i-\emptyset]} = \sum_{k=0}^{\infty}\sum_{l=k}^{\infty}(-1)^{l+k}\binom{\alpha[j-\emptyset i-\emptyset]}{l}\binom{l}{k}G(x)^{k}$$

Then

$$F(x) = \sum_{k=0}^{\infty} b_k\, G(x)^{k}$$

where

$$b_k = \sum_{i=j=0}^{\infty} \sum_{l=k}^{\infty} w_{i,j} (-1)^{l+k} \binom{\alpha[j-\emptyset i-\emptyset]}{l} \binom{l}{k}$$

Theorem (3.4.1)

Let X be a random variable with pdf (7). The expectation is given by:

$$\mathrm{E(X)} = \frac{\alpha}{\lambda} \sum_{t=i=j=k=0}^{\infty} \sum_{l=k}^{\infty} \frac{(-1)^{i+l+k} \binom{\emptyset i+\emptyset-1}{j}}{i!\,[j-\emptyset i-\emptyset]} \binom{\alpha[j-\emptyset i-\emptyset]}{l} \binom{l}{k} \binom{\alpha-1}{t} \frac{1}{(t+1)^2}$$

Proof:

$$E(X) = \int_0^{\infty} x f(x)\, dx$$

$$= \alpha\lambda \sum_{t=i=j=k=0}^{\infty} \sum_{l=k}^{\infty} \frac{(-1)^{i+l+k} \binom{\emptyset i+\emptyset-1}{j}}{i!\,[j-\emptyset i-\emptyset]} \binom{\alpha[j-\emptyset i-\emptyset]}{l} \binom{l}{k} \binom{\alpha-1}{t} \int_0^{\infty} x e^{-\lambda x(t+1)}\, \mathrm{d(x)}$$

$$= \frac{\alpha}{\lambda} \sum_{t=i=j=k=0}^{\infty} \sum_{l=k}^{\infty} \frac{(-1)^{i+l+k} \binom{\emptyset i+\emptyset-1}{j}}{i!\,[j-\emptyset i-\emptyset]} \binom{\alpha[j-\emptyset i-\emptyset]}{l} \binom{l}{k} \binom{\alpha-1}{t} \frac{1}{(t+1)^2}$$

Theorem (3.4.2):

Let X be a random variable with pdf (7). The moment generating function is given by:

$$E(e^{tX}) = \int_0^{\infty} e^{tx} f(x) dx$$

$$= \alpha\lambda \sum_{t=i=j=k=0}^{\infty} \sum_{l=k}^{\infty} \frac{(-1)^{i+l+k} \binom{\emptyset i+\emptyset-1}{j}}{i!\,[j-\emptyset i-\emptyset]} \binom{\alpha[j-\emptyset i-\emptyset]}{l} \binom{l}{k} \binom{\alpha-1}{t} \frac{1}{-\lambda(t+1)+t}$$

Proof:

$$E(e^{tX}) = \alpha\lambda \sum_{t=i=j=k=0}^{\infty} \sum_{l=k}^{\infty} \frac{(-1)^{i+l+k} \binom{\emptyset i+\emptyset-1}{j}}{i!\,[j-\emptyset i-\emptyset]} \binom{\alpha[j-\emptyset i-\emptyset]}{l} \binom{l}{k} \binom{\alpha-1}{t} \int_0^{\infty} e^{tx}\, e^{-\lambda x(t+1)}\, dx$$

$$= \alpha\lambda \sum_{t=i=j=k=0}^{\infty} \sum_{l=k}^{\infty} \frac{(-1)^{i+l+k} \binom{\emptyset i + \emptyset - 1}{j}}{i!\,[j - \emptyset i - \emptyset]} \binom{\alpha[j - \emptyset i - \emptyset]}{l} \binom{l}{k} \binom{\alpha - 1}{t} \int_0^{\infty} e^{x(-\lambda t - \lambda + t)}\, dx$$

$$E(e^{tX}) = \alpha\lambda \sum_{t=i=j=k=0}^{\infty} \sum_{l=k}^{\infty} \frac{(-1)^{i+l+k} \binom{\emptyset i + \emptyset - 1}{j}}{i!\,[j - \emptyset i - \emptyset]} \binom{\alpha[j - \emptyset i - \emptyset]}{l} \binom{l}{k} \binom{\alpha - 1}{t} \frac{1}{-\lambda(t+1)+t}$$

Theorem (3.4.3):

Let X be a random variable with pdf (7). The rth moment of X about origin is defined as

$$\mathrm{E}\,(X^r) = \int_0^{\infty} X^r f(x)\; dx$$

$$= \alpha\lambda \sum_{t=i=j=k=0}^{\infty} \sum_{l=k}^{\infty} \frac{(-1)^{i+l+k} \binom{\emptyset i + \emptyset - 1}{j}}{i!\,[j - \emptyset i - \emptyset]} \binom{\alpha[j - \emptyset i - \emptyset]}{l} \binom{l}{k} \binom{\alpha - 1}{t} \int_0^{\infty} x^r e^{-\lambda x(t+1)}\; dx$$

4. ESTIMATION

In this subsection, interest is to define the parameter estimation of the generalized odd gumbel type 2 exponential distribution by maximum likelihood estimation.

Let, X_1, X_2,…, X_n be i.i.d random variables of size n. Then the log- likelihood function for this distribution is

$$\ell = \mathrm{n}\ln\theta + n ln\emptyset + n ln\alpha + (-\alpha\emptyset - 1) + (-\alpha\emptyset - 1) \sum_{i=0}^{n} \log\left(1 - e^{-\lambda x}\right) + (\emptyset - 1) \sum_{i=0}^{n} log s_i \lambda \sum_{i=0}^{n} log e^{-\lambda x} - \theta \sum_{i=0}^{n} p_i \quad (17)$$

where

$$s_i = 1 - (1 - e^{-\lambda x})^{\alpha}\ and\ p_i = \left[\frac{(1-e^{-\lambda x})^{\alpha}}{s_i}\right]^{-\emptyset}$$

The maximum likelihood estimates can be obtained by taking derivatives of (17) w.r.t parameters and equate to 0. The exact solution for unknown parameters is not possible analytically so the estimates are obtained by solving nonlinear equations simultaneously. The solution of nonlinear system is easier by iterative techniques common as Newton Raphson approach.

5. APPLICATION

Now we use a real data set to show that the generalized odd Gumbel type 2 exponential distribution can be a better model.

We consider a data set of the life of fatigue fracture of Kevlar 373/epoxy that are subject to constant pressure at the 90% stress level until all had failed, so we have complete data with the exact times of failure. The data is as follows:

0.0251, 0.0886, 0.0891, 0.2501, 0.3113, 0.3451, 0.4763, 0.5650, 0.5671, 0.6566, 0.6748, 0.6751, 0.6753, 0.7696, 0.8375, 0.8391, 0.8425, 0.8645, 0.8851, 0.9113, 0.9120, 0.9836, 1.0483, 1.0596, 1.0773, 1.1733, 1.2570, 1.2766, 1.2985, 1.3211, 1.3503, 1.3551, 1.4595, 1.4880, 1.5728, 1.5733, 1.7083, 1.7263, 1.7460, 1.7630, 1.7746, 1.8275, 1.8375, 1.8503, 1.8808, 1.8878, 1.8881, 1.9316, 1.9558, 2.0048, 2.0408, 2.0903, 2.1093, 2.1330, 2.2100, 2.2460, 2.2878, 2.3203, 2.3470, 2.3513, 2.4951, 2.5260, 2.9911, 3.0256, 3.2678, 3.4045, 3.4846, 3.7433, 3.7455, 3.9143,4.8073, 5.4005, 5.4435, 5.5295, 6.5541, 9.0960.

Table 1
Estimated Parameters of the GOG2-E, GOLLE, BE, KWE and Exponential Distribution for Data Set

Model	**ML Estimator**	**Standard Error**	**Log-Likelihood**
Generalized odd gumbel type 2 exponential distribution.	$\alpha = 0.0003501$ $\theta = 1.0028780$ $\lambda = 0.5715554$ $\emptyset = 0.9223371$	0.2243701 0.0393994 0.0001048 0.1243789	**106.1954**
Generalized odd log logistic exponential distribution.	$\alpha = 2.636$ $\theta = 0.453$ $\lambda = 0.155$	0.380 0.114 0.07	120.752
Beta exponential distribution	a=1.679 b=1.508 $\lambda = 0.484$	0.374 6.760 1.981	122.227
Kumaraswamy exponential distribution	a=1.556 b=2.448 $\lambda = 0.328$	0.401 6.065 0.691	122.094
Exponential distribution	$\lambda = 0.510.510$	0.058	127.114

Table 2 shows the values of -2l, AIC, CAIC and BIC for the data set. The better distribution corresponds to smaller -2l, AIC, CAIC and BIC values.

Table 2
Criteria for comparison

Model	-2 ℓ	AIC	CAIC	BIC
GOG2-E	212.391	220.391	221.332	219.914
GOLLE	241.505	247.505	248.005	254.497
BE	244.455	250.455	250.621	257.447
Kw	244.188	250.188	250.521	257.180
Exp.	254.228	248.643	249.143	258.995

Table 2. shows that the values of AIC CAIC and BIC for new distributions. The values of AIC CAIC and BIC are small as compared to other four distributions. The smaller values of AIC, CAIC, BIC shows that the GO-GT2-E distribution is more flexible and reliable than GOG2-E, GOLLE, BE, KwE and exponential distributions. The new distribution is better fit to real data sets than GOG2-E, GOLLE, BE, KwE and exponential distributions.

6. CONCLUSION

In this paper, we propose a new class of distribution called the generalized odd gumbel type 2 exponential distribution. For this distribution we derive some of its mathematical properties. We also calculate the reliability analysis such as survival function, hazard rate function, survival function, reverse and cumulative hazard rate function, Quantile function and moment generating function. We use the maximum likelihood approach to estimate the parameters of the proposed model.

The graphs of the probability density function and hazard rate function shows that the distribution increases at the initial stage and then decreases. The graph of cumulative distribution shows increasing trend. The graph shows that the distribution is positively skewed and uni-model for different values of parameters α and $\emptyset$.

Application study is also conducted. Table 1 shows the summary statistics and the estimates of parameters for GOG2-E, GOLLE, BE, KwE and exponential distributions using MLEs for real data set. Table 2 shows that the information criteria for comparison. The smaller values of AIC CAIC and BIC shows that the GO-GT2-E distribution is more flexible and reliable than GOG2-E, GOLLE, BE, KwE and exponential distributions.

REFERENCES

1. Alizadeh, M., Ghosh, I., Yousof, H.M., Rasekhi, M. and Hamedani, G.G. (2017). The Generalized Odd Generalized Exponential Family of Distributions: Properties, Characterizations and Application. *Journal of Data Science*, 16, 443-466.
2. Aljarrah, M.A., Lee, C. and Famoye, F. (2014). The Generating T-X Family of Distribution Using Quantile Functions. *Journal of Statistical Distributions and Applications*, doi: 10.1186/2195-5832-1-2.
3. Alzaatreh, A, Lee, C, Famoye, F (2013). A new method for generating families of continuous distributions. *Metron*, 71(1), 63-79.
4. Bourguignon, M., Silva, R.B. and Cordeiro, G.M. (2014). The Weibull-G family of probability distributions. *Journal of Data Science*, 12(1), 53-68.
5. Cordeiro, G.M., Alizadeh, M., Ramires, T.G. and Ortega, E.M.M. (2016). The Generalized Odd Half-Cauchy Family of Distributions: Properties and Applications. *Communication in Statistics-Theory and Methods*, 46, 5685-5705.
6. Cordeiro, M.G., Alizadeh, M., Ozel, G., Hosseini, B. and Altun, E. (2017). The Generalized Odd Log-Logistic Family of Distributions: Properties, Regression Models and Applications. *Journal of Statistical Computation and Simulation*, 87(5), 908-932.
7. Cordeiro, G.M. and de Castro, M. (2011). A new family of generalized distributions. *Journal of statistical Computation and Simulation*, 81(7), 883-898.
8. Cordeiro, G.M., Alizadeh, M. and Ortega, E.M. (2014a). The exponentiated half-logistic family of distributions: Properties and applications. *Journal of Probability and Statistics*, http://dx.doi.org/10.1155/2014/864396
9. Cordeiro, G.M., Ortega, E.M., Popović, B.V. and Pescim, R.R. (2014). The Lomax generator of distributions: Properties, minification process and regression model. *Applied Mathematics and Computation*, 247, 465-486.
10. Hosseini, B., Afshari, M. and Alizadeh, M. (2017). The Generalized Odd Gamma-G Family of Distributions: Properties and Applications. *Austrian Journal of Statistics*, 47, 69-89. Doi: 10.17713/ajs.v4i2.580
11. Korkmaz, M.C., Alizadeh, M., Yousof, H.M. and Shafique, N. (2018). The Generalized Odd Weibull Generated Family of Distributions: Statistical Properties and Applications. *Pakistan Journal of Statistics and Operation Research*, 14(3), 541-556.
12. Tahir, M.H., Cordeiro, G.M., Alzaatreh, A., Mansoor, M. and Zubair, M. (2016). The Logistic-X family of distributions and its applications. *Communications in Statistics-Theory and Methods*, 45(24), 7326-7349.
13. Qoshja, A. and Muca, M. (2018). A New Modified Generalized Odd Log-Logistic Distribution With Three Parameters. *Mathematical Theory and Modeling*, 8(1), 2224-5804.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 327-334

A STUDY ON EFFECTS OF CO-CURRICULAR ACTIVITIES ON STUDENT'S SOCIAL BEHAVIORS: A SECONDARY LEVEL STUDY

Aulliyah Jaweria
University of Education, Lahore, Pakistan
Email: jaweriamunir@gmail.com

ABSTRACT

The Present study was designed to investigate the effect of co-curricular activities on student's social behaviors at secondary level. The main objective of the study to determine whether students who are involved in co-curricular activities are better than students who do not participate in co-curricular activities .The purpose was tried to find the truth about views of students about the effects of co-curricular activities as well. The data was collected by convenient sampling technique from the students of private and public secondary schools of Lahore through questionnaire. The researcher distributed 150 questionnaires which was developed by researcher with the help of supervisor. The Questionnaire comprised of 25 items. The collected data was tabularized and analyzed with the help of SPSS to calculate how much objectives researcher achieved through this study. The results of the study indicated co-curricular activities had a positive effect on overall development of students. Students become more confident, responsible and social by participating in co-curricular activities.

KEY WORDS

Co-curricular activities, effects, student's social behaviors, secondary level.

INTRODUCTION

Co-curricular activities are the fundamental piece of instructive framework. *Kumar et al. (2004)* remarked that co-curricular activities hold a position of extraordinary significance in the field of instruction for the inside and out advancement of youngsters. Notice has been made in different instructive books, commission reports and instructive arrangement with respect to the strategy, project, activities and noteworthiness of these activities. They encourage included that for social, physical and otherworldly advancement co-curricular activities are essential.

Co-curricular activities are the activities performed by students that don't fall in the domain of the normal educational programs of instructive organization (Wikipedia). Once these were viewed as additional curricular activities however because of their acknowledgment of their significance, now these are called co-curricular activities. Whether these activities have any connection with scholarly accomplishment or not. These are vital in their own privilege because of numerous reasons. Numerous educationists trust that these dynamic build social communication, upgrade authority quality, give a possibility of sound diversion, and make students self-taught and sure.

Bog and Kleitman (2002) tried whether support in co-curricular activities impacts scholastic results notwithstanding when the impacts of a student's capacity, school, individual and family attributes, and various different variables are controlled. They observe that joining more co-curricular activities and investing more energy taking an interest in them is connected with higher evaluations, more troublesome courses choose, additional time spent on homework, more universities connected to, a higher probability of beginning and completing school, and a higher last degree earned, notwithstanding when different elements are controlled. Size of the example for this study was 12084 students which further expanded the unwavering quality of the outcomes.

Thompson and Austin (2003) found no huge relationship between co-curricular activities and the scholastic evaluations *Mahoney et al. (2003)* found a constructive relationship between co-curricular activities and between individual abilities, high desire and better consideration level. *Hollway (2002)* examined impact on inspiration and thought that it was sure as well. Essentially *Bauer and Liang, (2003)* indicated beneficial outcome on basic considering, social and individual development.

PURPOSE OF THE STUDY

The purpose of the study was to investigate the effects of co-curricular activities on student’s social behavior at secondary level .The objective pursued in this study. As follow;

- To investigate the perception of students about the social behaviors and influence of co-curricular activities.

Following are the Research Questions

- What are the effects of co-curricular activities on development of the student’s social behavior?
- What are the perceptions of secondary school students about the effects of co-curricular activities?
- What students expect by participating in co-curricular skills they develop/gain?

METHODOLOGY

The researcher designs the procedure carefully, comprehensive questionnaire was developed for the completion of study.

The type of research was descriptive. Descriptive research involved collecting data in order to answer these questions. The population of the study included secondary level students from public and private school in Lahore district. In this study researcher selected four public and four private schools for collecting data about the effects of co-curricular activities on student’s social behavior at secondary level.

INSTRUMENT

Researcher developed a questionnaire to collect data. Researcher collected the data from male and female students of public and private secondary schools.

FINALIZATION OF INSTRUMENT

Researcher did finalization of instrument and get validation of the questionnaire from the experts and supervisor. The questionnaire consisted of 25 items. Questionnaires were distributed among 120 private and public school students and explains the procedure of filling the questionnaire and told them about five point liker scales, researcher collected data through personal visits of all selected school.

DATA ANALYSIS PROCEDURE

Data was analyzed with the help of computer program Statistical Package for Social Sciences (SPSS). Result of data is presented in different tables.

Table 5.1
Student who Participate in Co-Curricular Activities are more Social

	Frequency	Percent	Mean	Standard Deviation
D.A	2	1.3		
S.D.A	3	2.0		
N	17	11.3		
A	67	44.7		
S.A	61	40.7		
Total	150	100	4.21	.824

Table 5.1: Shows that the mean is calculated 4.21. Majority of the students (80%) agreed with the statement that “by participating in co-curricular activities they become more social”.

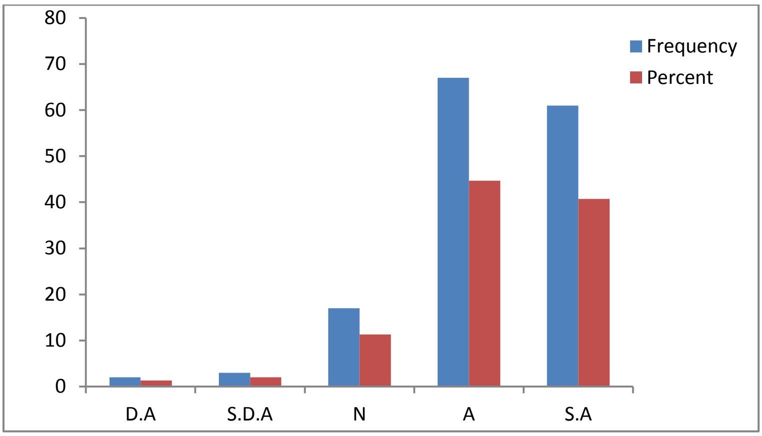


Table 5.2
Co-curricular Activities Enhance Student's Knowledge about the Society

	Frequency	Percent	Mean	Standard Deviation
S.D.A	2	1.3		
D.A	2	1.3		
N	12	12.0		
A	56	37.3		
S.A	72	48.0		
Total	150	100.0	4.29	.832

Table 5.2: Shows that the mean is 4.29. Most of the students (85%) agreed with the statement "co-curricular activities such as essay writing or debating enhance their knowledge about society".

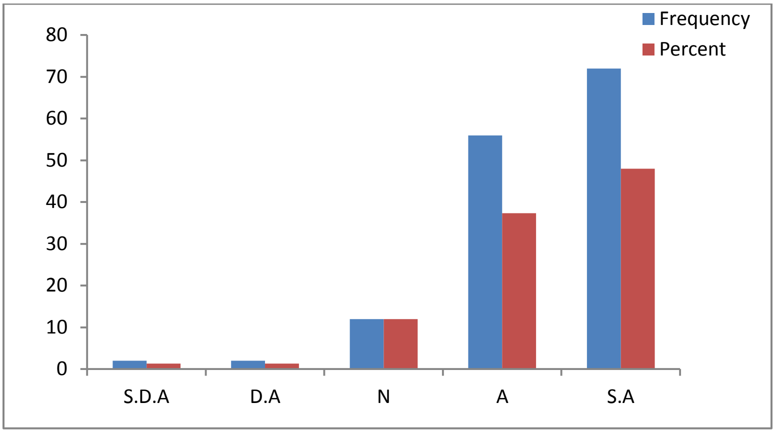


Table 5.3
These Activities Improve Critical Thinking of the Participant

	Frequency	Percent	Mean	Standard Deviation
S.D.A	3	2.0		
D.A	4	2.7		
N	28	18.7		
A	71	47.3		
S.A	44	29.3		
Total	150	100.0	3.99	.879

Table 5.3: Shows that the calculated mean is 3.99. Majority of the students (77%) agreed with the statement "co-curricular activities improve their critical thinking".

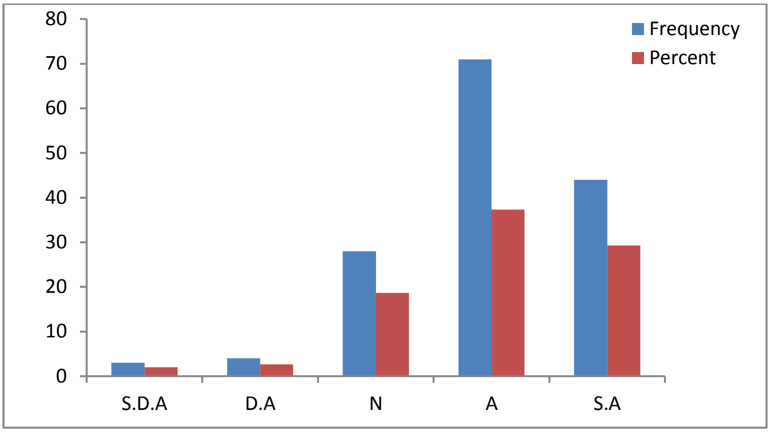


Table 5.4
Students who Participate Become more Confident

	Frequency	Percent	Mean	Standard Deviation
S.D.A	1	.7		
D.A	8	5.3		
N	20	13.3		
A	46	30.7		
S.A	74	49.3		
Missing	1	.7		
Total	150	100.0	4.23	.926

Table 5.4: Shows that the calculated mean is 4.23. This table shows that most of the students (79%) agreed with co-curricular activities increase their confidence level.

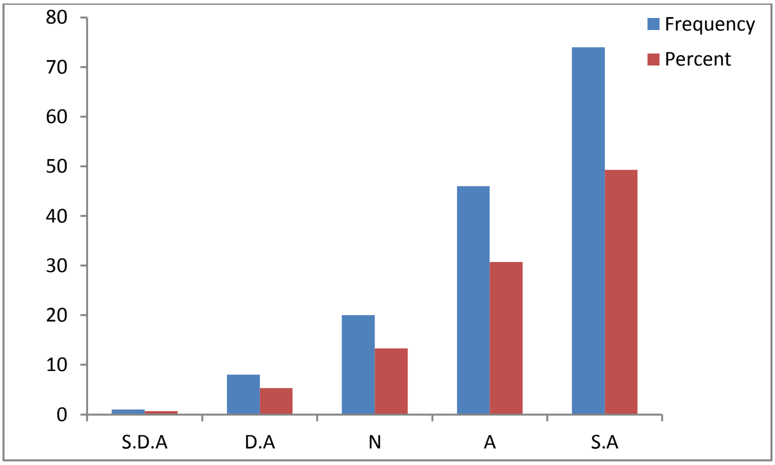


Table 5.5
Those Students Never Participate are Less Confident as Compare to the Participants

	Frequency	Percent	Mean	Standard Deviation
S.D.A	5	3.3		
D.A	9	6.0		
N	30	20.0		
A	59	39.3		
S.A	37	31.3		
Total	150	100.0	3.89	1.024

Table 5.5: Shows that calculated mean is 3.89. This table shows that majority of participants (70%) answered that feel more confident.

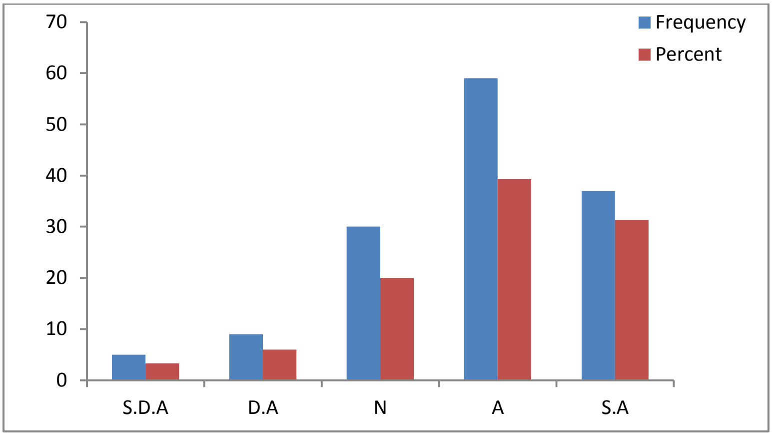
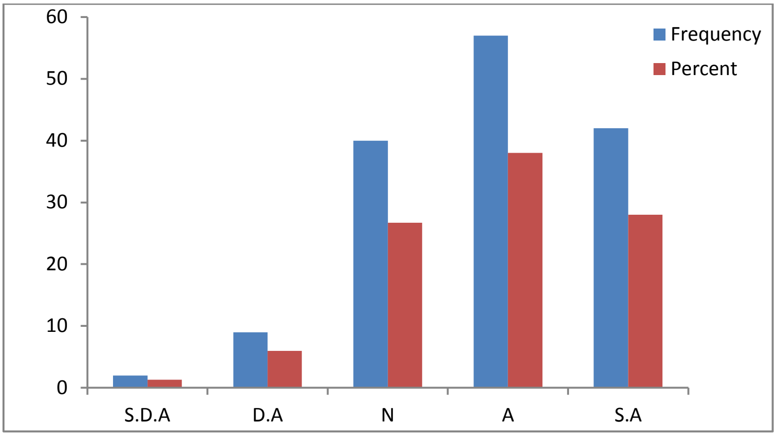


Table 5.6
Participants are More Easily Adjustable in any Situation

	Frequency	Percent	Mean	Standard Deviation
S.D.A	2	1.3		
D.A	9	6.0		
N	40	26.7		
A	57	38.0		
S.A	42	28.0		
Total	150	100.0	3.84	.966

Table 5.6: Shows that the calculated mean is 3.84. This table shows that most of the students (66%) answered that they easily adjustable in any situation.



CONCLUSION

Co-curricular activities have strongly effects the developing adaptation, self-confidence, honesty, sociability, understanding attitude, social obligation, and sense of responsibility of secondary school students. Co-curricular activities have significant impact on attitudes of secondary school students

The results obtained by the study reveal that co-curricular activities play a significant role in personality development of secondary school students and make them more confident, responsible and social in their lives.

The results of study support the findings of Maribeth (1990), Leslie (1992), Thomas and Morrison (1995). They reported the connection of athletics with personality traits. This study has verified their findings

Christopher (1998) reported that sports develop manners and sense of justice among school students. Wendy L. Lagon (2008) most of the time we are unable to identify the most influencing activity for our students but with the help of co-curricular activities and with students involvement we can see the bigger picture and can understand the needs of our child. The results of the study show that the students who participate in co-curricular activities behave inversely from those who don’t participate in these activities. In short co-curricular activities play an important and positive role in development of the students.

REFERENCES

1. Allar, B. (2008). *Core principals. Louisville Magazine*, 64-69.
2. Brown, R. (1999). *Extracurricular activity: How does participation encourage positive youth development*. *(Fact Sheet 99-32),* University of Nevada, Reno. Recuperado de http://www. unce. unr. edu/publications/files/cy/other/fs9932. pdf.
3. Casinger, J. (2010). *College extracurricular activities: the history of activities* [Article Directory].

4. CCSRI (2007). *Using Positive Engagement to Increase Student Achievement*. The Center for Comprehensive School Reform and Improvement (CCSRI), Washington, DC.
5. Davalos, D., Chavez, E. and Guardiola, R. (1999). The effects of extracurricular activity, ethnic identification, and perception of school on student dropout rates. Hispanic *Journal of Behavioral Sciences*, 21(1), 61-78.
6. Fredricks, J.A. and Eccles, J.S. (2006). Is extracurricular participation associated with beneficial outcomes? Concurrent and longitudinal relations. Developmental psychology, 42(4), 698-713.
7. Howie, L.D., Lukacs, S.L., Pastor, P.N., Reuben, C.A. and Mendola, P. (2010). Participation in activities outside of school hours in relation to problem behavior and social skills in middle childhood. *Journal of school Health*, 80(3), 119-125.
8. Logan, W.L. and Scarborough, J.L. (2008). Connections through clubs: Collaboration *and coordination of a school wide program. Professional School Counseling*, 12(2), 157-161.
9. McNeal Jr, R.B. (1999). Participation in high school extracurricular activities: Investigating school effects. *Social Science Quarterly*, 80(2), 291-310.
10. Morrissey, K. (2005). The relationship between out-of-school activities and positive youth development: An investigation of the influences of communities and family. *Adolescence*, 40, 67-85.
11. Sorge, C., Newsom, H. and Hagerty, J. (2000). Fun is not enough: Attitudes of Hispanic middle school students toward science and scientists. *Hispanic Journal of Behavioral Sciences,* 11(3), 332-346.
12. Zahida, H. (2012). Role of co-curricular activities for the performance of students at primary level in schools. *Interdisciplinary Journal of Contemporary Research in Business. Institute of Interdisciplinary Business Research*, 3(9).

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 335-346

FACTORS AFFECTING EXCLUSIVE BREASTFEEDING DURATION IN PAKISTAN: PARAMETRIC SURVIVAL MODELS

Hafiza Nasiha Sarwar[1], **Asifa Kamal**[2] and **Tooba Khan**[3]
Department of Statistics, Lahore College for Women University, Lahore, Pakistan
Email: [1]nasiha_935@yahoo.com
[2]asifa.k53@gmail.com
[3]tooba.khan@kinnaird.edu.pk

ABSTRACT

This study aims to determine factors significantly associated with exclusive breastfeeding duration in Pakistan. Duration of exclusive breastfeeding is the dependent variable and is built on evidence pertaining to the last child of the respondent from PDHS 2012-2013 and PDHS 2006-2007. This study assesses the socioeconomic and demographic factors associated with exclusive breastfeeding practice of mothers in Pakistan. Sixteen probable risk factors are included, such as region, maternal age, maternal occupation, wealth index, husband's education, respondent working, place of residence, prenatal visits to doctor, antenatal care by private doctor, assistance at delivery by doctor, place of delivery, size of baby at birth, gender of baby, watching television, delivery by caesarean section and preceding birth interval. Kaplan-Meier Survival Curves and Parametric Survival models are employed for the analysis of exclusive breastfeeding duration. For PDHS (2012-13 and 2006-2007) Weibull model and Exponential model are applied to the dataset and the factors that were significantly connected with duration of exclusive breastfeeding are region (Punjab), maternal age (15-19), wealth index (middle), place of residence (urban), antenatal care by private doctor and delivery by caesarean section.

1. INTRODUCTION

Breastfeeding plays a predominantly essential role in child survival. Exclusive breastfeeding has a direct contact with the health of infants. Breastfeeding still remains the preferred way for the majority of Pakistani women (Molla et al., 1993).

World Health Organization recommends exclusive breastfeeding for six months period (World Health Organization, 2009). During these six months of age, no additional liquids, semi-solid or breast milk substitute must be given to the infants except medicine. Exclusive breastfeeding is helpful for the health and welfare of infants and mothers. It is considered as an essential measure to shelter health and survival for infants (World Health Organization, 2013). It gives numerous paybacks to infants, from their overall fitness to the optimal advancement and growth (Khanna et al., 2010). Infants who are not exclusively breastfed for six months are more in the danger zone of different diseases. It is the secure and pure nourishment for a baby. Breastfed babies have less risk of coughing, asthma, depression, obesity, diabetes and gastrointestinal tract infections (Horta & Victora, 2013; Chung et al., 2007; Ajetunmobi et al., 2015; Monasta et al., 2010). These children have also enhanced intelligence and brain development (Gartner et al., 2005). Breastfed child always develop talking, sightseeing and walking skills faster.

Health practitioners should be familiar, that their approaches towards breastfeeding have a great impact on a woman's choice to breastfeed (Gilbert et al., 2013). Mothers trust their healthcare professional that's why they follow their opinion regarding breastfeeding. The practice of breastfeeding mostly lengthens the time between next pregnancies by delaying the ovary to its original position.

Duration of exclusive Breastfeeding for Pakistani infants is lower than WHO standards. There are numerous factors that affect the duration of exclusive breastfeeding among mothers. Objective of current paper is to examine factors affecting exclusive breastfeeding durations using the two national level surveys of Pakistan.

2. DATA AND METHODOLOGY

For this study, data from Pakistan Demographic and Health Survey (PDHS) 2012-2013 and 2006-2007 (sample consisted of last-born children in the two years preceding the survey) have been used to identify significant demographic, biological, public and health care factors affecting exclusive breastfeeding duration in Pakistan.

Variable for exclusive breastfeeding was generated through code available on DHS Users Forum. STATA 12.0 version was used for analysis of data extracted from PDHS (2006-07 and 2012-13) data file, named PKKR61FL. Total of 3946 and 5369 valid cases were obtained out of 13,558 and 10,023 for the study from PDHS (2012-13 and 2006-07). Relationship between exclusive breastfeeding duration and factors linked with exclusive breastfeeding duration are studied using Kaplan-Meier Survival Curves and Parametric Survival models.

3. RESULTS

Kaplan Meier Survival Plots (Appendix) shows that chances of exclusive breastfeeding falls more rapidly in PDHS (2012-13) for Punjab, Sindh and Baluchistan as compared to Khyber Pakhtunkhwa. For PDHS (2006-07) women in Baluchistan exclusively breastfeed their infants lower as compared to those who used to exclusively breastfeed in Punjab, Khyber Pakhtunkhwa and Sindh. It is depicted from Figure 4.3 and 4.4 that chances of exclusive breastfeeding are higher for younger women as compared to their older counterpart for both surveys. Difference becomes narrow for PDHS 2012-13 as compared to PDHS 2006-07. Survival plot 4.5 for PDHS (2012-13) shows that there is an increasing trend of exclusive breastfeeding in educated mothers' than uneducated mothers and survival plot (Figure 4.6) for PDHS (2006-07) shows that there is minor increasing trend of exclusive breastfeeding in educated mothers than uneducated mothers. It is depicted from Figure 4.7 that exclusive breastfeeding rates are not same for all categories of wealth index. Exclusive breastfeeding rates are lowest for poorest class as compared to middle and richest wealth index and it can be seen from Figure 4.8 that mothers who belonged to middle and poorest class had higher chances of exclusive breastfeeding as compared to richest mothers.

Survival plot for husband's education 4.9 shows that, exclusive breastfeeding rates are higher for those women whose husbands were educated. Kaplan-Meier survival plot 4.11 shows that mothers may experience a number of barriers to breastfeed their infants at workplaces. Working mothers has low chances of breastfeeding as compared to non-working women. Probability of exclusive breastfeeding is higher for those mothers who had visited for antenatal care than those who had never visited for recent survey. Figure 4.16 depicts the same that mothers with no antenatal visits have the lower survival rate of exclusive breastfeeding duration for later survey. Difference in the probabilities of exclusive breastfeeding between the mothers who had visited for antenatal care and those who had not is more for PDHS 2012-13 as compared to PDHS 2006-07. Mothers who had assistance at delivery by doctors had slightly low chances of exclusive breastfeeding as compared to those whose delivery was not assisted by doctor. It seems true that the transition from traditional to modern societies has prompted a move away from exclusive breastfeeding of long duration. But in Pakistani society effect is minor. For PDHS (2006-07) both categories of assistance at delivery by doctor work almost same. It is observed from Figure 4.19 that children who were born in government sector had low probability of exclusive breastfeeding as compared to private hospital and home deliveries. Kaplan-Meier survival plot 4.20 shows that all categories place of delivery shows nearly same trend for PDHS (2006-07). The Kaplan-Meier Survival plots of exclusive breastfeeding for various categories of size of baby for PDHS (2012-13) shows almost same trend and survival plot 4.22 shows that infants whose size was small at the time of birth had visibly low chances of exclusively breastfeeding as compared to children whose size at birth was average or large. According to Figure 4.23 in PDHS (2012-13) nominal difference in the Figures for exclusive breastfeeding is observed for both genders. Exclusively breastfeeding female infants is observed to be slightly higher than male infants and Kaplan-Meier Survival plots (Figure 4.24) of duration of exclusive breastfeeding for gender of baby shows nearly same trend in PDHS (2006-07).

For both the data sets it can be seen that chances of exclusive breastfeeding are observed to be less in cesarean birth as compared to children who were delivered normal. The Kaplan-Meier Survival plots of exclusive breastfeeding for various categories of media exposure and preceding birth interval shows almost same trend for PDHS (2012-13) and PDHS (2006-07).

Exclusive breastfeeding duration is response variable in this study. Those women who had not started weaning are treated as censored observations. The objective of the current study is to use appropriate model for studying the effect of various socioeconomic and demographic variables on exclusive breastfeeding duration. It necessitates determining the appropriate survival distribution for response variable. Exponential, Weibull, Gamma, Lognormal, and Generalized Gamma distribution are potential survival distribution which can be used to model duration of exclusive breastfeeding. Log likelihood, AIC and BIC are used to identify the best distribution for exclusive breastfeeding duration. The Weibull model for exclusive breastfeeding duration is identified as best model for PDHS (2012-13) and Exponential model for PDHS (2006-07).

Table 1
Parameter Estimates for Duration of Exclusive Breastfeeding, Weibull Model for (PDHS, 2012-13) and Exponential Model for (PDHS, 2006-07)

Parameters		(PDHS, 2012-13)		(PDHS, 2006-07)	
		Hazard Ratio	Sig.	Hazard Ratio	Sig.
Factors		Weibull Model		Exponential Model	
Region	Punjab	0.9858	0.001*	0.8457	0.023*
	Sindh	0.5585	0.006*	0.9511	0.145
	KPK	0.2210	0.101	0.721	0.101
	Baluchistan	1.000	-	1.000	-
Maternal Age	15-19	0.899	0.000*	0.9229	0.010*
	20-34	1.128	0.065	0.724	0.046*
	35-49	1.000	-	1.000	-
Maternal Education	Un-Educated	1.108	0.121	2.367	0.220
	Educated	1.000	-	1.000	-
Wealth Index	Poorest	0.9305	0.080	1.235	0.022*
	Middle	1.002	0.003*	0.702	0.003*
	Richest	1.000	-	1.000	-
Husband's Education	Un-Educated	1.264	0.06	2.214	0.0087*
	Educated	1.000	-	1.000	-
Respondent working	No	0.3152	0.103	0.512	0.1003
	Yes	1.000	-	1.000	-
Place of Residence	Urban	1.01	0.001*	1.001	0.001*
	Rural	1.000	-	1.000	-
Prenatal Visits to Doctor	No	2.065	0.241	1.154	0.0141*
	Yes	1.000	-	1.000	-
Antenatal Care by Private Doctor	No	0.9781	0.004*	1.561	0.0029*
	Yes	1.000	-	1.000	-
Assistance at Delivery by Doctor	No	0.835	0.147	0.835	0.147
	Yes	1.000	-	1.000	-
Place of Delivery	Home	0.831	0.071	0.980	0.021*
	Private Sector	1.02	0.004*	1.102	0.865
	Govt. Sector	1.000	-	1.000	-
Size of Baby at Birth	Large	1.767	0.901	1.160	0.901
	Average	0.472	0.004*	0.972	0.270
	Small	1.000	-	1.000	-
Gender of Baby	Male	1.090	0.34	1.023	0.223
	Female	1.000	-	1.000	-
Watching Television	No	0.619	0.068	0.923	0.168
	Yes	1.000	-	1.000	-
Delivery by Caesarean Section	No	0.014	0.014*	0.121	0.0001*
	Yes	1.000	-	1.000	-
Preceding Birth Interval	Less than 24 months	0.991	0.007*	0.910	0.107
	Greater than or equal to 24 months	1.000	-	1.000	-

Table 1 indicates that according to the Weibull model for PDHS (2012-13); region (Punjab, Sindh), maternal age (15-19), wealth index (middle), place of residence (urban), antenatal care by private doctor, place of delivery (private sector), size of baby at birth (average), delivery by caesarean section and preceding birth interval are significantly associated with duration of exclusive breastfeeding at 5% level of significance. However, remaining factors are not significantly associated with duration of exclusive breastfeeding. It can also be observed from Table 1 that according to the exponential model fitted using PDHS (2006-07); region (Punjab), maternal age (15-19, 20-34), wealth index (poorest, middle), husband's education, place of residence (urban), prenatal visits to doctor, antenatal care by private doctor, place of delivery (home) and delivery by caesarean section are significantly associated with duration of exclusive breastfeeding at 5% level of significance.

4. COMMENTS AND CONCLUSION

Pakistani mothers have lower breastfeeding practices than World Health Organization recommendations. This study indicated that lesser exercise of exclusive breastfeeding certainly projects the forbidding situation of awareness regarding significance of exclusive breastfeeding practice in Pakistan.

Survival analysis is used to study the trend of exclusive breastfeeding duration with different demographic and socioeconomic factors for both surveys. Maternal age (younger and middle) , antenatal care by private doctor, assistance at delivery by doctor, normal deliveries and preceding birth interval showed a good behaviour, that is, mothers exclusively breastfeed their infants for six months for PDHS 2012-13. Mothers who were in the younger or middle age group, taking antenatal care by private doctor, having large or average size of baby and normal delivery, exclusively breastfeed their infants for six months, for PDHS (2006-07).

Exponential and Weibull models are identified as best model to fit the current data for PDHS 2006-07 and PDHS 2012-13 respectively. Overall, findings of the present study suggest calls for promotion of exclusive breastfeeding programs in Pakistan among older mothers, especially those aged between 35-49 years of age. Previous evidence on influence of increasing maternal age on duration of breastfeeding is mixed, with studies reporting either positive (Bautista, 1997) or negative (Nath & Goswami, 1997) or no influence (Giashuddin & Kabir, 2004). Financial status could have great impact on physical as well as emotional values of mothers (Venancio & Monteiro, 2006). Generally it is a perception that when there is a rise in living standards then breast milk substitutes are used in a large variety meaning thereby that women of higher status chose not to exclusively breastfeed their infants. Comparison of socioeconomic status showed that mothers from lower socioeconomic class tend to breastfeed exclusively for shorter periods for PDHS 2006-07 while for PDHS 2012-13 findings are inverse. Urban women tend to terminate exclusive breastfeeding earlier as compared to rural women in both surveys.

Prenatal and antenatal contact with health professionals has been recognized as a factor that helps give confidence to women to establish early breastfeeding. Majority of women regularly attended their scheduled prenatal meetings since doctors made them aware of the significance of maintaining regular appointments for a better, healthy and successful pregnancy. Place of delivery is one of the predictors of exclusive breastfeeding practice. From PDHS (2006-07 and 2012-13), delivery at home is associated with breastfeeding of longer duration. Health practitioners who are involved in delivery of babies usually gave advertising materials as well as complimentary testers of infant formula to the mothers (Adair et al., 1993). Many hospitals do not inspire mothers to exclusively breastfeed their infant, which results in the discouragement of breastfeeding. Infants recognized to be non-average size are less likely to exclusive breastfeed by their mothers. Immature delivery may affect every condition of the infant's behavior and care. In different countries, similar pattern have adopted and infants with low birth weight are always breastfeed for minimum duration. Sometimes mothers may practice a number of different barriers to breastfeed their infant, particularly their lack of confidence, poor sucking, illness or baby, infant being kept in nursery for intensive care (Mulready & Sackoff, 2013). Mother wanted longer recovery time after cesarean birth and stay in hospital longer could have an impact on initiating breastfeeding. As revealed by PDHS (2012-13), preceding birth interval is a significant feature affecting exclusive breastfeeding duration in Pakistan.

Finally, it is concluded that in both surveys mothers who were Punjabi, young, not visited private doctor for antenatal care and who delivered babies through normal deliveries had longer exclusive breastfeeding duration. Urban women ceased breastfeeding early as compared to rural women and both surveys provided same evidence. Private medical centers, cesarean births reflect impact of urbanization on breastfeeding practices. Forces that work behind urbanization are education, financial autonomy and engagement in job market. Study recommends promotion of importance of exclusive breastfeeding at private health care centers, through gynecologists in the urban areas.

REFERENCES

1. Adair, L.S., Popkin, B.M. and Guilkey, D.K. (1993). The duration of breast-feeding: how is it affected by biological, socio-demographic, health sector, and food industry factors? *Demography*, 30(1), 63-80.
2. Ajetunmobi, O.M., Whyte, B., Chalmers, J., Tappin, D.M., Wolfson, L., Fleming, M., MacDonald, A., Wood, R. and Stockton, D.L. (2015). Breastfeeding is associated with reduced childhood hospitalization: evidence from a Scottish Birth Cohort (1997-2009). *The Journal of Pediatrics*, 166(3), 620-625.
3. Bautista, L.E. (1997). Factors associated with the initiation of breast feeding by women in the Dominican Republic. *Revista Panamericana de Salud Publica = Pan American Journal of Public Health*, 1(3), 200-207.
4. Chung, M., Raman, G., Chew, P., Magula, N., Trikalinos, T. and Lau, J. (2007). Breastfeeding and maternal and infant health outcomes in developed countries. *Evid. Technol. Asses*., 153(153), 1-186.
5. Gartner, L.M., Morton, J., Lawrence, R.A., Naylor, A.J., O'Hare, D., Schanler, R.J. and Eidelman, A.I. (2005). Breastfeeding and the use of human milk. *Pediatrics*, 115(2), 496-506.
6. Giashuddin, M.S. and Kabir, M. (2004). Duration of breastfeeding in Bangladesh. *Indian Journal of Medical Research*, 119, 267-272.

7. Gilbert, N.L., Auger, N., Wilkins, R. and Kramer, M.S. (2013). Neighborhood income and neonatal, post neonatal and sudden infant death syndrome (SIDS) mortality in Canada, 1991-2005. *Canadian Journal Public Health*, 104(3), 187-192.
8. Horta, B.L. and Victora, C.G. (2013). *Long-term effects of breastfeeding-a systematic* review, WHO Library Cataloguing-in-Publication Data, World Health Organization.
9. Khanna, R., Hota, P. and Lahariya, C. (2010). Health research strengthening and operational research needs for improving child survival in India. *Indian Journal of Pediatrics*, 77(3), 291-299.
10. Lee, E.T. and Wang, J. (2003). *Statistical methods for survival data analysis* (Vol. 476). John Wiley & Sons.
11. Molla, A., Badruddin, S.H., Khurshid, M., Molla, A.M., Rahaman, F.N., Durrani, S., Suria, A., Snyder, J.D. and Hendricks, K. (1993). Vitamin A status of children in the urban slums of Karachi, Pakistan, assessed by clinical, dietary, and biochemical methods. *The American Journal of Tropical Medicine and Hygiene*, 48(1), 89-96.
12. Monasta, L., Batty, G.D., Cattaneo, A., Lutje, V., Ronfani, L., Van Lenthe, F.J. and Brug, J. (2010). Early-life determinants of overweight and obesity: a review of systematic reviews. *Obesity Reviews*, 11(10), 695-708.
13. Mulready-Ward, C. and Sackoff, J. (2013). Outcomes and factors associated with breastfeeding for less than 8 weeks among preterm infants: Findings from 6 states and NYC, 2004–2007. *Maternal and Child Health Journal*, 17(9), 1648-1657.
14. Nath, D.C. and Goswami, G. (1997). Determinants of breastfeeding patterns in an urban society of India. *Human Biology*, 69, 557-573
15. Venancio, S.I. and Monteiro, C.A. (2006). Individual and contextual determinants of exclusive breast-feeding in Sao Paulo, Brazil: a multilevel analysis. *Public Health Nutrition*, 9(1), 40-46.
16. World Health Organization. (2009). Acceptable medical reasons for use of breast-milk substitutes (No. WHO/FCH/CAH/09.01). Geneva: World Health Organization.
17. World Health Organization. (2013). Nutrition. Retrieved from: http://www.who.int/nutrition/topics/infantfeeding_recommendation/en/

APPENDIX

Figure 4.1 Kaplan-Meier Survival Plot for Region

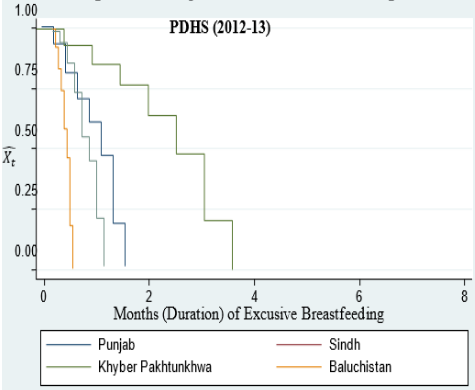


Figure 4.2 Kaplan-Meier Survival Plot for Region

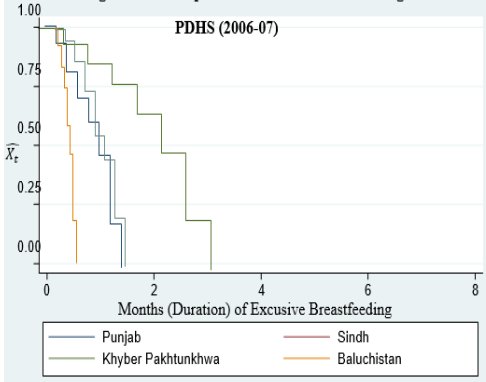


Figure 4.3 Kaplan-Meier Survival Plot for Maternal Age

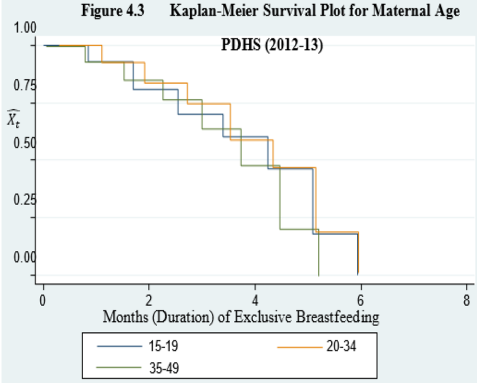


Figure 4.4 Kaplan-Meier Survival Plot for Maternal Age

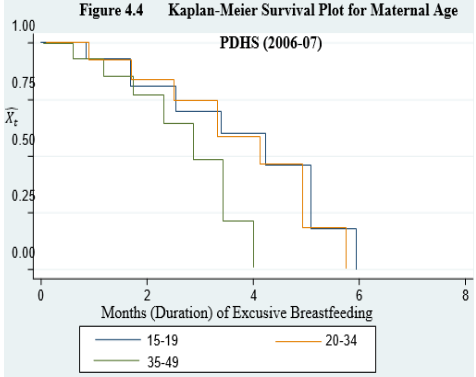


Figure 4.5 Kaplan-Meier Survival Plot for Maternal Education

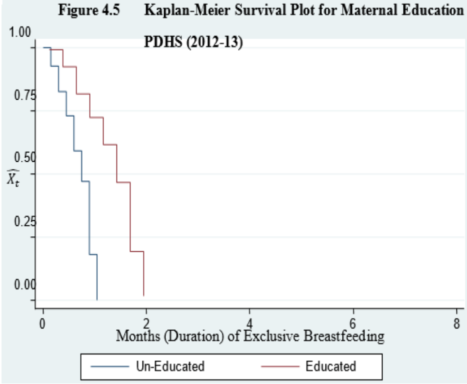
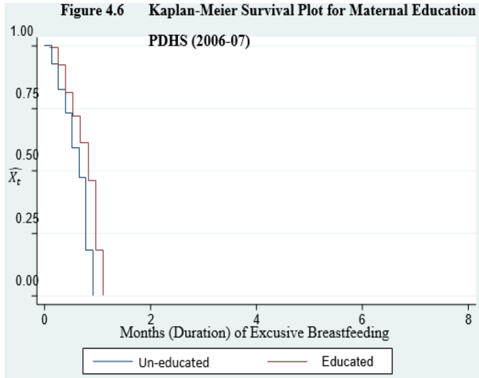


Figure 4.6 Kaplan-Meier Survival Plot for Maternal Education



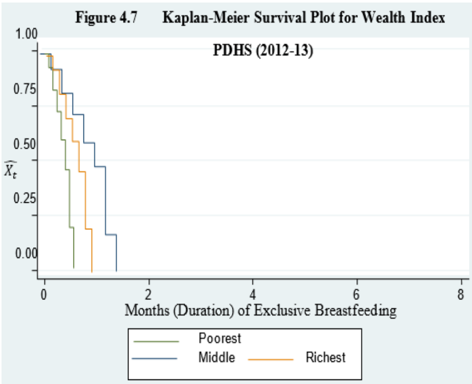


Figure 4.7 Kaplan-Meier Survival Plot for Wealth Index PDHS (2012-13)

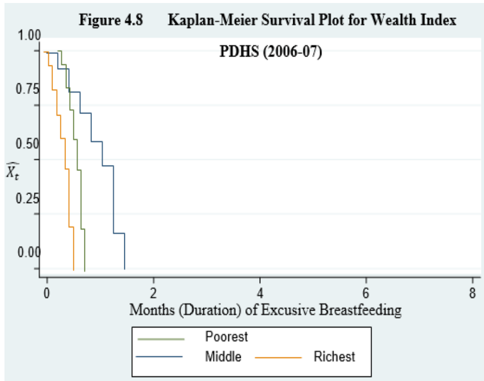


Figure 4.8 Kaplan-Meier Survival Plot for Wealth Index PDHS (2006-07)

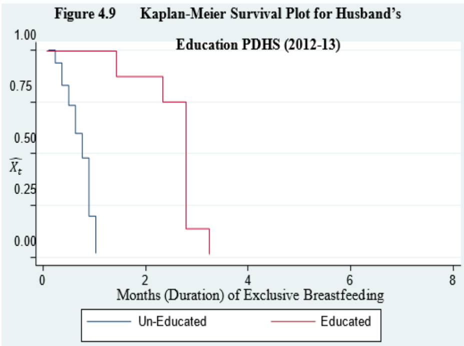


Figure 4.9 Kaplan-Meier Survival Plot for Husband's Education PDHS (2012-13)

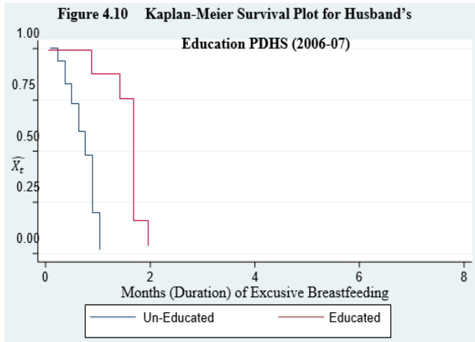


Figure 4.10 Kaplan-Meier Survival Plot for Husband's Education PDHS (2006-07)

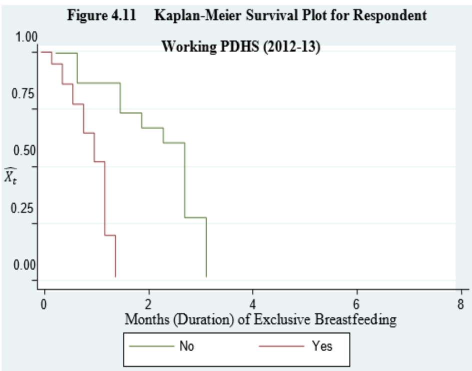


Figure 4.11 Kaplan-Meier Survival Plot for Respondent Working PDHS (2012-13)

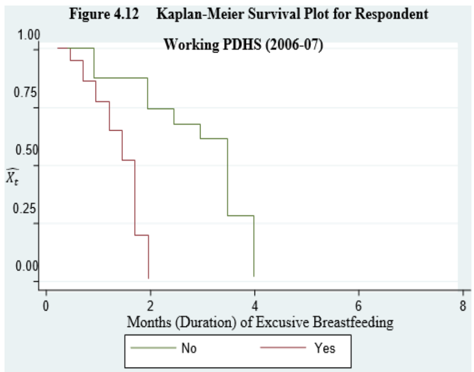


Figure 4.12 Kaplan-Meier Survival Plot for Respondent Working PDHS (2006-07)

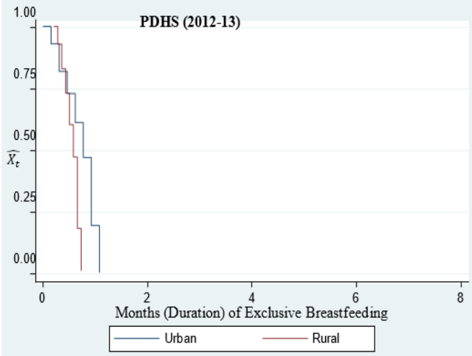


Figure 4.13 Kaplan-Meier Survival Plot for Place of Residence PDHS (2012-13)

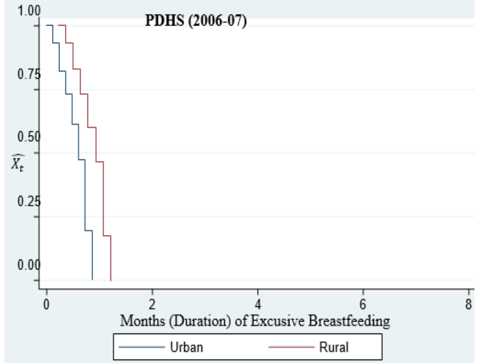


Figure 4.14 Kaplan-Meier Survival Plot for Place of Residence PDHS (2006-07)

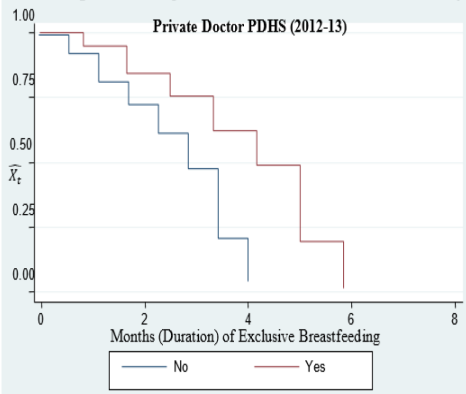


Figure 4.15 Kaplan-Meier Survival Plot for Antenatal Care by Private Doctor PDHS (2012-13)

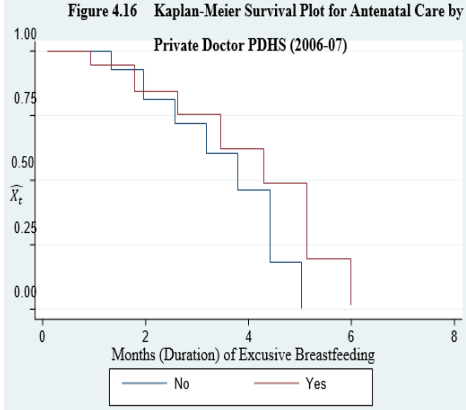


Figure 4.16 Kaplan-Meier Survival Plot for Antenatal Care by Private Doctor PDHS (2006-07)

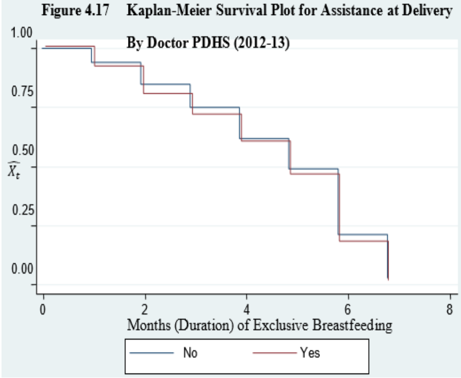


Figure 4.17 Kaplan-Meier Survival Plot for Assistance at Delivery By Doctor PDHS (2012-13)

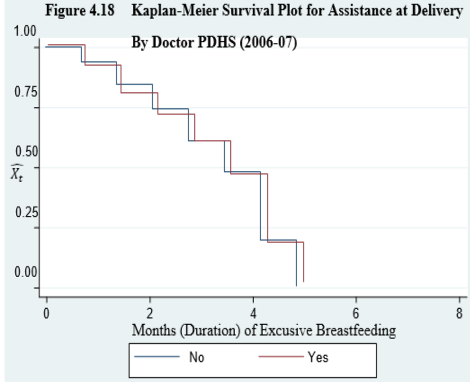


Figure 4.18 Kaplan-Meier Survival Plot for Assistance at Delivery By Doctor PDHS (2006-07)

Figure 4.19 Kaplan-Meier Survival Plot for Place of Delivery

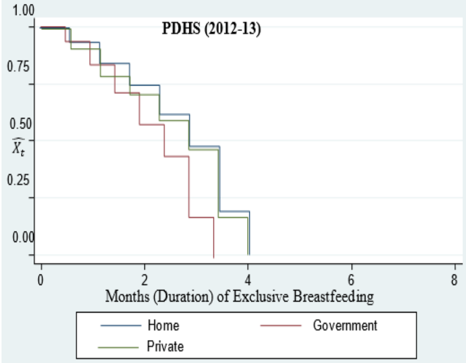


Figure 4.20 Kaplan-Meier Survival Plot for Place of Delivery

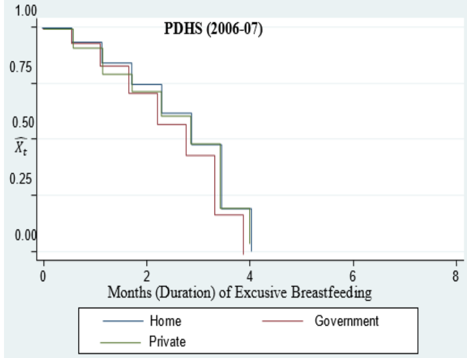


Figure 4.21 Kaplan-Meier Survival Plot for Size of Baby at Birth

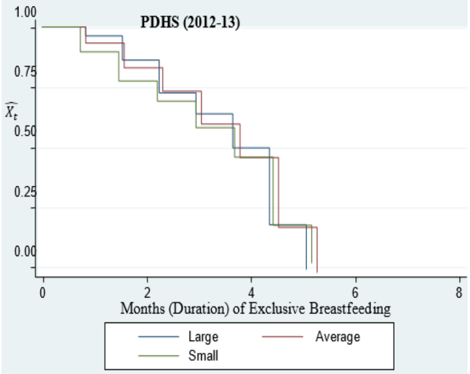


Figure 4.22 Kaplan-Meier Survival Plot for Size of Baby at Birth

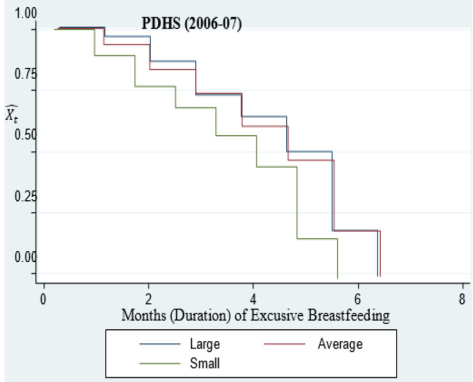


Figure 4.23 Kaplan-Meier Survival Plot for Gender of Baby

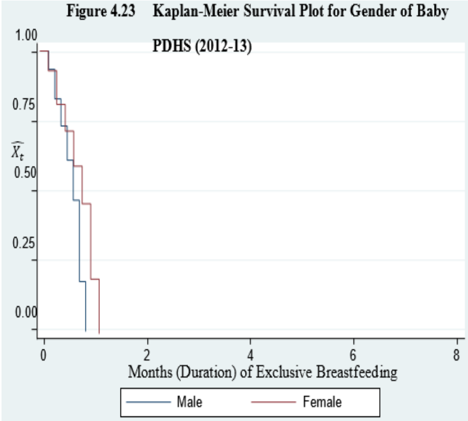
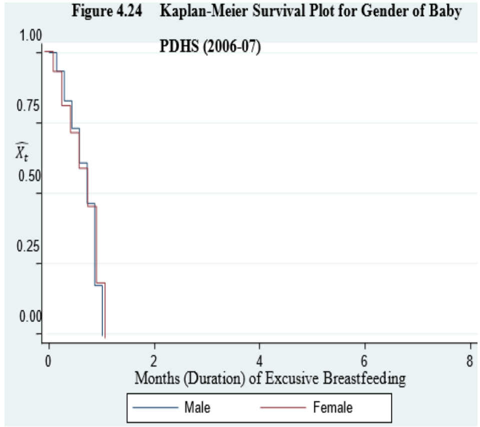


Figure 4.24 Kaplan-Meier Survival Plot for Gender of Baby



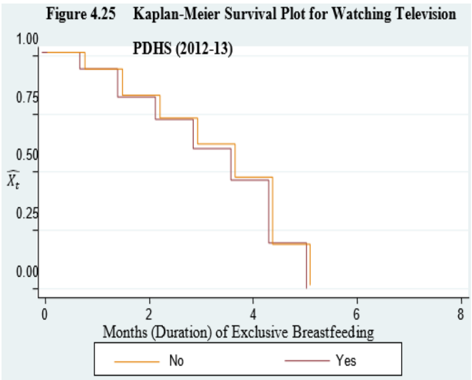


Figure 4.25 Kaplan-Meier Survival Plot for Watching Television PDHS (2012-13)

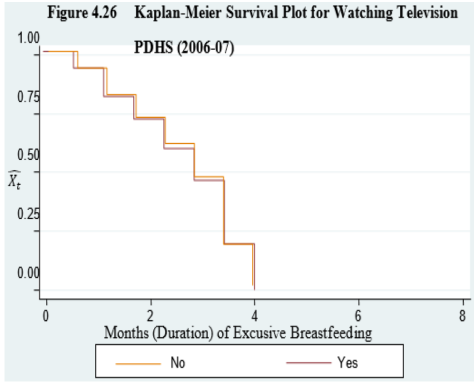


Figure 4.26 Kaplan-Meier Survival Plot for Watching Television PDHS (2006-07)

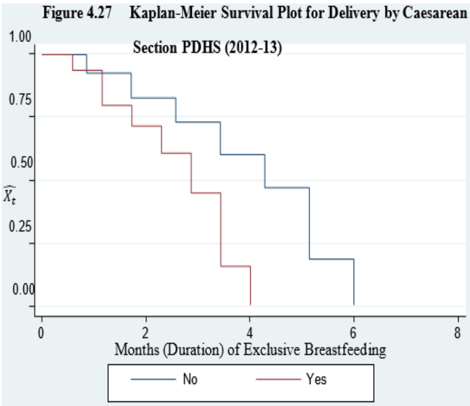


Figure 4.27 Kaplan-Meier Survival Plot for Delivery by Caesarean Section PDHS (2012-13)

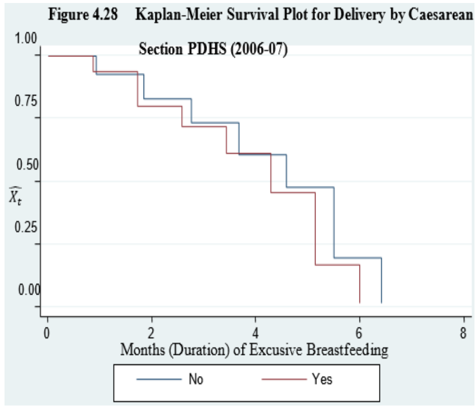


Figure 4.28 Kaplan-Meier Survival Plot for Delivery by Caesarean Section PDHS (2006-07)

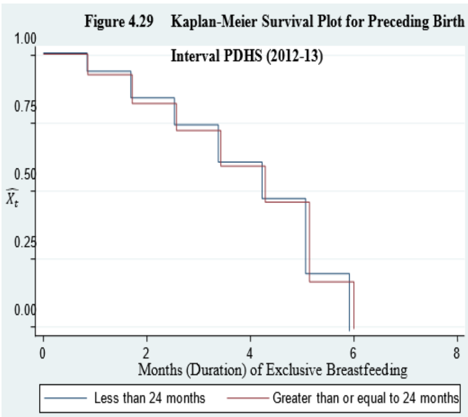


Figure 4.29 Kaplan-Meier Survival Plot for Preceding Birth Interval PDHS (2012-13)

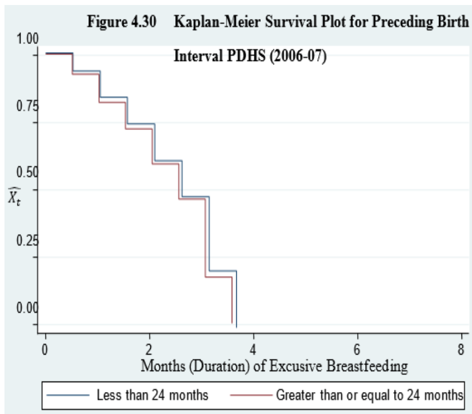


Figure 4.30 Kaplan-Meier Survival Plot for Preceding Birth Interval PDHS (2006-07)

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 347-352

COOPERATIVE LEARNING: EFFECTS ON MATHEMATICS STUDENTS' ACHIEVEMENT IN PRIVATE SCHOOLS

Muhammad Shahid Zulfiqar[1] and **Syeda Tehzeeb Alvi**[2]
[1] University of Education, Lahore, Pakistan.
Email: shahidzac@yahoo.com
[2] Lahore College for Woman University, Lahore, Pakistan.
Email: tehzeebarshad@yahoo.com

ABSTRACT

Cooperative Learning strategy is being used at all grade levels and in all subject areas across the world but its potential as an alternative to traditional teaching strategy has not been fully explored in most schools in Pakistan. Thus, this experimental study was conducted to explore the effects of Cooperative Learning on academic achievements of 9th grade maths students studying in private schools of Lahore. Further, the students' academic achievement was compared on the basis of high and low level of achievement in mathematics. The students (50 students) were randomly assigned to both (x25 to each) experimental and control group. The instrument to measure the students' achievement in math consisted of maths achievement tests which have 30 items. The results showed that the students taught by the cooperative learning showed higher achievement in math than those students taught by the traditional method. As it is proved through the results that cooperative learning enhances the students' academic achievement in mathematics. So it should be deployed while teaching mathematics so the students' achievement in math could be maximized.

KEY WORDS

Cooperative Learning, academic achievement, private schools.

INTRODUCTION

Education as an important element for moral, social, political, cultural and economic development of a nation. Those nations which have taken bold initiatives towards education have revolutionized themselves; they have introduced new innovations and perform miracles in recent days. Responding to global challenges Govt. of Pakistan has adopted vision 2030" as the planning commission which states that Pakistani education system should be able to provide quality education to the children and youth of Pakistan and it may enable them to be real potentials and contributors to the development of nation (Govt. of Pakistan, 2009).

An educational system is explicitly based on the quest 'what to teach and how to teach? 'What to teach? 'Means the learning material. The continuum of learning material swings from linguistic to scientific knowledge. The choice of contents and subject matter from the multifarious branches of knowledge is subjected to societal needs. However mathematics as branch of knowledge has been inevitable ingredient of core curriculum at basic stages in every society of ancient times. Mathematics is still a subject that is considered difficult and boring to many students. According to Woodard (2004), weaker students feel anxiety towards mathematics and this anxiety affects their performance in mathematics. The quality of education that teachers provide to students is dependent upon what teacher do in their classroom (Zakaria & Iksan, 2007)

Research has proved that constructivism helps the teacher to be successful in their classrooms (Casas, 2006). Cooperative Learning refers to a family of instructional strategies/models in which students work together to enhance the learning of each other. Cooperative learning was designed and implemented in order to develop social strategies and social attitudes in students, and to improve social relations within and between groups. In addition there is a large cluster of cooperative learning models aimed at cognitive development e.g. in mathematics. Sometimes cooperative learning is directed at both the social and the cognitive side of human development (Gillies, 2008)

Cooperative learning is a teaching arrangement that refers to small groups of students working together to achieve a common goal (Kagan, 1994). The basic element are positive interdependence, individual accountability, equal participation, simultaneous interaction. Cooperative learning is the ways of organizing the learning environment of a classroom. In cooperative learning environment, the goals of separate individuals become so linked that there is a positive correlation between them.

According to Johnson (1989) in cooperative learning, students tend to enjoy mathematics and this enjoyment motivates them to learn. Melihan and Sirri (2011) concludes that the cooperative learning method is more effective than the traditional teaching method in the academic success of students. Woolfolk (2004) has explained the Cooperative Learning "as a system of learning in which students are in a mix abilities group and they are rewarded as a group". Researchers have argued about the superiority and the effectiveness of cooperative learning over traditional method. In today's educational setting, cooperative learning is urgently needed to promote social relationship among students.

Robertson et al., (1999) reported cooperative learning as visible and effective instructional technology for teaching and learning mathematics. It helps to make mathematics exciting and enjoyable for both students and teachers. Cooperative strategies can be integrated at any grade level and for any mathematics topic. Moreover many positive effects are noted by teachers and students. Students learn to cooperate with others and to communicate in the language of mathematics. The classroom atmosphere tends to be relaxed and informal, help is readily available, questions are freely asked and answered, and even the shy students find it easy to be involved. Student tends to become friends with their group members and teachers-students relationship tends to be more relaxed. In addition many students maintain a high level of interest in the mathematics activities and have an opportunity to pursue the more challenging aspects of mathematics, while they achieve, at least, as much information and skills as in more traditional approach.

Although cooperative learning method have become a major research field for past three decades (Stane, 2000), yet a very few studies have been carried out to investigate the use of cooperative learning in science classes (Colossi, 1998).Research on Cooperative Learning at school level has generally been in English, Humanities and

Social Sciences. Furthermore only a few studies have been conducted in South Asia. Studies conducted in Pakistan are even very few.

STAD (Student Team Achievement Division) model of cooperative learning was selected by the researchers due to its wide ranging and diversified outcomes. It promotes interaction among students towards science and increase social interpersonal skills (Balkfakih, 2003).

STAD has not been investigated so far in Pakistan as an alternative method for teaching maths at high school level. The present study designed to explore the effects of Cooperative Learning on academic achievement of 9^{th} grade maths students.

STATEMENT OF THE PROBLEM

Cooperative Learning strategy is being used at all grade levels and in all subject areas across the world but its potential as an alternative to traditional teaching strategy has not been fully explored in most schools in Pakistan. Thus an experimental study was conducted to explore the effects of Cooperative Learning on students' maths academic achievement.

SIGNIFICANCE OF THE STUDY

The effects of Cooperative learning on maths students' achievement was investigated in this study. This will be helpful for the educators to know that to what extent cooperative learning stands to prove effective in teaching the maths and attaining maximum students' achievement.

HYPOTHESIS OF THE STUDY

The study aimed to explore

1. There is no significant difference between the mean achievement scores of maths students taught through cooperative learning and those taught by traditional method of teaching.
2. There is no significant difference between the mean scores of high achievers and low achievers of the control and experimental groups on post-test.

METHODOLOGY

The researchers deployed the Quasi-experimental design. The sample for this study included 50 students of 9^{th} class taken randomly from two sections of The Educators School, Gulshan-e-Ravi Campus Lahore. These students were further divided into two groups (experimental group & control group) each of which comprised of 25 students. Treatment group were exposed to cooperative learning while the control group was taught by the traditional method. The researchers used the mathematics achievement tests developed by themselves in order to measure the students' maths achievement. The instrument (maths achievement tests) were validated through the expert opinion whereas checked for the reliability, which was 0.916 taken through Chronbac's alpha after piloting the test on 15 students of 9^{th} class. The instrument consists of 30 items. The data were collected from all the students included in the study.

Table 1
Pre-and-Post –Test Equivalent Group Design

Group	Pre-Test	Treatment	Post-Test
A (n=25)	O1	X	O2
B (n=25)	O1	C	O2

The above mentioned table shows the quasi-experimental design of this study. A represents the experimental group while B represents the control group. The pre-test is represented by O1 whereas the post test is represented by O2 for the experimental and control group respectively. The cooperative learning treatment is represented as x whereas the traditional treatment is represented by C.

ANALYSIS AND RESULTS

The data were obtained in the form of raw scores of pre and post-tests. The data were analysed by applying the *t*-test to explore the significant difference between the means scores of both the experimental and control groups on the variable of pre-test and post-test scores. On the other hand, to examine the treatment effects of high and low levels of achievement of the groups, factorial design (2*2 Analysis of variance) was applied. For this purpose, students of both groups were divided into two halves, i.e. high achievers (above the means score) and low achievers (below the means score).This division was made on the basis of pre-test results.

Table 2
Independent Sample t –test Comparing the Group Performance on Pre-test

	Groups	N	μ	σ	*t- value*	Sig
	Experimental Group	25	21.88	206		
Pre Test					.264	.793
	Control Group	25	21.72	2.20		

The results of Table 2 shows that there was no significant difference found (t= .264 sig= .793) among the achievement of experimental group (Mean= 21.88, SD= 2.06) and control group (Mean= 21.72, SD= 2.20) score in the pre-test.

Table 3
Independent Sample t –test Comparing the Group Performance on Post-test

	Groups	N	μ	σ	*t- value*	Sig
	Experimental Group	25	32.68	5.22		
Pre Test					9.04	0.00
	Control Group	25	22.20	2.50		

The results of Table 3 indicates that the significant difference was found between the mean scores of experimental group and control group (t=9.04, sig=0.00). The aforementioned results indicate experimental group performance (Mean=32.68, SD=5.22) in math is better than control group (Mean=22.20, SD=2.50).

Table 4
Control Group in Post-Test on Applying ANOVA

	Source of Variation	Df	Sum of Square	Mean Square	F
	Treatment	1	1873.48	1873.48	5.96
Post Test					
	Level				

Significant at 0.05 level

It is revealed that the teaching methods used were efficient and the experiment conducted falls under the high level of significance. Teaching method (treatment), achievement level and the interaction between the factors remained significant at 0.05 level of probability. Results shows that the high achievers and low achievers of the experimental group performed significantly better than high and low achievers of the control group. Figure 1 shows the gap difference narrowing towards the high achievers, which revealed that the performance of experimental group was significantly better than control group for low achievers as compared to high achievers. Thus cooperative learning approach promises to be more effective for low achievers. Previous research support this findings.

DISCUSSIONS

The present study supported the findings of (Anderson, 2003), (Zakaria, Chin, & Daud, 2010), (Hossain & Tarmizi, 2013) and (Remillard, 2015) that co-operative learning enhances the students' academic achievement in mathematics. So it should be deployed while teaching mathematics so the students' achievement in math could be maximized.

CONCLUSIONS

It is concluded through the results of this study that the students taught by the cooperative learning showed higher performance/ achievement in math than those students taught by the traditional method. So, it is proved through the results that cooperative learning enhances the students' academic achievement in mathematics. As cooperative learning brings many advantages like maximum class involvement, good educational results, so it should be deployed while teaching mathematics so the students' achievement in math could be maximize.

REFERENCES

1. Anderson, S.A. (2003). *The effect of cooperative learning on the mathematical achievement of fifth grade students.* (Unpublished Thesis of Master of Science in teaching) Rowan University.
2. Casas, M. (2006). Implementing constructivist web-based learning and determining its effectiveness on a teacher preparation course. The Journal of Educators Online, 3(2), 1-17.
3. Emmer, E.T. and Gerwels, M.C. (2002). Cooperative learning in elementary classrooms: Teaching practices and lesson characteristics. *The Elementary School Journal*, 103(1), 75-91.
4. Gillies, R., Ashman, A. and Terwel, J. (2008). The Teacher's Role in Implementing Cooperative Learning in the Classroom. *Computer Supported Collaborative Learning Series*. Vol. 8, 2008, New York: Springer.
5. Hossain, A. and Tarmizi, R.A. (2013). Effects of cooperative learning on students' achievement and attitudes in secondary mathematics. *Procedia - Social and Behavioral Sciences*, 93, 473-477.
6. Johnson, D.W. and Johnson, R. (1989). *Cooperation and competition: Theory and research*. Edina, MN: Interaction Book Co.
7. Kagan, S. (1994). *Cooperative Learning.* San Clemente, CA: Resources for Teachers, Inc.
8. Melihan, U. and Sirri, A. (2011). The effect of cooperative learning method on the students' success and recall levels of the 8th grade students learning in permutation and probability subject. *Journal of Kirsehir Education Faculty*, 12, 1-16.
9. Remillard, H.A. (2015). *The effect of cooperative learning on middle school math students.* (Unpublished Master's Thesis) Heritage University.
10. Government of Pakistan (2009). *Vision 2030*, Planning Commission of Pakistan, Islamabad.
11. Woodard, T. (2004). The Effects of Math Anxiety on Post-Secondary Developmental Students as Related to Achievement, Gender, and Age. Inquiry, 9(1), 1-5.
12. Woolfolk, A. (2004). *Educational Psychology*, Prentice Hall.
13. Zakaria, E., Chin, L.C. and Daud, M.Y. (2010). The effects of cooperative learning on students' mathematics achievement and attitude towards mathematics. *Journal of Social Sciences*, 6(2), 272-275.
14. Zakaria, E. and Iksan, Z. (2007). Promoting cooperative learning in science and mathematics education: A Malaysia perspective. *Eurasia Journal of Mathematics, Science & Technology Education*, 3, 35-39.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 353-364

IMPACT OF INTERNATIONAL GEOPOLITICS ON THE SUCCESS OF C – PEC WITH INTERNAL ISSUES OF PAKISTAN ACTING AS MEDIATOR

Muhammad Rashid Salaria[1] and **Zahid Saeed Sheikh**[2]
Shaheed Zulfiqar Ali Bhutto Institute of Sciences and Technology (SZABIST),
Islamabad, Pakistan.
Email: [1]rashid@szabist-isb.edu.pk
[2]callxahid@gmail.com

ABSTRACT

China – Pak Economic Corridor (C-PEC) encourages a shift from classical geopolitics to critical geopolitics where in focus is on rational regionalism and strategic alternatives. This study examines the general perception / opinion of university students from the disciplines of international relations (IR), social sciences (SS), & development studies (DS), personnel of Pakistani print, electronic & social media, with regard to international geopolitics creating problems in Pakistan in terms of terrorism, political instability and cross-border intrigues to impede the success of C-PEC. This study takes into special account the mediatory role of internal issues of Pakistan between the relationship of international geopolitics and the success of C-PEC. The study is based on a representative sample of 160 plus individuals. The data obtained through online questionnaire has been analyzed with the help of statistical tests of correlation and regression. The results reveal that international geopolitics is related directly and indirectly with the progress of C-PEC and explains significant amount of variability in the dependent variable. Said differently, international geopolitics has full potential of affecting the success of C-PEC directly as well as indirectly via its impact on internal problems/issues of Pakistan. Thus, partial mediatory role of internal problems/issues of Pakistan in the relationship of international geopolitics with the success of C-PEC is established in this study. Areas of further research have been suggested in the study.

1. INTRODUCTION

On the paradigm shift in the international relations, one of the Greek philosophers had once said, '*One cannot step into the same river water twice'.* This rightly signifies that nothing is permanent except change in nation to nation relations.

International relations have undergone drastic changes over time. Foes have become friends, and friends have turned out to be foes. Countries that once went on wars against each other are today joining unions together. Furthermore, the erstwhile super powers have retreated to the background while the new emerging powers which were once considered third world nations are taking the reins of world order. This phenomenon has given birth to new form of international geopolitics on global scale.

Today's truth in the context of international relations is that Pakistan and China have become closest allies. This is primarily because both the countries have common mutual interests as well as good collaborative skills to deal with each other. Pakistan's geographic location is highly strategic to China as it helps China to connect with the Middle East and Central Asia directly. On one hand, China is likely to suffer tremendously if Pakistan plans to exclusively join the US camp, and, on the other hand, Pakistan sees itself at a loss if China decides to strengthen its ties with India keeping in mind the high trade flow between India and China. As such, all-weather friendship of Pakistan and China is strategically very important for both the countries in the current scenario.

The mega initiative and project of China – Pak Economic Corridor (C-PEC) is about laying down the foundation of China's idea of One Belt-One Road' (OBOR), creating nexus of Asia with Europe, Central Asia, Middle East and Africa. The C-PEC is a real game changer which carries new hopes, novel welfares strategies and much desired harmony in the local, regional and international geo-politics.

China-Pakistan Economic Corridor (C-PEC) is a landmark strategic opening of the world trade corridor in the 21st century which connects the European and African markets with the rising power centers of China and Asia via an alternate land route from Gwadar to Kashgar. This corridor prompts a shift from classical geopolitics to critical geopolitics with the great objective of laying down the foundation of rational regionalism and searching for strategic alternatives. China is the rising world power of the future and 21st century is the century of Asia. The peaceful rise of China, with its global outreach, is merely drawing on one dictum i.e. to search for alternatives and not to put all eggs into a single basket. Mankind has been utilizing geographical settings since ancient times in its greater interest. Lands on the banks of rivers turned into epicenters of civilizations.

Needless to reemphasize, the C-PEC is a high-significance initiative as it goes through the most geostrategic areas in South Asia. It serves as an exchange connect between China, Middle East, Central Asia and Europe through Pakistan and is full of potential to create multiple socio-economic activities with a large number of new jobs absorbing skilled and non-skilled labor in various occupations and offering huge opportunities of earning income and eliminating poverty, thus raising standards of living of people not only in Pakistan and China but also in other nations of the region. There are, however, a few challenges posed by certain local and worldwide partners that require to be tactfully managed and amicably sorted-out for achievement of big objectives.

2. OBJECTIVES OF THE STUDY

This study aims to find out answers to the stated-below questions by analyzing statistically the impact of current international geopolitics on the success of C-PEC and the effects of the latter on various strata of Pakistani society and on different domains of economic activity including commerce and industry in Pakistan. The specific objectives of the study are:

- To determine the general perception / opinion of public whether or not the current international geopolitics is instrumental in hindering the success of C-PEC directly as well as through igniting internal problems in Pakistan.
- To determine as to what extent international geopolitics is likely to aggravate the internal problems in Pakistan.
- To determine the extent to which internal problems of Pakistan are likely to impact the success of C-PEC.

3. RESEARCH QUESTIONS

The research questions were structured to check relationship of variables covered in this research. One issue pertained to checking of the strength of relationship of independent variable (international geopolitics) with the dependent variable (success of C-PEC) given the likelihood of the mediatory variable (internal problems of Pakistan) to affect C-PEC in its own right. Following questions were formulated for testing of various hypotheses in the study:

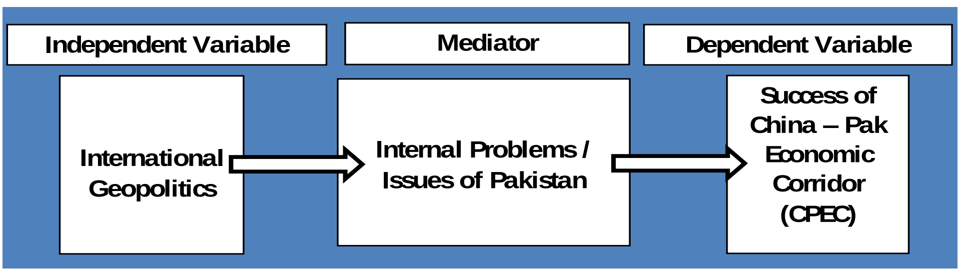
- How the international geopolitics will affect the success of C-PEC overtly or covertly or both ways?
- How the intrigue of international geopolitics will aggravate Pakistan's internal issues/problems?
- How Pakistan's internal issues/problems are potentially dangerous for the success of C-PEC?

4. SIGNIFICANCE OF THE STUDY

In-depth study of the impact of international geopolitics on the success of China-Pak Economic Corridor (C-PEC) constitutes the rationale of the present research. There is no denying the fact that the global powers and neighboring countries are and have been intriguing to make the C-PEC process unsuccessful due to the assumed hegemony of China in the region and outrunning US in economic as well as in strategic fronts. Specially, India's stance has persistently been negative in this process as it is expected to enable Pakistan to supersede her traditional rival on the one hand while on the other, China, India's arch rival, is all set to cut short the Indian size, putting it in the category of sub regional power from the status of a regional power. Such elements are potential threats to the success of C-PEC. The results of the study are thus of great value for national strategic planners, policy makers, development researchers and practitioners.

5. THEORETICAL / CONCEPTUAL FRAMEWORK

Based on the above-cited theory and extensive literature review, the following conceptual framework is developed for the study:



6. HYPOTHESES

It can be assumed that international geopolitics has a direct and indirect impact on the success of C-PEC. Furthermore, internal problems of Pakistan can also potentially impede the prospects of C-PEC. The amicable solutions of all thorny issues of international and domestic nature provide a proper environment to save the mega project from all adverse effects. It is, thus, hypothesized that:

H_1: International geopolitics contributes positively to impede the success of C-PEC.

H_2: International geopolitics contributes positively to promote and aggravate internal issues/problems of Pakistan.

H_3: Internal issues / problems of Pakistan contribute positively to impede the success of C-PEC.

7. METHODOLOGY

This is a quantitative and descriptive study aimed at describing the relationship of variables under study in depth. In order to assess the impact of international geopolitics on Pakistan with regard to C-PEC and its successful completion and operations afterwards, it is attempted to predict the mediation caused by internal issues to hinder the process C-PEC. The target population for the study consists of university students of International Relations (IR), Social Sciences (SS), & Development Studies (DS) and Journalists of Pakistani Print, Electronic & Social Media. Individuals are thus the units of analysis for the present study. Non-probability convenience sampling technique was used to collect the data. Data was collected through a structured questionnaire floated online, followed by phone calls and personal visits. Two hundred (200) questionnaires were floated and one hundred sixty one (161) were received back from the respondents in complete form. The response rate, thus, was 80.5%.

One dependent variable (DV), one independent variable (IV) and one mediating variable (MV) were measured on 0 – 5 point ratio scale (0 “No Agreement” and 5 “Complete Agreement”). Software used to analyze data was SPSS (version 24.0). Correlation and Regression tests were applied to measure the relationship between the variables under study. Cronbach Alpha test was run to determine the internal consistency and reliability of the questionnaire. The content and face validity of the questionnaire were assessed by the experts of the relevant areas.

8. DATA ANALYSIS AND RESULTS

Reliability Statistics: Table 1 summarizes and portrays the internal consistency / reliability statistics of items of questionnaire. Independent variable (International Geopolitics) had seven (7) items; mediating variable (Internal Issues / Problems of Pakistan) had five (5) and dependent variable (Success of C-PEC) had six (6) items. The reliability of the items was assessed separately, analyzing the consistency within the items of each variable. Reliability evaluation through Alpha Reliability of IV (International Geopolitics) was 0.666 for 7 items, showing a consistency within 7 items of the IV. The Alpha Reliability of MV was 0.752 for6 items, showing no element of inconsistency among the items. DV consisting of 5 items was evaluated through

Cronbach Alpha Reliability Test and it was revealed that consistency within the items of DV was0.723, showing consistency of items as quite good.

Table 1
Reliability Statistics

Variables	Cronbach Alpha (α)	No. of Items
International Geopolitics (IV)	0.666	7
Internal Issues of Pakistan (MV)	0.752	6
Success of C-PEC (DV)	0.723	5

Correlation Analysis: Table 2 reveals the nature of the relationship between the independent variable i.e. International Geopolitics and the Success of C-PEC as dependent variable. This relationship is positive and has high statistical significance. Likewise, internal problems of Pakistan are significantly related with International Geopolitics and with the success of C-PEC.

This dimension of the research is quite interesting. As indicated in the table, the relationship of International Geopolitics and Internal Problems of Pakistan (MV) is positive and statistically significant at significance level p = 0.01. Similarly, the correlation coefficient of Internal Problems of Pakistan and the success C-PEC is positive and statistically significant at the significance level of p = 0.01. This means that the impact of international geopolitics is quite high on the internal problems of Pakistan and the success of C-PEC is conditional to an effective mechanism which needs to be put in place to address the influence of international geopolitics and internal issues of Pakistan in order to make the mega project of C-PEC a thorough success.

Table 2
Correlation Matrix

		International Geopolitics (IV)	Success of C-PEC (DV)	Internal Issues of Pakistan (MV)
International Geopolitics (IV)	Pearson Correlation	1		
Success of C-PEC (DV)	Pearson Correlation	.716**	1	
Internal Issues of Pakistan (MV)	Pearson Correlation	.677**	.738**	1

**** Correlation is significant at the 0.01 level (2-tailed).**

Regression Analysis: Regression tests were carried-out to measure the variations in success of C-PEC (DV) as explained by the variation in international geopolitics (IV). Mediating role of internal issues of Pakistan (MV) in relationship between international geopolitics and the success of C-PEC was also measured. Regression results of each variable are analyzed and interpreted as follows:

Table 3 indicates that the variation in the success of China – Pak Economic Corridor (C-PEC) is explained to the extent of 51% (R^2 = 0.512) through variation in international geopolitics. The remaining 49% variation is unexplained. This shows the overall strength of international geopolitics as a predictor of the success of C-PEC.

Step 1: International Geopolitics versus Success of China - Pak Economic Corridor (C-PEC) (DV on IV)

Table 3
Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.716[a]	0.512	0.509	0.533	0.512	167.831	1	160	0.000
a. Predictors: (Constant), Mean_GP1									

ANOVA establishes the fitness of the regression model with F = 167.831 which is significant at level 0.000. This shows the model fits very well for the prediction of dependent variable. In other words the independent variable, international geopolitics is a significant predictor of dependent variable, the success of China – Pak Economic Corridor (C-PEC).

Table 3.1
ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	47.705	1	47.705	167.831	.000[b]
	Residual	45.479	160	0.284		
	Total	93.183	161			
a. Dependent Variable: Mean_SCPEC						
b. Predictors: (Constant), Mean_GP1						

In Table 3.2, the marginal effect of international geopolitics (IV) on the success of C-PEC (DV) is studied. International geopolitics is highly statistically significant (t = 12.955, p = 0.000) which means that ceteris paribus, one unit increase in international geopolitics (IV) will impede the success of C-PEC (DV) by .757 units. In other words, as it was hypothesized, the international geopolitics can be a great hurdle in the way of the success of China – Pak Economic Corridor.The results testify this hypothesis.

Table 3.2
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	99.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	0.969	0.227		4.277	0.000	0.378	1.559
	Mean_GP1	0.757	0.058	0.716	12.955	0.000	0.605	0.909

a. Dependent Variable: Mean_SCPEC

Table 4 indicates that the variation in internal issues/problems of Pakistan is explained to the extent of 46% (R^2= 0.458) through variation in international geopolitics. The remaining 54% variation is unexplained. This shows the strength and potential of international geopolitics as a predictor of aggravation in the internal issues/problems of Pakistan.

Step 2: International Geopolitics versus Internal Issues / Problems of Pakistan (MV on IV)

Table 4
Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.677[a]	0.458	0.454	0.608	0.458	135.137	1	160	0.000

a. Predictors: (Constant), Mean_GP1

ANOVA establishes the fitness of the regression model with F = 135.137 which is significant at level 0.000. This shows the validity and predictive position of the independent variable in the model. In other words, international geopolitics as an independent variable is a significant predictor of aggravation in the internal issues/problems of Pakistan.

Table 4.1
ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	**Regression**	49.976	1	49.976	135.137	.000[b]
	Residual	59.171	160	0.370		
	Total	109.147	161			

a. Dependent Variable: Mean_Int_Issues
b. Predictors: (Constant), Mean_GP1

In Table 4.2 the marginal effect of international geopolitics (IV) on internal issues/problems of Pakistan (MV) is studied. International geopolitics is highly statistically significant (t = 11.625, p = 0.000) which means that ceteris paribus, one unit increase in international geopolitics (IV) will aggravate the internal problems of Pakistan (MV) by .775 units. It was hypothesized that international geopolitics contributes positively to aggravate the internal issues/problems of Pakistan, which is verified by the results of this table.

Table 4.2
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	99.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	0.812	0.258		3.144	0.002	0.139	1.486
	Mean_GP1	0.775	0.067	0.677	11.625	0.000	0.601	0.949

a. Dependent Variable: Mean_Int_Issues

Table 5 reveals that the variation in the success of China – Pak Economic Corridor (C-PEC) is explained to the extent of 55% (R^2= 0.545) through variation in internal issues of Pakistan. The remaining 45% variation is unexplained. This shows the potential of internal issues of Pakistan as a predictor of impediment in the success of China – Pak Economic Corridor (C-PEC). It was hypothesized that internal issues /problems of Pakistan are positive contributors to impede the success of C-PEC, in terms of security threats, subversive activities, disruption in economic moves, etc. This hypothesis is also testified by the results.

Step 3: Internal Issues / Problems of Pakistan versus Success of China - Pak Economic Corridor (C-PEC) (MV on DV)

Table 5
Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.738[a]	0.545	0.542	0.515	0.545	191.738	1	160	0.000

a. Predictors: (Constant), Mean_Int_Issues

ANOVA in Table 5.1 establishes the fitness of the regression model with F = 191.738 which is significant at level 0.000. This shows the validity and predictive position of the independent variable in the model. In other words, the factor of internal issues/problems of Pakistan, treated here as an independent variable, is a significant predictor of dependent variable, the success of China – Pak Economic Corridor (C-PEC).

Table 5.1
ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	**Regression**	50.796	1	50.796	191.738	.000[b]
	Residual	42.388	160	0.265		
	Total	93.183	161			

a. Dependent Variable: Mean_SCPEC
b. Predictors: (Constant), Mean_Int_Issues

In Table 5.2, the marginal effect of the internal issues of Pakistan (MV) on the success of China – Pak Economic Corridor (C-PEC) (DV) is examined. Internal issues/problems of Pakistan are highly statistically significant (t = 13.847, p = 0.000) which means that ceteris paribus, one unit increase in internal issues of Pakistan (MV) will impede the success of C-PEC (DV) by 0.682 units. The hypothesis under test here was that the internal/issues/problems of Pakistan adversely affect the success of C-PEC which is supported by the results.

Table 5.2
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	99.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	1.285	0.190		6.770	0.000	0.790	1.780
	Mean_Int_Issues	0.682	0.049	0.738	13.847	0.000	0.554	0.811

a. Dependent Variable: Mean_SCPEC

Table 6 reveals that the variation in the success of China – Pak Economic Corridor (C-PEC) is explained to the extent of 63% (R^2 – coefficient of determinant = 0.631) through variation in international geopolitics and in internal issues/problems of Pakistan. The remaining 37% variation is unexplained. This shows the join strength of international geopolitics and internal issues/problems of Pakistan as predictors of the success of China – Pak Economic Corridor (C-PEC).

Step 4: International Geopolitics and Internal Issues of Pakistan versus Success of China - Pak Economic Corridor (C-PEC) (IV & MV on DV)

Table 6
Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.794[a]	0.631	0.626	0.465	0.631	136.008	2	159	0.000

a. Predictors: (Constant), Mean_Int_Issues, Mean_GP1

ANOVA in Table 6.1 establishes the fitness of the regression model with F = 136.008 which is significant at level 0.000. This shows the validity and predictive position of the independent variables in the model. In simple terms, the hostile international geopolitics and its concomitant effect on internal issues/problems of Pakistan are the significant predictors of variation in the dependent variable, the success of China – Pak Economic Corridor (C-PEC).

Table 6.1
ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	58.808	2	29.404	136.008	.000[b]
	Residual	34.375	159	0.216		
	Total	93.183	161			

a. Dependent Variable: Mean_SCPEC
b. Predictors: (Constant), Mean_Int_Issues, Mean_GP1

Table 6.2 gives the partial effects of international geopolitics (IV) and the internal issues/problems of Pakistan (MV) on the success of C-PEC (DV). These effects are highly statistically significant ($t = 6.088$, $p = 0.000$), ($t = 7.167$, $p = 0.000$) respectively, This means that ceteris paribus, one unit increase in international geopolitics may impede the success of C-PEC by .421 units provided there is no change in internal issues/problems of Pakistan at the same time. On the other hand, one unit increase in internal issues/problems of Pakistan will impede the success of C-PEC by .443 units if status quo prevails in the arena of international geopolitics. As indicated in the table, international geopolitics is a potent hurdle in the way of the success of C-PEC and its whole process directly as well as indirectly. The indirect impact of international geopolitics is in terms of its role in creating and aggravating internal issues/problems of Pakistan and exploiting them to the extent possible. These results establish the hypothesis that the international geopolitics is a great impediment to the success of C-PEC, both directly as well as indirectly in the existing format of things in Pakistan.

Table 6.2
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	99.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	0.617	0.204		3.031	0.003	0.086	1.148
	Mean_GP1	0.421	0.069	0.398	6.088	0.000	0.241	0.602
	Mean_Int_Issues	0.433	0.060	0.469	7.167	0.000	0.276	0.591

a. Dependent Variable: Mean_SCPEC

9. COMMENTS & CONCLUSIONS

The present conceptual research attempts to contribute new knowledge to the existing literature on China – Pak Economic Corridor (C-PEC). It is a quantitative study comprising of in-depth literature review and empirical evidences with primary data obtained through online survey. The idea was to test the hypotheses that there exists a highly strong nexus between international geopolitics and the grand initiative of C-PEC as the changing international geopolitics in the region and beyond is likely to adversely affect the success of C-PEC. Furthermore, the mediatory role of internal issues / problems of Pakistan between the relationship of international geopolitics and the success of C-PEC was to be tested in all clarity. The study establishes beyond doubt the likelihood of a strong adverse effect of international geopolitics on the successful operation of C-PEC provided no remedial mechanism exists on the ground. The study also reveals the partial mediating role of internal issues / problems of Pakistan in this regard. The results further reveal that the indirect impact of international geopolitics on the success of C-PEC will be more serious and dangerous as compared to its direct impact. The reason is that China, for her own reasons, may use her clout as future world player to offset to an extent the effects of international geopolitics on the success of C-PEC. But to cope with the thorny internal issues / problems of Pakistan, especially in Baluchistan and other parts, and to sort out their permanent solutions amicably, will be a great challenge for the state. Given the enthusiasm of Pakistani public and full determination of the relevant stakeholders to see the C-PEC a real success, everything is likely to go smoothly regarding this mega project.

In the future, we recommend research for incorporating other important factors such as socio-economic development of stakeholder countries, international relations, geostrategic aspects and environmental issues in order to establish the authenticity of the relationships analyzed in this study. The present study will not only enrich the existing literature on the subject, but it will also practically help researchers, practitioners, development sector personnel and government officials of the concerned countries to reinforce their determination to reap positive and useful results out of this landmark initiative. Undoubtedly, the grand project of C-PEC is going to be a big step forward to further strengthen the friendship of Pakistan and China, leading to huge mutual benefits for both the countries and their citizens.

10. ACKNOWLEDGEMENT

All praises are for the Almighty Allah (SWT) Who has granted us the ability to perform this task and finish it to the best of our abilities. We are extremely thankful to all the people who provided data and thus participated in this study so actively. We are also very thankful to our families who were always there to support us. Last but not the least; we are thankful to our colleagues and friends for their help and support.

REFERENCES

1. Khan, A.U. (2014). Pak-China economic corridor: The hopes and reality. *Regional Studies*, 33(1), 45-63.
2. Ambassador Zukang, Sha (2015), *China-Pakistan Bilateral Relations, Seminar at the NUST Global Think Tank Network,* October 19, 2015.
3. Gupta, A. and Shrivastav, S. (2014). Evolution of US rebalancing strategy: Implications for India. In Muni, S.D. and Chadha, V. (Ed.) *Asian Strategic Review*, 5-28.
4. Bannerji, D. (2010). Addressing Nuclear Dangers: Confidence Building between India-China-Pakistan. *India Review,* 9(3), 345-363.
5. Baron, R.M. and Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173.
6. BBC (2015). *China's Xi Jinping agrees $46bn superhighway to Pakistan*, www.bbc. com/news/world-asia-323770888.

7. Bhattacharjee, D. (2015). *China Pakistan Economic Corridor* Available at SSRN: https://ssrn.com/abstract=2608927 or http://dx.doi.org/10.2139/ssrn.2608927
8. Bhaumik, P. (2014). Assessing the Barriers to Trade in Services in South Asia. *Global Business Review*, 15(4), 795-814.
9. Bijan, Z. (2005). China's Peaceful rise to great-power status. *Foreign Aff.*, 84, 18.
10. Bilal, Muhammad (2014). *Pak-China Economic Corridor: Geo-Strategic Importance of Pakistan*; PRIME Institute (Policy Research Institute of Market Economy).
11. Buskirk, T. (2016). *Target Population and Sampling Frame in Survey Sampling,* http://www.theanalysisfactor.com/target-population-sampling-frame/
12. Campbell, M.J., Machin, D. and Walters, S.J. (2007). *Medical Statistics: A Textbook for the Health Sciences* (4th ed.). Chichester, UK: Wiley.
13. Centre for Pakistan and Gulf Studies (CPGS) (2014). *Foreign Policy Roundtable Series, Pakistan – China Relations: Prospects & Challenges*, Roundtable Report V.
14. Chaturvedy, R.R. and Malone, D.M. (2009). *India and its South Asian Neighbours*, ISAS Working Paper No. 100. Institute of South Asian Studies, Singapore.
15. Chhetri, P.S. (2015). Political and Economic Dynamics of CPEC within the Larger Framework of China's One Belt, One Road Initiative, *Proceedings of International Conference on CPEC* held at GC University, Lahore.
16. Ebrahim, Z.T. (2015). China‘s new silk road: What‘s in it for Pakistan? daily *Dawn*, *http://www.dawn.com/news/1177116*
17. Chinadaily.com.cn (2014). *China's Xi Proposes Security Concept for Asia*, Xinhua, 05/21/2014.
18. Clinton, H. (2011). America‘s Pacific Century, *Foreign Policy*, 56-63.
19. Domínguez, G. (2015). Author Interview with Andrew Small: *"Bilateral relations: The Sino-Pakistani axis: Asia's 'little understood' relationship"*, Deutsche Welle. http://www.dw.de/the-sino-pakistani-axis-asias-little-understood-relationship/a-18194448
20. Elahi, M.M. (2015). Heartland and Rimland Theories in CPEC Perspective: Strategic Interplay in 21st Century; *Proceedings of International Conference on CPEC* held at GC University, Lahore.
21. Downs, E. (2012). China buys into Afghanistan. *SAIS Review of International Affairs*, 32(2), 65-84.
22. Farooq, Y. (2014). Is the New silky Road really an Economic Corridor for China and Pakistan? *The Express Tribune,*
23. Fazil, M.D. (2016). Pakistan: What Stands in CPEC's Way? *The Diplomat*. http://thediplomat.com/2016/02/pakistan-what-stands-in-cpecs-way/

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 365-372

ANTECEDENTS OF PATIENTS' SATISFACTION IN HEALTH CARE CENTERS: A CASE STUDY OF ISLAMABAD AND RAWALPINDI

Muhammad Rashid Salaria[1] and **Maryam Shakeel**[2]
Shaheed Zulfiqar Ali Bhutto Institute of Sciences and Technology (SZABIST), Islamabad, Pakistan.
Email: [1]rashid@szabist-isb.edu.pk
[2]maryam.shakeel@hotmail.com

ABSTRACT

Health care sector of any country is very important, Good quality of health care services gives hope and provides relief to the people of the state. Well managed and good quality health care services help in maintaining and sustaining a healthy human capital that plays a big role in the development of the country. The objective of this research study is to analyze whether the patients in Rawalpindi and Islamabad are satisfied with the health care services they are receiving from the hospitals. For this purpose SERVQUAL instrument is used to measure patients' satisfaction about the quality of health care facilities provided in the public and private hospitals in Islamabad and Rawalpindi. Because of the nature of the study, only those respondents were included in the study, who had some experience of the health care facilities provided in the hospitals. A sample of 180 individuals was selected for the research study. Data was analyzed with the help of correlation and regression techniques. Results show that, if empathy, responsiveness and timeliness are present in the hospitals, patients are quite well satisfied. Number of Doctors does not matter to the patients, as a factor if the other three elements are present. Results further show that responsiveness is the key element of patient satisfaction followed by empathy and timeliness.

KEYWORDS

Hospitals, Health Care, Service Quality, Patient Satisfaction.

1. INTRODUCTION

Health care sector of any country whether a country is developed or is developing, requires particular attention by the government. This is because the health care facilities are very important for the people of the country, which help in maintaining a required healthy human capital.

Health institutions are supposed to work as a system to provide the required health care services to the people. These services need to be safe, high quality dependable and accessible to all. (Malik, 2013). Three types of hospitals are working in Pakistan, private, public and Combined Military Hospitals (CMH). Although there are sufficient numbers of hospitals in the country but still the quality of health care services provided in these hospitals is not up to the mark and patients have to face lot of issues regarding health care facilities. Many research studies have been done on service quality and these research studies indicate that service quality is very important for the success and reputation of service providers including hospitals.

In the present era, a tough competition, both inter and intra-organizational, and progress in the technology are also playing important role for organizations to keep their service quality up to the mark and keep on making new improvements successful. Successful organizations and field units working under them play an extremely important role in the improvement and the development of the economy of the country. This is because the more productive the organizations are there are more chances of economic development and social welfare of the state. Pakistan is a developing country and in this country health care sector is generally over looked and is not given proper fundings, as compared to other sectors. So in this scenario, only a few health care institutions are providing really good and efficient service quality to the patients. The service quality provided and the satisfaction of the patients play very important role for the survival of health care organizations, as these two factors are key issues for development planning and strategic decision making of the health care organizations. Patients' perception of a hospital not only affects its profitability but also affects the reputation of the hospital. The rising population pressure and the socio-political needs of the masses are also demanding a changed and innovative health care system in the country.

2. PURPOSE AND OBJECTIVE

The objective of this study is to examine, the relationship between the patients' satisfaction and the facilities provided in hospitals of Rawalpindi and Islamabad, in terms of a good number of doctors, empathy, responsiveness and timeliness. Health sector is very essential for any country, so it is really important to satisfy the patients. Hospitals whether they are public or private, they have to provide the best possible services to the patients to satisfy them completely. It is, therefore very important to analyze the factors that contribute to Patients' Satisfaction, so that the health centers work efficiently and work hard to give best health care facilities to the patients.

3. CONCEPTUAL FRAMEWORK

This study examines the effects of the factors like number of doctors, empathy, responsiveness and timeliness as independent variables on Patients' Satisfaction which is taken as a dependent variable. The following is the conceptual framework of this research work.

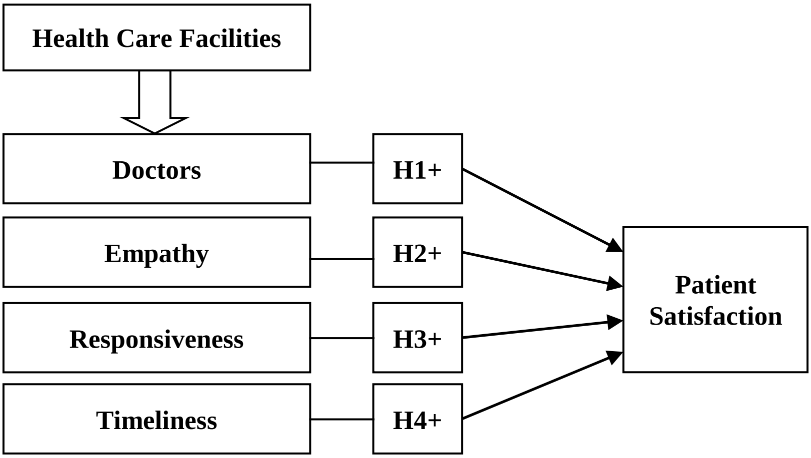


Figure 1: Antecedents of Patients' Satisfaction in Health Care Centers

Based on the above conceptual and theoretical framework, following hypotheses are formulated and tested in this research work.

4. HYPOTHESIS

H1: Number of Doctors has a positive Impact on Patients' Satisfaction.
H2: Empathy has a positive impact on Patients' Satisfaction.
H3: Responsiveness has a positive impact on Patients' Satisfaction.
H4: Timeliness has a positive impact on Patients' Satisfaction.

5. METHODOLOGY

This study is quantitative in nature and aims to analyze the antecedents of Patient Satisfaction in health care Centers in Islamabad and Rawalpindi. Exploratory Research Design has been utilized in the research, to explore the role and contribution of factors which are assumed to be necessary for satisfaction of patients in the public and private hospitals of Islamabad and Rawalpindi. Convenience sampling technique was used for the collection of data for the study. Data was collected through questionnaires that were distributed to 200 respondents, out of these 180 (90%) filled in and useable questionnaires were received back, which were considered enough for the study. The questionnaire was based on the five-point Likert-scale, comprising of four variables to measure Patient satisfaction. The questionnaire focused on the number of doctors available, the concern that doctors showed towards patients and the concern of the hospital staff and nurses towards patients. Questionnaire also focused on whether, everything was done timely and whether the patients feel comfortable or not, and whether the Patients were satisfied with the overall health care attention provided by the hospitals. The questionnaire was structured and close ended which was used for this quantitative research.

Content Validity refers to the extent to which a measure represents all facets of a given social construct. It measures the significance of the questions in the questionnaire with respect to its link with the area under study. The questionnaire for the Survey was discussed with the professionals and experts who found it as suitable for examining the health care facilities provided by the Public and Private hospitals, and whether Patients were satisfied with the services. Furthermore a small-scale pre-test was conducted before the actual study. The results of the pilot study were positive, which is an indication that the questionnaire measured properly what it was supposed to measure. It clearly represented the content and study population. Cronbach's alpha was used to assess instrument's internal consistency.

6. DATA ANALYSIS AND FINDINGS

Required descriptive statistics were examined for a proper data analysis. Pearson Correlation coefficients were calculated and examined in depth to see the relationship between study variables. As for demographic data, out of total respondents, 50% were male and 50% female. About 38% of the respondents were in the age bracket of between 26 to 35 years, 22% were less than 25 years, between 36-45 years were also 22% and 18% were above 45 years old.

A correlation matrix (table below) was prepared and used to investigate inter-relations among the four antecedents of patients' satisfaction and their direct relationship with the dependent variable. It was noted that multicollinearity was not a problem when patients' satisfaction factors were entered in regression model. The multiple regression indicated that these four patients' satisfaction antecedents could explain the variation in the response variable quite well. Suitability of the data for multiple regression was assessed by investigating the relationship between the independent variables and the overall patients' satisfaction, coupled with thorough study of residuals.

Correlation Matrix

	Doctors	Empathy	Responsiveness	Timeliness	Patient Satisfaction
Doctors	1	.418*	.450**	.430**	.426**
Empathy	.418**	1	.965**	.948**	.951**
Responsiveness	.450**	.956**	1	.960**	.956
Timeliness	.430**	.948**	.960**	1	.940**
Patient Satisfaction	.426**	.951**	.956**	.940**	1

** Correlation is Significant at the 0.01 level (2-tailed).

Regression Analysis
Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.963[a]	.928	.926	.44577

a. Predictors: (Constant), Timeliness, Doctors, Empathy, Responsiveness.

ANOVA

Model	Sum of Squares	Df	Mean Square	F	Sig.
Regression	447.613	4	111.903	563.137	.000[b]
Residual	34.775	175	.199		
Total	482.388	179			

a. Dependent Variable: Patient Satisfaction
b. Predictors: (Constant), Timeliness, Doctors, Empathy, Responsiveness

ANOVA test established the validity and fitness of the regression model which provided a strong reason to the researchers to believe that at least one of the four distinct patients' satisfaction factors, examined simultaneously, were contributing significant predictability for the overall patients' satisfaction in health care centers. Results (F=563.137, p=.000) shows that this model fits very well for the prediction of response variable.

Regression model summary provided the value of R square as 0.93. R square = 0.93 indicates that the fitted regression model with the above said predictors explains 93% of the variation in dependent variable that is patient satisfaction.

Coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
(Constant)	-.412	.168		-2.447	.015
Doctors	.007	.053	.003	.135	.893
Empathy	.400	.087	.371	4.583	.000
Responsiveness	.472	.103	.425	4.563	.000
Timeliness	.181	.078	.178	2.336	.021

a. Dependent Variable: Patient Satisfaction
b. Predictors: (Constant), Timeliness, Doctors, Empathy, Responsiveness

Studying antecedents of Patients' Satisfaction individually, the results show that all independent variables (except Doctors) are significantly related with the Dependent Variable that is Patient Satisfaction. The table shows that one unit increase in "Empathy" enhances the patients' satisfaction by .40 units, when other variables are held constant. This coefficient is statistically significant. It was hypothesized in the study that Empathy has a positive impact on patients' satisfaction. The results established this hypothesis. The table further shows that, one unit increase in "Responsiveness" enhances the patients' satisfaction by .47 units, when there is no change in other independent variables. This coefficient is statistically significant. It was hypothesized in the study that responsiveness has a positive impact on patients' satisfaction. The results established this hypothesis also. Likewise, as table shows, one unit increase in "Timeliness" enhances the patients' satisfaction by .18 units, when other variables are held constant. This coefficient is statistically significant. The hypothesis under test was that Timeliness has a positive impact on patients' satisfaction which was fully established by the results. The results of the study, however, do not support the hypothesis that number of doctors has a positive impact on Patients' Satisfaction, Although the regression coefficient of variable "Doctors" is positive, but it is not statistically significant when other variables remain in the model simultaneously. It appears that the patients main concern is not number of doctors itself, but they feel more satisfied when factors like empathy, responsiveness and timeliness are there in adequate amount and quality in the hospitals.

Consequently, as per the results of the study, the position of the hypothesis tested is as follows:

H1: Number of medical doctors has a positive impact on Patients' Satisfaction... Rejected

H2: Empathy has a positive impact on Patients' Satisfaction… Accepted

H3: Responsiveness has a positive impact on Patients' Satisfaction… Accepted

H4: Timeliness impacts on Patients' Satisfaction positively… Accepted

7. CONCLUSION

Patients' satisfaction is essential for a proper cure. It reflects on the quality of hospital management and provision of health care facilities to a patient during the medical treatment. Although many things matter in a good treatment but certain elements are must. These include proper and standardized medical practices and existence of qualified doctors and paramedical staff. A high quality health care facility is a dire need of Pakistan, a nation with a low rank recorded in Human development Index. (Human Development Index, 2015.)

The issue of patient satisfaction and service quality is getting a lot of attention in health sector strategic planning and development. Opinion of the Patients' about the services provided by the hospitals or clinics not only affects the image of the hospitals but also the profitability of the hospitals. The study examines in depth the role of the number of doctors in a hospital along with the standards of empathy, responsiveness and timeliness prevailing there.

The results of this research show that all variables studied, except number of doctors, are significantly related to the dependent variable, that is patients’ satisfaction. This shows that patients are not concerned with the number of doctors as such as long as the other factors are available in the hospitals. Providing good health care is very important and it is one such profession in which high quality knowledge and practices are required to satisfy the patients and for the overall well-being of the masses. A developing nation like Pakistan, confronting the monster of massive poverty and having most minimal position in worldwide HDI positioning, needs a great deal of endeavors for building up a strong healthcare framework both in the public and private health care centers. This study is expected to give proper understanding of the issue and can be useful for the government attention to improve the existing quality of the hospitals and to meet the healthcare needs of the people in a befitting manner.

REFERENCES

1. Andaleeb, S.S. (2000). Public and private hospitals in Bangladesh: service quality and predictors of hospital choice. *Health Policy and Planning*, 15(1), 95-102.
2. Andaleeb, S.S. (2001). Service quality perceptions and patient satisfaction: a study of hospitals in a developing country. *Social Science & Medicine*, 52(9), 1359-1370.
3. Asubonteng, P., McCleary, K.J. and Swan, J.E. (1996). SERVQUAL revisited: a critical review of service quality. *Journal of Services Marketing*, 10(6), 62-81.
4. Azam, M., Rahman, Z., Talib, F. and Singh, K.J. (2012). A critical study of quality parameters in health care establishment: developing an integrated quality model. *International Journal of Health Care Quality Assurance*, 25(5), 387-402.
5. Babakus, E. and Mangold, W.G. (1992). Adapting the SERVQUAL scale to hospital services: an empirical investigation. *Health Services Research*, 26(6), 767-780.
6. Donabedian, A. (1980). *The Definition of Quality and Approaches to Its Assessment. Explorations in Quality Assessment and Monitoring*, Vol. 1, Health Administration Press, Ann Arbor.
7. Kudsk, K.A., Reddy, S.K., Sacks, G.S. and Lai, H.C. (2003). Joint Commission for Accreditation of Health Care Organizations guidelines: too late to intervene for nutritionally at-risk surgical patients. *Journal of Parenteral and Enteral Nutrition*, 27(4), 288-290.
8. Lamiraud, K., Booysen, F. and Scheil-Adlung, X. (2005). The impact of social health protection on access to health care, health expenditure and impoverishment: A case study of South Africa. *Global Campaign on Social Security and Coverage for All, Geneva, International Labour Offices*.
9. Larsson, G. and Wilde-Larsson, B. (2010). Quality of care and patient satisfaction: A new theoretical and methodological approach. *International Journal of Health Care Quality Assurance*, 23(2), 228-247.
10. Levitt, T. (1981). Marketing intangible products and product intangibles. *Cornell Hotel and Restaurant Administration Quarterly*, 22(2), 37-44.
11. McKinley, R.K. and Roberts, C. (2001). Patient satisfaction with out of hours primary medical care. *BMJ Quality & Safety*, 10(1), 23-28.
12. Mostafa, M.M. (2005). An empirical study of patients' expectations and satisfactions in Egyptian hospitals. *International Journal of Health Care Quality Assurance*, 18(7), 516-532.

13. Moulin, H. (2004). *Fair division and collective welfare*. MIT Press.
14. Padma, P., Rajendran, C. and Sai, L.P. (2009). A conceptual framework of service quality in healthcare: perspectives of Indian patients and their attendants. *Benchmarking: An International Journal*, 16(2), 157-191.
15. Parasuraman, A., Zeithaml, V.A. and Berry, L.L. (1988). Servqual: A multiple-item scale for measuring consumer perception of service quality. *Journal of Retailing*, 64(1), 12-40.
16. World Health Organization, (. (2010). *Western Pacific Regional Strategy for Health Systems Based on the Values of Primary Health Care: Universal Coverage for Better Health Outcomes*. WHO Library Cataloguing in Publication Data.
17. Wu, C.C. (2011). The impact of hospital brand image on service quality, patient satisfaction and loyalty. *African Journal of Business Management,* 5(12), 4873-4882.
18. Yoo, D.K. and Park, J.A. (2007). Perceived service quality: Analyzing relationships among employees, customers, and financial performance. *International Journal of Quality & Reliability Management*, 24(9), 908-926.
19. Younger, S.D. (1999). The relative progressivity of social services in Ecuador. *Public Finance Review*, 27(3), 310-352.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 373-384

RISK FACTORS OF CARDIAC DISEASES IN ISLAMABAD CAPITAL TERRITORY (ICT)

Nagina Zulfiqar[1], **Sidra Razzaq**[2] and **Samina Satti**[3]
Department of Statistics, University of Wah, Wah Cantt, Pakistan
Email: [1]nagina.zulfiqar@uow.edu.pk
[2]sidra.razzaq@uow.edu.pk
[3]samina.satti@uow.edu.pk

ABSTRACT

Cardiovascular disease (CVD) is responsible for more number of deaths worldwide. The muscles and vessels of heart and blood transporting roads become vulnerable portion in most of the CVD. The role of hypertension and cholesterols of different density triglycerides in induction and progression of cardiovascular disease is discussed in this present review.

There have been several clinical studies examining the factors associated with cardiovascular disease (CVD) in patients. IHT, AMI and MVCAD are among the most common causes of death and disability in the world. The Pakistan (including India, Bangladesh, Sri Lanka, and Nepal) has among the highest rates of cardiovascular disease (CVD) globally. In this review, we discuss the existing data on the prevalence of CVD and its risk factors in the Islamabad Capital Territory (ICT) Pakistan.

Description about the Variables under Study

Variable Name	Description
Gender	Gender of patients; male/female.
Age	Age of the patients.
Weight	Weight of the patients in lb.
HTN	Hypertension
DM	Diabetes mellitus.
Systolic BP	Upper blood pressure level of patients.
Diastolic BP	Lower blood pressure level of patients.
Heart rate	Heart beat measures of the patients/pulse rate.
Sats	Saturation %
Creatinine	Level of creatinine mg/gl
Potassium serum	Blood potassium level mEq/L
Sodium serum	Blood sodium level mEq/L
CKMB	Creatine Kinase MB (CK-MB) level in ng/ml
Troponin	Level of cardiac-specific troponin in blood.
Calcium	Calcium level in bones Mg/dl
High Density Lipoprotein	High proportion of protein in blood plasma.
ECG	ECG of heart.
Cholesterol	Blood cholesterol level mg/dL
Diagnoses	IHD: ischemic heart disease; AMI: Acute myocardial infarction; MVCAD: Multi-vessel coronary artery disease

Abbreviations:

- HTN: Hypertension
- CVD: Cardiovascular Disease
- DM: Diabetes Mellitus
- CHD: Coronary Heart Disease
- IHD: Ischemic Heart Disease
- AMI: Acute Myocardial Infarction
- MVCAD: Multi-Vessel Coronary Artery Disease

1. INTRODUCTION

The concept of risk factors is straightforward; however, the term has been frequently misinterpreted by the general public and media. We define risk factors as specific pathophysiological mechanisms or clinical entities underlying diseases that promote increased morbidity and mortality. The term risk factors was first introduced by the team of investigators of the Framingham Heart Study who identified specific ‘‘factors of risk’’

Cardiovascular Disease (CVD) is the most common cause of death in the Asia and includes coronary heart disease (angina/heart attack), stroke (where normal blood supply to part of the brain is cut off, damaging the area affected); mini stroke (known as transient ischemic attack or TIA and peripheral arterial disease (narrowing of arteries usually in the legs). Non modifiable and modifiable risk factors can increase the probability of developing CVD.

CVD includes a wide range of disorders which includes diseases of the cardiac muscle and of the vascular systems. Potential risk factors for CVD include hypertension, tobacco use, physical inactivity, elevated low-density lipoprotein cholesterol, diabetes, overweight/obesity and a cluster of interrelated metabolic risk factors. Framingham Heart Study (FHS) links the presence of high cholesterol, tobacco usage, hypertension and diabetes mellitus to future CVD.

Cardiovascular disease (CVD) is not only a serious threat to human health, but also an important contributor to the total costs of medical care worldwide. According to the statistics, 30% of people die from CVD. As a major cause of mortality, CVD is increasing at an alarming rate. Pakistani population has one of the highest risks of coronary heart disease (CHD) in the world. In Pakistan, 30 to 40 percent of all deaths are due to cardiovascular diseases (CVD).

2. LITERATURE REVIEW

Cardiovascular disease can be defined as the development of pathology that occurs in the vascular system. CVD is associated with one or more characteristics of an individual that increases the likelihood of developing a disease (Kramer, Newton, & Sivarnjan Froelicher, 2008). CV risk factors are associated with an increased risk of developing cardiovascular disease. In general, CV risk factors include demographic characteristics, family history of CVD, physical inactivity, lipids and lipoprotein, obesity, hypertension and diabetes (Kramer and at all, 2008). Knowledge describes a familiarity, awareness or understanding of facts, theoretically acquired by a person (Cavell, 2002).

Pakistan is facing a dual burden of both communicable and non-communicable diseases. The 2013 global burden of disease report predicted that the 30% of the worldwide deaths are related to CVD. A few of the estimates about the common illness among Pakistani adult population includes 41% hypertension, 21% tobacco use, 17.3% high cholesterol, 21% obesity, 10% diabetes mellitus (DM), and dyslipidemia (males, 34%; females, 49%), and 2.8% stroke. These estimates are rising in the country, and the rate of NCDs and communicable diseases is almost equal. In Pakistan, population data for heart disease is limited; two population studies were conducted in 1965 and 1973 that showed that the prevalence of heart disease is between 0% and 3.7% in rural and urban areas. Data from the 1994 National Health Survey of Pakistan (NHSP) on health problems shows a high incidence of risk factors for CVD in both rural and urban populations. Other small-scale studies also showed high prevalence of CVD risk factors in Pakistan, however, these studies have limitations in sample size and the self-reported data-collection methods. A cross-sectional study conducted at a hospital of Karachi identified the CVD-related risk factors, including family history of ischemic heart disease (IHD), age, body mass index (BMI), smoking, sedentary lifestyle, total cholesterol, DM, low-density lipoprotein (LDL), high-density lipoprotein (HDL), and triglycerides. The relationship between smoking and CVD is unequivocal; it is the major health risk in today's world. According to the 2015 World Health Organization (WHO) report, 22.2% of men and 2.1% of women smoke in Pakistan. It is generally noted that an individual's SES and illiteracy are directly associated with risk behaviors and an unhealthy lifestyle. Thus, resource constraint led people to use money on alcohol and smoking rather than on education and healthy food, which cause unfavorable effects on cardiovascular system. (Rubina Barolia, Amber Hussain Sayani (Aga Khan University School of Nursing and Midwifery Karachi, Pakistan), Journal of Pakistan Medical Association (JPMA), November, 2017.)

3. OBJECTIVE OF THE STUDY

The major objective of the study is to compare risk factors for cardiovascular disease and to examine the prevalence, incidence, predisposing factors for hypertension, hazards as an ingredient of the CVD risk profile, and implication of this information for prevention and treatment.

Defining Variables:

In present study CVD Diagnoses **(**IHD: ischemic heart disease, AMI: Acute myocardial infarction, MVCAD: Multi-vessel coronary artery disease) is dependent variable and all the remaining variables are taken as independent.

Hypothesis:

There is insignificant relation between independent variables (Gender, Age, Weight, HTN, DM, Systolic BP, Diastolic BP, Heart rate, Sats, Creatinine, Potassium serum, Sodium serum, CKMB, Troponin, Calcium, High density lipoprotein, ECG, Cholesterol) and the dependent variable cardiovascular disease.

4. RESEARCH METHODOLOGY

The present Heart Study is a community-based study which was designed to enquire into and provide information about the risk factors of cardiovascular disease in early life. We implemented a retrospective cohort survey at Shifa International Hospital Ltd., with a total sample size of 250. In this study no sampling technique was adopted. The statistical analysis will be performed in SPSS Version 20.0. BP, cholesterol, weight variables are taking as continues variables which are set into interval scale; while HTN, DM and diagnoses are set as categorical variables, so that we can get more precise results. Since our dependent variable (diagnoses) is categorical therefore we apply “Logistic Regression Analysis” on our obtained data set.

5. RESULTS

Multinomial logistic regression is used to model nominal outcome variables, in which the log odds of the outcomes are modelled as a linear combination of the predictor variables.

5.1. Nominal Regression model

5.1.1. Case Processing Summary

		N	Marginal Percentage
CV Diagnosed Disease	IHD	155	62.0%
	AMI	56	22.4%
	MVCAD	39	15.6%
Age of the Patients (Binned)	<= 36	67	26.8%
	37 – 52	58	23.2%
	53 – 64	73	29.2%
	65+	52	20.8%
ECG of the Heart	Yes	240	96.0%
	No	10	4.0%
Diabetes Mellitus	Yes	222	88.8%
	No	28	11.2%
Hypertension	Yes	197	78.8%
	No	53	21.2%
Gender of the Patients	Male	161	64.4%
	Female	89	35.6%
Valid		250	100.0%
Missing		0	
Total		250	
Subpopulation		250[a]	

a. The dependent variable has only one value observed in 250 (100.0%) subpopulations.

The case processing summary Table **5.1.1**, shows the total number of observations in each category of categorical variable with respective marginal percentages. The dependent variable CVD diagnosed disease, shows that IHD have more prevalence as compare to AMI & MVCAD.

5.1.2. Model Fitting Information

Model	**Model Fitting Criteria**	**Likelihood Ratio Tests**		
	-2 Log Likelihood	**Chi-Square**	**df**	**Sig.**
Intercept Only	460.671			
Final	429.727	30.944	40	.847

The model fitting information Table **5.1.2**; have chi-square value 30.944 and the significance value 0.847 under the likelihood ratio tests column. -2 log likelihood is 460.671 when we have intercept only in the model and 429.727 with parameters in it. Thus the model with parameters is better fit model than the model without parameters.

5.1.3. Pseudo R-Square

Cox and Snell	.116
Nagelkerke	.138
McFadden	.067

In Pseudo R-square Table **5.1.3**, there are three pseudo R-squared values. The values of R-square shows the proportion of variance of the CVD diagnoses approximately 11% to 13% explained by Gender, Age, Weight, HTN, DM, Systolic BP, Diastolic BP, Heart rate, Sats, Creatinine, Potassium serum, Sodium serum, CKMB, Troponin, Calcium, High density lipoprotein, ECG and Cholesterol.

5.1.4. Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	429.727[a]	.000	0	.
Weight	429.983	.256	2	.880
SistolicBP	431.365	1.637	2	.441
DiastolicBP	430.506	.779	2	.677
HeartRate	430.784	1.057	2	.590
Sats	431.000	1.273	2	.529
Creatinine	430.896	1.169	2	.557
PotassiumSerum	430.031	.304	2	.859
SodiumSerum	429.940	.213	2	.899
CKMB	430.861	1.134	2	.567
Troponin	432.932	3.205	2	.201
Calcium	430.689	.961	2	.618
High Density Lipoprotein	432.234	2.507	2	.285
Cholestrol	436.345	6.617	2	.037
Newage	434.391	4.663	6	.588
ECG	431.580	1.853	2	.396
DM	430.287	.560	2	.756
HTN	431.103	1.376	2	.503
Gender	430.194	.466	2	.792

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

In Likelihood ratio tests Table **5.1.4**; in the sig. column the values are as 0.880, 0.441, 0.677, 0.590, 0.529, 0.557, 0.589, 0.899, 0.567, 0.201, 0.618, 0.285, 0.037, 0.588, 0.396, 0.756, 0.503 and 0.792 respectively. As all the sig. values are greater than p-value 0.05, except the sig. value of cholesterol i.e. 0.037, thus the CVD has a significant overall association with cholesterol.

5.1.5. Parameter Estimates

CV Diagnosed Disease[a]		B	Std. Error	Wald	df	Sig.	Exp (B)	95% Confidence Interval for Exp (B)	
								Lower Bound	Upper Bound
IHD	Intercept	-.149	4.968	.001	1	.976			
	Weight	.000	.005	.007	1	.931	1.000	.991	1.010
	Sistolic BP	-.003	.014	.055	1	.815	.997	.970	1.024
	Diastolic BP	.003	.017	.030	1	.862	1.003	.969	1.038
	Heart Rate	.007	.007	1.030	1	.310	1.007	.993	1.022
	Sats	.018	.023	.648	1	.421	1.018	.974	1.065
	Creatinine	.290	.319	.829	1	.363	1.337	.716	2.497
	Potassium Serum	-.009	.059	.024	1	.878	.991	.883	1.113
	Sodium Serum	.007	.018	.172	1	.678	1.007	.973	1.043
	CKMB	.027	.035	.580	1	.446	1.027	.958	1.101
	Troponin	.042	.053	.603	1	.438	1.042	.939	1.158
	Calcium	.048	.111	.189	1	.664	1.049	.844	1.305
	High Density Lipoprotein	.009	.008	1.157	1	.282	1.009	.993	1.026
	Cholestrol	-.009	.004	4.681	1	.030	.991	.983	.999
	[Newage=1]	-.667	.551	1.468	1	.226	.513	.174	1.510
	[Newage=2]	.424	.634	.448	1	.503	1.529	.441	5.295
	[Newage=3]	.058	.570	.010	1	.920	1.059	.346	3.240
	[Newage=4]	0[b]	.	.	0	.	.	.	.
	[ECG=1]	.469	.932	.254	1	.614	1.599	.257	9.930
	[ECG=2]	0[b]	.	.	0	.	.	.	.
	[DM=1]	-.426	.686	.386	1	.534	.653	.170	2.507
	[DM=2]	0[b]	.	.	0	.	.	.	.
	[HTN=1]	-.614	.547	1.260	1	.262	.541	.185	1.581
	[HTN=2]	0[b]	.	.	0	.	.	.	.
	[Gender=1]	-.093	.403	.054	1	.817	.911	.413	2.008
	[Gender=2]	0[b]	.	.	0	.	.	.	.

CV Diagnosed Disease[a]		B	Std. Error	Wald	df	Sig.	Exp (B)	95% Confidence Interval for Exp (B)	
								Lower Bound	Upper Bound
AMI	Intercept	-1.982	6.282	.100	1	.752			
	Weight	.002	.006	.178	1	.673	1.002	.991	1.014
	Sistolic BP	-.017	.016	1.131	1	.288	.983	.952	1.015
	Diastolic BP	-.010	.020	.227	1	.634	.990	.952	1.030
	Heart Rate	.006	.008	.599	1	.439	1.006	.990	1.023
	Sats	.041	.041	1.007	1	.316	1.042	.962	1.128
	Creatinine	.376	.364	1.069	1	.301	1.456	.714	2.969
	Potassium Serum	.017	.058	.088	1	.767	1.017	.909	1.139
	Sodium Serum	.009	.020	.191	1	.662	1.009	.970	1.048
	CKMB	.032	.037	.765	1	.382	1.033	.961	1.110
	Troponin	-.031	.069	.205	1	.650	.969	.846	1.110
	Calcium	.117	.127	.855	1	.355	1.124	.877	1.442
	High Density Lipoprotein	.004	.010	.162	1	.688	1.004	.985	1.023
	Cholestrol	-.011	.005	5.513	1	.019	.989	.980	.998
	[Newage=1]	-.288	.649	.197	1	.657	.750	.210	2.675
	[Newage=2]	.485	.736	.434	1	.510	1.625	.384	6.878
	[Newage=3]	.099	.671	.022	1	.883	1.104	.296	4.112
	[Newage=4]	0[b]	.	.	0	.	.	.	.
	[ECG=1]	1.634	1.345	1.476	1	.224	5.127	.367	71.612
	[ECG=2]	0[b]	.	.	0	.	.	.	.
	[DM=1]	-.166	.786	.045	1	.833	.847	.182	3.952
	[DM=2]	0[b]	.	.	0	.	.	.	.
	[HTN=1]	-.476	.616	.596	1	.440	.621	.186	2.080
	[HTN=2]	0[b]	.	.	0	.	.	.	.
	[Gender=1]	.150	.478	.099	1	.753	1.162	.455	2.966
	[Gender=2]	0[b]	.	.	0	.	.	.	.

a. The reference category is: MVCAD.
b. This parameter is set to zero because it is redundant.

In above Parameter estimates Table **5.1.5.**; the MVCAD is taken as reference category. B- are the values for the logistic regression equation predicting from the CVD from Gender, Age, Weight, HTN, DM, Systolic BP, Diastolic BP, Heart rate, Sats, Creatinine, Potassium serum, Sodium serum, CKMB, Troponin, Calcium, High density lipoprotein, ECG, Cholesterol. They are in log-odds units. Similar to OLS regression, the logistic prediction equation is:

$$\text{Pij}=e^{\sum_{j=1}^{k}\alpha+\beta kj*Xkji}/\sum_{j=1}^{k} e^{\sum_{j=1}^{k}\alpha+\beta*Xkji}$$

where i=cases, j=categories, and k is the number of explanatory variables i.e.; HTN, DM, weight/obesity, BP (systolic, diastolic) and cholesterol.

IHD relative to MVCAD:

The values under sig. column are 0.976, 0.931, 0.815, 0.862, 0.310, 0.421, 0.363, 0.878, 0.678, 0.446, 0.438, 0.664, 0.282, 0.030, 0.226, 0.503, 0.920, 0.614, 0.534, 0.262 and 0.817 respectively. The significance value shows that only cholesterol has significant effect on IHD.

AMI relative to MVCAD:

The values under sig. column are 0.752, 0.673, 0.288, 0.634, 0.439, 0.316, 0.301, 0.767, 0.662, 0.382, 0.650, 0.355, 0.688, 0.019, 0.657, 0.510, 0.883, 0.224, 0.833, 0.440 and 0.753 respectively. The significance value shows that only cholesterol has significant effect on AMI.

5.2 After Illumination of In-Significant Factors Reduced Nominal Regression Model

5.2.1. Case Processing Summary

		N	Marginal Percentage
CV Diagnosed Disease	**IHD**	155	62.0%
	AMI	56	22.4%
	MVCAD	39	15.6%
Valid		250	100.0%
Missing		0	
Total		250	
Subpopulation		38[a]	

a. The dependent variable has only one value observed in 7 (18.4%) subpopulations.

After illumination of in-significant variables the case processing summary Table **5.2.1** shows the total number of observations in each category of categorical variable with respective marginal percentages. The table shows the same results as shown with significant factors i.e.; the dependent variable CVD diagnosed disease, shows that IHD have more prevalence as compare to AMI & MVCAD.

5.2.2. Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	Df	Sig.
Intercept Only	180.233			
Final	175.009	5.224	2	.073

After the illumination of insignificant values the reduced model only remains with intercept and explanatory variable cholesterol. The model fitting information Table **5.2.2** shows that chi-square value is 5.224 and the significance value 0.073 under the likelihood ratio tests column. -2 log likelihood is 180.233 when we have intercept only in the model and 175.009 with parameters in it. Thus the model with parameters is better fit model than the model without parameters.

5.2.3. Pseudo R-Square

Cox and Snell	.021
Nagelkerke	.025
McFadden	.011

In Pseudo R-square Table **5.2.3**; there are three pseudo R-squared values for reduced model. The values of R-square shows the proportion of variance of the CVD diagnoses approximately 1.1% to 2.5% explained by the cholesterol.

5.2.4. Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	187.449	12.439	2	.002
Cholestrol	180.233	5.224	2	.073
The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.				

The above likelihood ratio test Table **5.2.4** for reduced model shows the same effect as full model.

5.2.5. Parameter Estimates

CV diagnosed disease[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B)	
								Lower Bound	Upper Bound
IHD	Intercept	3.444	1.085	10.067	1	.002			
	Cholestrol	-.008	.004	3.880	1	.049	.992	.984	1.000
AMI	Intercept	2.797	1.198	5.447	1	.020			
	Cholestrol	-.009	.004	4.367	1	.037	.991	.982	.999
a. The reference category is: MVCAD.									

The parameter estimates Table **5.2.5** also shows that the cholesterol has significant effect on IHD and AMI.

6. CONCLUSION

The overall study shows that the risk factor cholesterol is highly significant in both diagnoses (IHD, AMI relative to MVCAD) rather than other factors. Thus we conclude that the cholesterol has more influence on CVD than other factors.

7. SUGGESTIONS/RECOMMENDATIONS

Policy intervention programs should be implemented to promote healthy dietary patterns and physical activity in school and work place. Culture norms should be identified that might hinder CVD prevention efforts. Research infrastructure and training of clinical researchers should be developed. High-quality national surveillance programs should be developed that document causes of death and disability to guide policy and decision makers. Assessment of the knowledge of the CVD burden in Islamabad among lay people, health care workers, and policy makers. Reliable statistics should be generated on prevalence and incidence of CVD and its risk factors and CVD-related mortality through the initiation of large cohort studies and trial registries. Studies should be documenting child-hood risk factors, such as obesity, in urban and rural settings. Large registries and case control studies to document the burden of stroke and its determinants in Islamabad. Continuing efforts to address non-atherosclerotic manifestations of CVD, including rheumatic heart disease and tuberculosis pericarditis.

8. ACKNOWLEDGEMENT

The authors wish to thank anonymous referees for their careful reading and constructive suggestions which leads to the improvement over an earlier version of the paper. Thanks to the Department of Statistics, University of Wah, Wah Cantt, Pakistan for their logistic and financial support for this research.

REFERENCES

1. Burger, A., Pretorius, R., Fourie, C.M. and Schutte, A.E. (2016). The relationship between cardiovascular risk factors and knowledge of cardiovascular disease in African men in the North-West Province. *Health SA Gesondheid*, 21(1), 364-371.
2. Barolia, R. and Sayani, A.H. (2017). Risk factors of *cardiovascular disease and* its *recommendations* in Pakistani context. *The Journal of the Pakistan Medical Association*, 67(11), 1723-1729.
3. Mahmooda, S.S., Levy, D., Vasan, R.S. and Wang, T.J. (2014). The Framingham Heart Study and the epidemiology of cardiovascular diseases: A historical perspective. *The Lancet*, 383(9921), 999-1008.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 385-390

CHRONIC KIDNEY DISEASE: A STATISTICAL ANALYSIS

Fizza Yaqub, Mansoor Shoukat Khan, Amna Nazeer[§]
and **Javeria Nawaz Abbasi**
COMSATS University, Islamabad, Pakistan
[§]Corresponding author Email: nazeeraamna@yahoo.com

ABSTRACT

Chronic kidney disease (CKD) is a life-threatening condition. It presents with many associated co-morbidities like diabetes mellitus, hypertension, muscular and joint diseases, CVD and other chronic conditions. Fibroblast growth factor-23 (FGF23), a protein, is secreted by the osteocytes and osteoblasts while responding to high levels of phosphorous in serum and also due to high levels of active form of vitamin D. Serum FGF23 has been proposed to be the initial adaptive response in early CKD patients for the protection of the organism from adverse effects of phosphate retention. In this article we aimed at assessing the levels of FGF23 in CKD patients of both gender with 18 years of age and above who were on vitamin D supplementation with those patients without vitamin D supplementation. Both genders are equally on risk of getting CKD and no single group is found on significant threat. Based on the findings of multiple linear regression with dependent variable as levels of FGF23, we conclude that age has negative impact on levels of FGF23. we also have noticed that the FGF23 levels were significantly higher in CKD patients who were on vitamin D supplementation than those without vitamin D supplementation.

1. INTRODUCTION

The kidneys are the bean-shaped organs located just below the ribs cage. The kidneys are powerful chemical factories that executes life sustaining jobs e.g. remove waste products and excess fluid from the body, balance the body's fluids, produce an active form of vitamin D that promotes strong and healthy bones, release hormones that regulate blood pressure and control the production of red blood cells, etc. Kidneys maintain a healthy balance of water, salts, and minerals in the blood. Without this balance, nerves, muscles, and other tissues in the body may not work normally.

The term "chronic kidney disease (CKD)" means lasting damage to the kidneys that can get worse over time. CKD is a condition characterized by a gradual loss of kidney function over time usually for three months or longer. CKD is mostly caused by other health problems that have done permanent damage to the kidneys over time. Diabetes and high blood pressure are among the most common causes of CKD. Chronic kidney disease usually gets worse slowly and symptoms may not appear until kidneys are badly damaged. CKD is categorized into five stages least harming to severe damage. Kidney failure is the last stage of CKD that is relieved either through dialysis or kidney transplant.

The eGFR (estimated glomerular filtration rate) is a blood test that measures how well the kidneys filter waste from the blood. The stages of kidney disease are based on the eGFR number. Fibroblast growth factor 23 or FGF23 is a protein and a bone-derived hormone that is responsible for phosphate and vitamin D metabolism through a novel bone-kidney axis. Serum FGF23 level was found to increase once estimated glomerular filtration rate (eGFR) is lower than 90 mL/min/1.73m2,16 and was found to be associated with an increase of CVD, left ventricular hypertrophy, and mortality rate in both non-dialysis and dialysis CKD patients. FGF23 is thought to be an early biomarker of disordered phosphorus metabolism in the initial stages of chronic kidney disease (Isakova et al. 2011).

Vitamin D deficiency is highly prevalent among patients with chronic kidney disease (Kandula et al. (2011)). Vitamin D regulates multiple signaling pathways that are linked to renal injury (Andress, 2006). Vitamin D stimulates FGF23 secretion and is inhibited by increased FGF-23 (Hu et al., 2012). Many studies have been conducted investigated the relation between FGF23 and vitamin D. Shimada et al. (2004) concluded FGF-23 serum as a potent regulator of phosphate and vitamin D homeostasis. Perwad et al. (2007) demonstrated that FGF-23 plays an important role in regulating Pi and vitamin D metabolism in vivo and acts directly to regulate 1α-hydroxylase and 24-hydroxylase gene expression in renal epithelial cells in vitro. Nakano et al. (2012) set vitam D deficiency and high level of serum FGF23 as a predictor of progression of CKD stages. Trummer et al. (2018) found that Vitamin D3 supplementation had no significant effect on FGF23 in the entire study cohort. However, they observed an increase of FGF23 concentrations in subgroups with low baseline 25(OH)D. Lerch et al. (2019) conducted study with 80 vitamin D deficient children with CKD. They found the supplements increased fibroblast growth factor 23 in children with advanced CKD.

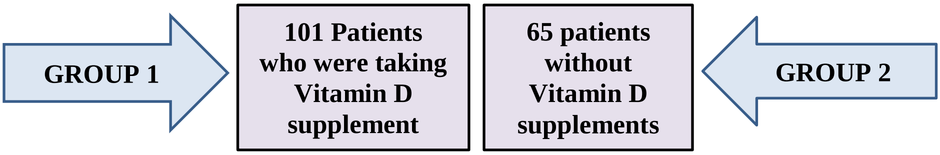
Chronic kidney disease (CKD) is progressively increasing in south Asian countries like Pakistan. Pakistan ranks eight in kidney disease causing 20,000 deaths every year. Chronic kidney disease is rapidly growing in Pakistan. Consuming junk and low quality food, self-medication or excessive use of medicine, low water intake, obesity, diabetes, hypertension, and renal stones are a few common causes of kidney disease in Pakistan. These facts tend us to the analysis of CKD patients. However, the main objective of the study is to determine the relationship between vitamin D intake and FGF23 metabolism in CKD patients.

2. MATERIAL AND METHODS

Data Source: The study was carried out at the Armed forces Institute of Pathology (AFIP) in the department of chemical pathology, the Armed forces institute of urology (AFIU) department of nephrology and Military Hospitals Rawalpindi over a period of 12 months from January 2017 to December 2017.

Sampling: Total of 166 patients were included in the study by non-probability convenience sampling was used.

Protocol of Study: A total of 166 patients with Chronic Kidney Disease were included in the study. These were divided into two groups, group A consisted of those who were on vitamin D supplements while group B consisted of those who were not given vitamin D supplements.



Cases were selected as per following criteria:

All patients who fulfilled the following criteria were included:

- All adult patients with age 18- years and above.
- Patients with CKD stages 1-5
- Patients of both gender

The patients who fulfilled the following criteria were excluded.

- All lactating and pregnant females.
- Patients of acute kidney injury.
- Patients with poor compliance and G.I disturbances like malabsorption.

Statistical Analyses: The statistical analyses were performed with STATA and MS EXCEL. Descriptive statistical graphical tools are used to describe the data. For comparisons between the groups, Student's t-test was used. Multiple linear regression was used to check the effect of vitamin D supplement intake, age and gender on the FGF23. All the tests were two sided, and a P value less than 0.05 was considered as significant.

3. RESULTS

According to the gender wise distribution provided in figure 1, there were 71 (42.7%) females and 95 (57.2%) males. Of these 71 females 42 (59.1%) belonged to group 1 whereas 29 (40.8%) patients fulfilled the criteria for group 2. Similarly amongst the males 59 (62.1%) patients belonged to group 1 category whereas 36 (37.8%)patients belonged to group 2.

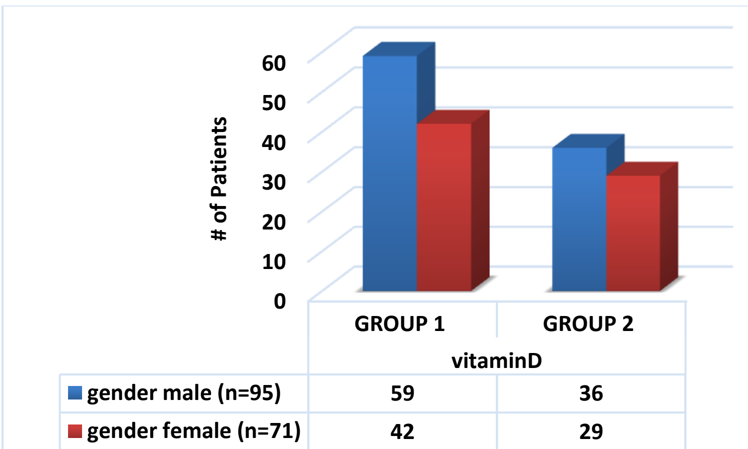


Figure 1: Gender Wise Distribution of Group 1 and Group 2.

The mean age of the patients was 53.51 years ± 16.010 with the youngest being at 18 years and the oldest at 95 years. However, the difference in mean age among the two groups is not statistically significant (52.79 years versus 54.64 years with p-value = 0.4681). Figure 2 represents the age-wise distribution of patients in both groups.

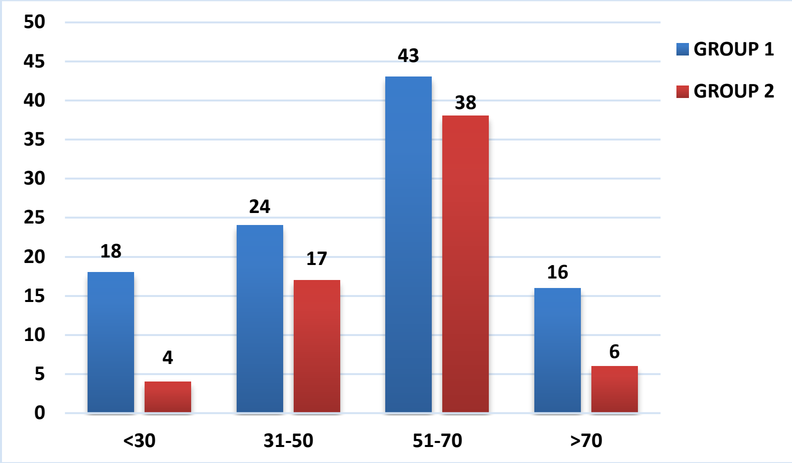


Figure 2: Representation of Group 1 and Group 2 into Different Age Groups

Figure 3 shows the distribution of CKD stages in group 1 (with vitamin D supplements) and group 2 (without vitamin D supplements) at the time of examination. It is very clear the number of patients in both groups was the CKD patients of stage 5.

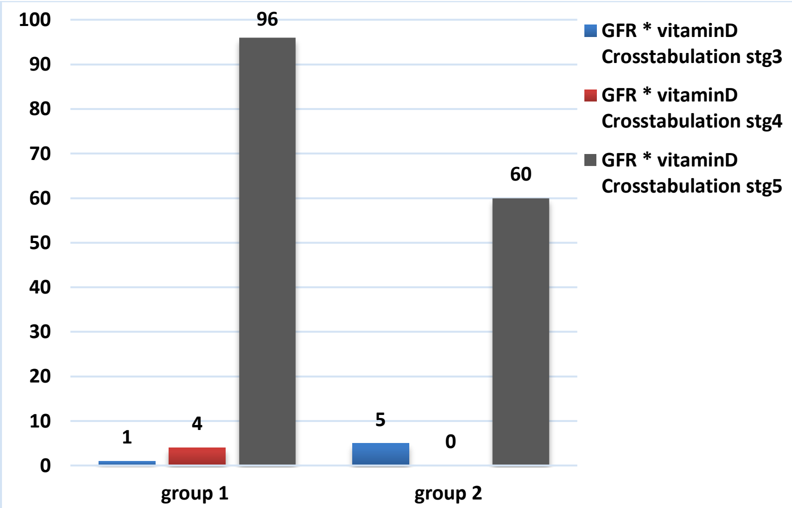


Figure 3: Distribution of CKD Stages in Group 1 and Group 2

Table 1 contains the results of the multiple linear regression with levels of FGF23 as dependent variable and age, gender and intake of vitamin D supplements are taken as independent variables.

Table 1
Regression Analysis of Levels of FGF23 and Study Parameters

Variables	Coefficients	Standard Errors	P-Value	95% C.I.
Age	-31.8649	8.2096	0.000	-48.07656 to -15.65325
Gender	-234.8025	264.5559	0.376	-757.2253 to 287.6203
Vitamin D	750.8202	268.1936	0.006	221.214 to 1280.426

The results show that gender is insignificantly related to levels of serum FGF23. While age and intake of vitamin D supplements have highly significant impact on the levels of serum FGF23.

4. COMMENTS AND CONCLUSION

Chronic kidney disease is a life-threatening condition. It presents with many associated co-morbidities like diabetes mellitus, hypertension, muscular and joint diseases, CVD and other chronic conditions. FGF23 a protein, is secreted by the osteocytes and osteoblasts while responding to high levels of phosphorous in serum and also due to high levels of active form of vitamin D. FGF23 has been proposed to be the initial adaptive response in early CKD patients for the protection of the organism from adverse effects of phosphate retention. In this project we aimed at assessing the levels of FGF23 in CKD patients who were on vitamin D supplementation with those patients without vitamin D supplementation. All adult patients with above 18 years age, patients with CKD stages 1-5 and both genders were included in the study. The pregnant women and those lactating, patients with acute kidney failure or injury and those with poor compliance and GI disturbances like malabsorption were excluded from the study. The difference between mean age of CKD patients among the groups was not statistically significant. However, patients with age 51 years to 70 years in both groups have great threat of having kidney disease.

Results of regression show that both genders are equally on risk of getting chronic kidney disease and there is no single group on significant threat. We have noticed that the FGF23 levels were significantly higher in CKD patients who were on vitamin D supplementation than those without vitamin D supplementation. The raise in FGF23 levels in the vitamin D supplementation group in this study proves that this complexity must be emphasized with great concern in chronic kidney patients. The protective or aggressive role of this protein bound needs further research in terms of different stages of CKD. The current study has many advantages as this was a comparative study and a reasonable number of cases were enrolled in the two groups. The detailed laboratory parameters related to CKD patients can also assessed.

REFERENCES

1. Andress, D.L. (2006). Vitamin D in chronic kidney disease: a systemic role for selective vitamin D receptor activation. *Kidney international*, 69(1), 33-43.
2. Hu, P., Xuan, Q., Hu, B., Lu, L., Wang, J. and Qin, Y.H. (2012). Fibroblast growth factor-23 helps explain the biphasic cardiovascular effects of vitamin D in chronic kidney disease. *International Journal of Biological Sciences*, 8(5), 663-671.
3. Isakova, T., Wahl, P., Vargas, G.S., Gutiérrez, O.M., Scialla, J., Xie, H., Appleby, D., Nessel, L., Bellovich, K., Chen, J. and Hamm, L. (2011). Fibroblast growth factor 23 is elevated before parathyroid hormone and phosphate in chronic kidney disease. *Kidney International*, 79(12), 1370-1378.
4. Kandula, P., Dobre, M., Schold, J.D., Schreiber, M.J., Mehrotra, R. and Navaneethan, S.D. (2011). Vitamin D supplementation in chronic kidney disease: a systematic review and meta-analysis of observational studies and randomized controlled trials. *Clinical Journal of the American Society of Nephrology*, 6(1), 50-62.
5. Lerch, C., et al. (2019). Vitamin D increases fibroblast growth factor 23 in children with advanced CKD. *Nephrol Dial Transplant.* doi:10.1093/ndt/gfy012.
6. Nakano, C., Hamano, T., Fujii, N., Matsui, I., Tomida, K., Mikami, S., Inoue, K., Obi, Y., Okada, N., Tsubakihara, Y. and Isaka, Y. (2012). Combined use of vitamin D status and FGF23 for risk stratification of renal outcome. *Clinical Journal of the American Society of Nephrology*, *7*(5), 810-819.
7. Perwad, F., Zhang, M.Y., Tenenhouse, H.S. and Portale, A.A. (2007). Fibroblast growth factor 23 impairs phosphorus and vitamin D metabolism in vivo and suppresses 25-hydroxyvitamin D-1α-hydroxylase expression in vitro. *American Journal of Physiology-Renal Physiology*, F1577-F1583.
8. Shimada, T., Hasegawa, H., Yamazaki, Y., Muto, T., Hino, R., Takeuchi, Y., Fujita, T., Nakahara, K., Fukumoto, S. and Yamashita, T. (2004). FGF-23 is a potent regulator of vitamin D metabolism and phosphate homeostasis. *Journal of Bone and Mineral Research*, 19(3), 429-435.
9. Trummer, C., Schwetz, V., Pandis, M., Grübler, M.R., Verheyen, V., Gaksch, M., Zittermann, A., März, W., Aberer, F., Steinkellner, J., Friedl, C., Brandenburg, V., Voelkl, J., Alesutan, I., Obermayer-Pietsch, B., Pieber, T.R., Tomaschitz, A. and Pilz, S. (2018). Effects if Vitamin D supplementation on FGF23: A randomize controlled trial. *European Journal of Nutrition*, 1-7.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 391-402

DISTANCE EDUCATION FOR DEVELOPMENT COMMUNITY IN KHAIRPUR CITY

Vijay Kumar[1] and **Faisal Afzal Siddiqui**[2]
Department of Statistics, Shah Abdul Latif University, Khairpur, Pakistan
Email: [1]brijbalvijaykrishna456@gmail.com
[2]brc.khi@gmail.com

ABSTRACT

The teacher Education has been required continuously on raising all over world for the qualified and quality teachers. The global necessary for teacher education is now greatly increased in the 21st century. Education as a necessary raised access results for globalization strategy. The development process was serving as a crucial part by Distance Education. The study is planned to measure the insight of the teaching, learning community about community development with special source to Distance Education. It was descriptive study, that targeted sample of the size was students, teachers, experts and member of the community. The targeted data of the sample size was analyzed by using statistical techniques served by SPSS. The findings reflected that observes Distance Education as a development tool. It is need to notice that the targeted audience has prominent performance by using the Distance Educational facility.

KEYWORDS

Distance, Education, Development, Community.

INTRODUCTION

The qualified and quality teacher has been required continuously on raising all over world. The global necessity for teacher education is now greatly increased in the 21st century (Midgley Hall Hardiman & Narine 1986). The teacher education has acquired a great significance, throughout. In order to participate in the growing concern of teacher's quality and lack of teacher, it is important to check the origin of the problem there is the type preparation of teacher and training which provide (Tesoriero and Ife 2006). Education as a necessary raised access results for globalization strategy was emphasized by 35.1 percent of respondents, Education as a necessary strategy for responding to globalization was emphasized by 35.1 percent of respondents, while 32.3 percent noted that increased access of the Internet and technology is necessary, and 25.8 percent mentioned for jobs or employment (Community Development outlook Survey 2013).

In our society we are a community and want to be better, there are many people who shared their experience, treasure and talent with us and gave their time to us to solve the problems for development in our community for several generations. There are many peoples in our society who are ready to deployed and use their skills to help for making the community better (Community Development outlook Survey 2013). The Local communities that can’t be connect and participate to larger and increasingly global online community upgraded to marginalize.

In this condition the Distance Education is going to play an important role in the education of teacher and training incipient since the United Nation Relief and Works Agency (UNRWA/UNESCO). Education as field of discipline was initiated in 1960. Education as a necessary raised access results for globalization strategy. Distance Education makes train teacher more willingly than conventional approaches (Ife 1995), to use innovative information and communication technologies and media. The role of distance education shifting technology enhanced open education from traditional education.

Distance Education resources are providing initial new opportunities. Development of community is playing broad term to perform the practices of civic activities involved citizens and professionals to build stronger and more resilient local communities.

METHODOLOGY

Distance learning or Distance Education for a community development is based on an outlook survey that targeted population. In this study the community arranged Teacher, Students, Employees, Employers, Owners experts and community members that were working in district Khairpur contributing for any form of teaching learning system or process to the development. Community Development recognized a number of different approaches including that Community Economic Development (CED), Community Capacity Building Social capital formation, Political participatory development, Community Participation and funding communities directly approaches. Often an individual basis (Mehmood, Akhter and Chaudhry 2016). The Distance Education or learning is a mode of delivering education and instructions to students, that are not presented in a cultural setting likewise classroom. Distance Education delivers, “access to learning when the information sources and the learners are separated through line and distance or may be both”.

Distance Educational courses demand for physically on-site presence for any case, containing Examination, combined or hybrid for study. Massive open online courses (MOOCs), highly standard aimed that interactive participation and open access web via and other technologies of network are developing recently in Distance Education (Christenson and Robinson, 1989).

POPULATION

The population of the teachers, students, employee’s experts and members of a community in Khairpur were targeted in this study, that form of teaching learning process or community development.

SAMPLE

In this study we have chosen total size of the sample was 100, that was targeted purpose by used purposive sample technique the selected sample is teachers, students, employers, employees, experts and community members in this study.

INSTRUMENT DEVELOPMENT

The collection of data used instrument for the purpose of development in community with consent and help of experts and experienced.

PROCEDURES

The researcher found the data with the help of questionnaire filled from sample size with their cooperation. The questionnaire has filled by an outlook survey to get proper results, permission and clearance approaching targets.

DISCUSSION AND DATA ANALYSIS

Gender		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	72	72.0	72.0	72.0
	Female	28	28.0	28.0	100.0
	Total	**100**	**100.0**	**100.0**	

In this study the classification for this study is the size of the Sample is 72 Male (72%) and 28 Female (28%).

Status		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Single	81	81.0	81.0	81.0
	Married	19	19.0	19.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case 81% Single were correctly classified and 19% were correctly classified for the Married

Language		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Sindhi	62	62.0	62.0	62.0
	Urdu	38	38.0	38.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case 62% Sindhi were correctly classified and 38% were correctly classified for the Urdu.

Age		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	13	3	3.0	3.0	3.0
	15	3	3.0	3.0	6.0
	16	10	10.0	10.0	16.0
	17	11	11.0	11.0	27.0
	18	10	10.0	10.0	37.0
	19	12	12.0	12.0	49.0
	20	13	13.0	13.0	62.0
	21	8	8.0	8.0	70.0
	22	5	5.0	5.0	75.0
	23	8	8.0	8.0	83.0
	24	5	5.0	5.0	88.0
	25	1	1.0	1.0	89.0
	26	2	2.0	2.0	91.0
	27	2	2.0	2.0	93.0
	28	2	2.0	2.0	95.0
	30	1	1.0	1.0	96.0
	32	1	1.0	1.0	97.0
	35	3	3.0	3.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case 03% 13 year aged, 03% 15 year aged, 10% 16 year aged,11% 17 year aged,10% 17 year aged,10% 18 year aged,12% 19 year aged, 13% 20 year aged, 8% 21 year aged,5% 22 year aged, 8% 23 year aged, 5% 24 year aged, 1% 25 year aged, 2% 26 year aged, 2% 27 year aged, 2% 28 year aged, 1% 30 year aged, 1% of 31 year aged were correctly classified and 3% were correctly classified for the 35 year aged.

Monthly Income		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3000	3	3.0	3.0	3.0
	7000	1	1.0	1.0	4.0
	8000	2	2.0	2.0	6.0
	10000	3	3.0	3.0	9.0
	11000	3	3.0	3.0	12.0
	12000	10	10.0	10.0	22.0
	12500	1	1.0	1.0	23.0
	13000	5	5.0	5.0	28.0
	14000	11	11.0	11.0	39.0
	14500	1	1.0	1.0	40.0
	15000	19	19.0	19.0	59.0
	15500	3	3.0	3.0	62.0
	16000	4	4.0	4.0	66.0
	17000	1	1.0	1.0	67.0
	18000	8	8.0	8.0	75.0
	19000	2	2.0	2.0	77.0
	20000	4	4.0	4.0	81.0
	22000	2	2.0	2.0	83.0
	23000	2	2.0	2.0	85.0
	25000	2	2.0	2.0	87.0
	26000	1	1.0	1.0	88.0
	30000	7	7.0	7.0	95.0
	40000	2	2.0	2.0	97.0
	45000	1	1.0	1.0	98.0
	99000	1	1.0	1.0	99.0
	125000	1	1.0	1.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case 3% 3000 of monthly income, were correctly classified, 1% 7000 of monthly income, 2% 8000 of monthly income, 3% 10000 of monthly income,3% 11000 of monthly income, 10% 12000 of monthly income, 1% 12500 of monthly income, 5% 13000 of monthly income, 11% 14000 of monthly income, 1% 14500 of monthly income, 19% 15000 of monthly income, 3% 15500 of monthly income, 4% 16000 of monthly income, 1% 17000 of monthly income, 8% 18000 of monthly income, 2% 19000 of monthly income, 4% 20000 of monthly income, 2% 22000 of monthly income, 2% 23000 of monthly income, 2% 25000 of monthly income, 1% 26000 of monthly income, 7% 30000 of monthly income, 2% 40000 of monthly income, 1% 45000 of monthly income, 1% 99000 of monthly income, and 1% were correctly classified for the 125000 of monthly income of the sample size.

Family Members		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	4	1	1.0	1.0	1.0
	5	1	1.0	1.0	2.0
	6	5	5.0	5.0	7.0
	7	4	4.0	4.0	11.0
	8	18	18.0	18.0	29.0
	9	24	24.0	24.0	53.0
	10	11	11.0	11.0	64.0
	11	7	7.0	7.0	71.0
	12	7	7.0	7.0	78.0
	13	5	5.0	5.0	83.0
	14	2	2.0	2.0	85.0
	15	7	7.0	7.0	92.0
	16	1	1.0	1.0	93.0
	17	2	2.0	2.0	95.0
	18	2	2.0	2.0	97.0
	19	2	2.0	2.0	99.0
	20	1	1.0	1.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case 1% 4 Family Members, were correctly classified, 1% 5 Family Members, 5% 6 Family Members, 4% 7 Family Members, 18% 8 Family Members, 24% 9 Family Members, 11% 10 Family Members, 7% 11 Family Members, 7% 12 Family Members, 5% 13 Family Members, 2% 14 Family Members, 7% 15 Family Members,1% 16 Family Members, 2% 17 Family Members, 2% 18 Family Members, 2% 19 Family Members and 1% were correctly classified for the 20 Family Members,

Educated Family Members		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	1	1.0	1.0	1.0
	2	8	8.0	8.0	9.0
	3	27	27.0	27.0	36.0
	4	18	18.0	18.0	54.0
	5	12	12.0	12.0	66.0
	6	16	16.0	16.0	82.0
	7	10	10.0	10.0	92.0
	8	3	3.0	3.0	95.0
	9	2	2.0	2.0	97.0
	12	1	1.0	1.0	98.0
	13	1	1.0	1.0	99.0
	20	1	1.0	1.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case 1% 1 Educated Family Member, were correctly classified, 8% 2 Educated Family Members, 27% 3 Educated Family Members, 18% 4 Educated Family Members, 12% 5 Educated Family Members, 16% 6 Educated Family Members, 10% 7 Educated Family Members, 3% 8 Educated Family Members, 2% 9 Educated Family Members, 1% 12 Educated Family Members, 1% 13 Educated Family Members and 1% were correctly classified for the 20 Educated Family Members.

Uneducated Family Members		**Frequency**	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	0	9	9.0	9.0	9.0
	1	9	9.0	9.0	18.0
	2	3	3.0	3.0	21.0
	3	7	7.0	7.0	28.0
	4	5	5.0	5.0	33.0
	5	19	19.0	19.0	52.0
	6	8	8.0	8.0	60.0
	7	10	10.0	10.0	70.0
	8	8	8.0	8.0	78.0
	9	5	5.0	5.0	83.0
	10	12	12.0	12.0	95.0
	11	1	1.0	1.0	96.0
	12	4	4.0	4.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case 9% 0 Uneducated Family Members were correctly classified, 9% 1Uneducated Family Member, 3% 2 Uneducated Family Members, 7% 3Uneducated Family Members, 5% 4Uneducated Family Members, 19% 5Uneducated Family Members, 8% 6Uneducated Family Members, 10% 7Uneducated Family Members, 8% 8Uneducated Family Members, 5% 9Uneducated Family Members, 12% 10Uneducated Family Members, 1% 11Uneducated Family Members and 4% were correctly classified for the 12 Uneducated Family Members.

Idea about Distance Education		**Frequency**	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	yes	100	100.0	100.0	100.0

In this study the classification for this study is the size of the 100 Sample is 100% have idea about Distance Education.

Lesson Delivered		**Frequency**	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	In one time	17	17.0	17.0	17.0
	Periodically	83	83.0	83.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case 17% students wants there course in one time for Distance Education were correctly classified and 83% students wants there course in Periodically were correctly classified for the Distance Education.

Use Library		**Frequency**	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Yes	25	25.0	25.0	25.0
	No	29	29.0	29.0	54.0
	Sometimes	46	46.0	46.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case 25% students are who use library for their study were correctly classified, 29% students who don’t use library for the study of Distance Education and 46% students who use library sometimes for Distance Education were correctly classified.

PC at Home		**Frequency**	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Yes	73	73.0	73.0	73.0
	No	20	20.0	20.0	93.0
	Sometimes	7	7.0	7.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case 73% candidates who have their own PC at their home for Distance Education course were correctly classified and 20% candidates don’t have their own PC at home were correctly classified.

Join Distance Educational Institute		**Frequency**	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Yes	48	48.0	48.0	48.0
	No	10	10.0	10.0	58.0
	No Idea	42	42.0	42.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case 48% answered yes to join a Distance Educational Institute were correctly classified, 10% answered don’t like to join a Distance Educational Institute and 42% who have no idea to join a Distance Educational Institute were correctly classified.

Distance Education Importance		**Frequency**	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Yes	57	57.0	57.0	57.0
	No	1	1.0	1.0	58.0
	Sometimes	35	35.0	35.0	93.0
	No Idea	7	7.0	7.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case 57% answered that Distance Education is important to Develop a Community were correctly classified, 1% is not important for development, 35% is sometimes it may be important for the Development and 7% has no idea were correctly classified for the important to development a community.

Suggested for Distance Education		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	19	19.0	19.0	19.0
	No	64	64.0	64.0	83.0
	No Thinking	17	17.0	17.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case 19% suggested to others to join Distance Educational course were correctly classified, 64% has not suggested to any other for Distance Education and 17% has no thinking about suggested for Distance Education were correctly classified.

Interested for Level		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Beginners	35	35.0	35.0	35.0
	Intermediate	21	21.0	21.0	56.0
	Advanced	36	36.0	36.0	92.0
	Technical	8	8.0	8.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case candidates were interested to joint Distance Education 35% Beginners level were correctly classified, 21% Intermediate level, 36% Advanced level and 8% were correctly classified for the technical level.

Necessary for Development		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	57	57.0	57.0	57.0
	No	2	2.0	2.0	59.0
	Sometimes	19	19.0	19.0	78.0
	May be	22	22.0	22.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy for Distance Education in this case 57% Necessary for Development were correctly classified, 2% has unnecessary to Development, 19% sometimes for Development and 22% may be Necessary for Development were correctly classified for the Distance Education.

Own Idea for Distance Learning		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Easy	49	49.0	49.0	49.0
	Difficult	29	29.0	29.0	78.0
	Easy but time consume	20	20.0	20.0	98.0
	Time consume	2	2.0	2.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy for the Distance Learning/Education in this case 49% Easy were correctly classified, 29% Difficult, 20% Easy but time consumed and 2% were correctly classified for the time consumed.

Categories		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Student	67	67.0	67.0	67.0
	Teacher	5	5.0	5.0	72.0
	Employer	11	11.0	11.0	83.0
	Employee	17	17.0	17.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case for the sample size 67% Students were correctly classified, 5% teachers, 11% Employer and 17% were correctly classified for the Employees.

Helpful for Conventional Course		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	71	71.0	71.0	71.0
	Disagree	1	1.0	1.0	72.0
	Strongly Agree	19	19.0	19.0	91.0
	Strongly Disagree	1	1.0	1.0	92.0
	Neutral	8	8.0	8.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case 71% Agree were correctly classified, 1% disagree, 19% strongly agree, 1% strongly disagree and 8% were correctly classified for the Neutral.

Lesson Format		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Online video	9	9.0	9.0	9.0
	DVD video	4	4.0	4.0	13.0
	Written notes	44	44.0	44.0	57.0
	Video and Written	32	32.0	32.0	89.0
	Tutorial	11	11.0	11.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case 9% candidates want their lessons in online video format were correctly classified, 4% DVD videos format, 44% candidates want their lessons in written notes were correctly classified, 32% candidates want video and written format and 11% were correctly classified for the Tutorial.

Courteously and Professionally		**Frequency**	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Agree	57	57.0	57.0	57.0
	Disagree	10	10.0	10.0	67.0
	Strongly Agree	11	11.0	11.0	78.0
	Strongly Disagree	3	3.0	3.0	81.0
	Neutral	19	19.0	19.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case 57% Agree with treated courteously and professionally were correctly classified, 10 disagree 11% strongly agree 3% strongly disagree and 19%% were correctly classified for the Neutral.

Satisfied to Answer and solutions		**Frequency**	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Agree	57	57.0	57.0	57.0
	Disagree	21	21.0	21.0	78.0
	Strongly Agree	17	17.0	17.0	95.0
	Neutral	5	5.0	5.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case candidates were satisfied 57% agree with their answers and solutions were correctly classified, 21% disagree 17% strongly agree and 5% were correctly classified for the neutral.

Fit Life Style and Study Habits Work		**Frequency**	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Agree	38	38.0	38.0	38.0
	Strongly Agree	48	48.0	48.0	86.0
	Neutral	14	14.0	14.0	100.0
	Total	**100**	**100.0**	**100.0**	

The statistic assumes sampling adequacy in this case 38% agree with Distance Education is to fit their life were correctly classified, 48% strongly agreed with Distance Education is to fit their life were correctly classified and 14% were correctly classified for the neutral. So 48% candidates were agreed with Distance Education fit their life style and study habits.

FINDINGS

After Data analysis we find the information that drawn following:

- According to this study the majority of student, teachers, experts, and members of the community were agreed with initial learning education for professional teacher skill, and necessary for development.
- Most of the students are strongly agreed with distance education is most important to develop a community and necessary for initial teacher education.
- The majority of the students, teachers, experts and community members agree to improve some weakness and upgraded to teaching learning process.
- The most sample were strongly agreed with the development of community is necessary by distance education.
- Distance education is going to upgrade from conventional education the technology enhanced open education.
- The most sample were wanted to join distance educational institute and the majority were interested to joint at advanced level.
- The most majority were students and they felt it is very easy and satisfied to their answer and solution and also agreed with fit their life style and study habits work.
- Our research showed that the interest of the growing Education, teacher education and Development of the community is sustainable throw Distance Education.

REFERENCES

1. Christenson, J.A. and Robinson, J.W. (1989). *Community development in perspective*. Iowa State University Press.
2. Federal Reserve Bank of St. Louis (2013). *Community Development Outlook Survey*. Federal Reserve Bank of St. Louis-Little Rock Branch 501-324-8495.
3. Ife, J.W. (1995). *Community development: Creating community alternatives-vision, analysis and practice*. Longman Australia.
4. Mehmood, T., Akhter, M. and Chaudhry, A.H. (2016). Community Development through Open Learning and Distance Education. *Secondary Education Journal*, 1(1), 1-14.
5. Midgley, J., Hall, A., Hardiman, M. and Narine, D. (1986). *Community participation, social development and the state*. Methuen.
6. Tesoriero, F. and Ife, J. (2006). *Community development: community-based alternatives in an age of globalization*, Pearson Education Australia.

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 403-408

ENVIRONMENTAL IMPACTS OF THE USE OF DISPOSABLE CUPS IN CAFETERIA OF NCBA&E, LAHORE

Muhammad Mohsan[1] and **Mahmoona Ashraf**[2]
National College of Business Administration and Economics, Lahore, Pakistan
Email: [1]mahmoonaashraf1@gmail.com
[2]mohsan771@gmail.com

ABSTRACT

The study was conducted in cafeteria of National College of Business Administration and Economics University Lahore. Before disposable cups, shared cups were in use which were replaced by disposable cups because of hygiene. Now a days different types of disposable cups are in use, paper cups, PET (polyethylene terephthalate), Poly coated cups, Polystyrene cups and foam cups. In College of Business Administration and Economics University Lahore (NCBA&E, Lahore) 150, 200 foam cups (made from Styrofoam) are in use on daily basis. Styrofoam has health as well as environmental issues. 200 cups are being in used in daily basis, 500 on monthly and 60000 on annually basis. Foam cups are made up of Polystyrene which is carcinogenic. In Pakistan landfilling and dumping are better methods where high amount of foam cups reduces the lifespan of dump/landfill sites and are less recyclable. In landfill sites Styrene react under high temperature increasing cost of leachate treatment and reduces gas productivity. Under sunlight photo degradatioon occur causing powdery form which contaminate soil. Shared cups are costly but it is a onetime investment and more hygienic.

1. INTRODUCTION

Use of disposable cups is very common at public places like markets, hospital canteens, and cafeterias of schools, colleges and universities, airports, railways stations etc. The culture of disposable cups was introduced in early twenties century. Before the use disposable cups people used shared cups and glasses at public places. About hundred years back this concept was introduced that shared cups and glasses are not hygienic and causing transfer of disease from one person to the others. After this, people being concerned about their health started the use of disposable cups. In this century different types of disposable cups were introduced with respect to their use and quality of materials. The common types with respect to their materials are paper, plastic and foam cups. These categories are further subdivided into biodegradable, degradable, compostable, recycles, made from paper, plastic, polystyrene and extended polystyrene. Where there are benefits of using these cups, we must also keep in mind that these are terribly effecting the environment. In order to study the impacts of disposable cups we must know some of the most used types of disposable cups for coffee and tea.

2. LITERATURE REVIEW

Paper Cup A product that is biodegradable can be naturally broken down by microorganisms, such as bacteria or fungi, and absorbed back into the ecosystem. These products are composed primarily of naturally-occurring constituent parts, and are often considered "green" or "eco-friendly" products. PET (polyethylene terephthalate) part of the polyester family, it is used to make synthetic fibers as well as food and beverage containers. Products made with PET are lightweight and are proficient at blocking gases, solvents, and moisture. They are also strong and impact-resistant. Products made from PET can also be recycled. Poly-Coated Cups Featuring a polymer coating for added insulation, poly-coated cups are great for serving hot or cold beverages alike! These cups may have a single or double poly-coating, with each additional layer offering additional rigidity. These cups come in a wide range of colors and sizes, and many feature tightly rolled rims for worry-free, leak-proof drinking. Polystyrene available in solid color and clear options, polystyrene cups are made from a high impact material that is sturdy, shatterproof, and flexible. These cups feature smooth rolled rims and come in a variety of styles, sizes, and shapes to meet your unique needs. Polystyrene plastic cups, like most plastic cups, are only designed for use with cold beverages though, so keep that in mind when deciding on the perfect cup. Foam Cups a great insulator, foam cups make an ideal option for hot or cold beverages. These cups hold their shape better than paper cups, yet still offer your establishment a low-cost, lightweight vessel.

Paper Cup PET Poly Coated Polystyrene Foam Cup

Figure 1: Types of Disposable Cups

Area selected for this study is cafeteria of National College of Business Administration and Economics University Lahore (NCBA&E, Lahore). More than five hundred students visit the cafeteria of the university daily and take food, drinks, coffee and tea. Out these five hundred, 150 to 200 students use to take coffee and tea which served in disposable cups. Ceramic cups (Reusable) are also available but students prefer to be served in disposable cups. According to the manager of the cafeteria only 5 to 10 students comes whose prefer to use ceramic cups. The cups used in the cafeteria are foam cups. The material used to mad these cups is called Styrofoam. *"Styrofoam" is actually just a brand name.* The white material making up your coffee cup is actually called expanded polystyrene and is very dangerous to the environment.

Polystyrene is a type of plastic. It can be expanded into the foam material that is known as Styrofoam by expanding it with 95% air. It is often used for insulating hot food and beverages. Due to cost effectiveness, no toxicity and high shelf life it is good for business. The Styrofoam downside is, it is not environments friendly and remains in environment for generations. Styrofoam is not biodegradable but it can be chemically broken down into small pieces. These pieces ultimately covers the major portion of landfill. According to Max Roman the phenomenon of breaking down of Styrofoam vary from few years to one million years depending upon the environmental conditions.

3. DISCUSSION

One downside of Styrofoam is health issues. Extended polystyrene (Styrofoam) is combination of polystyrene and air. Polystyrene is derived from styrene which is not friendly to human health. According to National Research Council of USA, styrene classified as carcinogen is found in the tissues of almost every person. Because polystyrene is part of almost all kinds of plastics used for packaging, hot and cold drinks and many others.

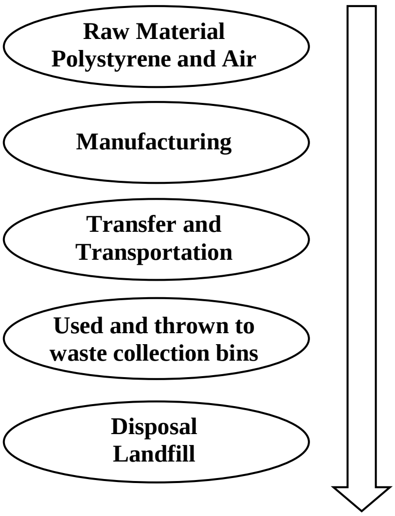


Figure 2: Life Cycle of Disposable Cups

The other downside of Styrofoam is environmental impacts because of non-biodegradability and less ability to recycling. As mentioned above about 200 cups are used on daily basis. For a long term calculations more than 5000 in a month and above 60000 in a year. This is huge amount of cups of cups used in a very small area. From this we can picture what a higher amount it will be if we consider all the institutions or the consumption of foam cups in the whole city. Being a developing country we don't have much resources to tackle such a pollutant. In simple it is a major part of waste disposed every day. Having $50kg/m^3$ ($0.05g/cm^3$) is also not a good indicator to the disposal process. Since the disposal of waste is not a single phase process, it comprises of collection, transfer and disposal. At every stage less density is problematic. Less density means more volume with less mass. Hence less density causes reduction on capacity of waste collection container/bins and transfer truck. After collection waste is transfer to the disposal station. In a country like Pakistan due to economic problems, dumping or landfill are the only best suited processes for waste disposal. In both of these processes density matters a lot. Less density results in reducing the life of dump/landfill site. High amount of disposable foam cups in waste results in low life span of dump/landfill site.

Waste is considered an asset which contributing a lot to the economic culture. In our system many people consider the collection of waste is a bright business because of recyclables present in the solid waste. Foam cups are recyclable but the process of recycling of cups is much costly and complicated which is not good for people in this business. Recycling of foam is also not considerable even in the developed countries like UK. According to a report of THE INDEPENDENT their only three recycling facilities for foam cups. Yearly 2.5 billion disposable cups are used in UK, out of these less than one percent are recycled.

Even after the disposal the story does not end. Two long term outcomes from landfill site are leachates and Gas Production. Landfill site is closed by placing a thick layer of soil. The temperature of waste keeps on increasing under biological processes occurring under the layer of soil. Styrene in the foam cups becomes reactive under the higher temperature, leachate which is already highly contaminated becomes more dangerous. It also increases the cost of leachates treatment. Gas from the biological process of the waste is also a source of energy. Larger amount of foam cups also reduce the gas production ability of the waste. Styrofoam (foam cup) is also sensitive before sunlight. When is exposed to sunlight continuously for months photo degradation can occur which converts it to the powder form. This powder form may affect the fertility of land.

Foam cups are light weight which also makes its use more negative for the environments. Due to light weight wind can easily carried them away from the disposal system causing littering of street and clogging of sewerage systems. During travels and picnics throwing of the cups may affect the marine life.

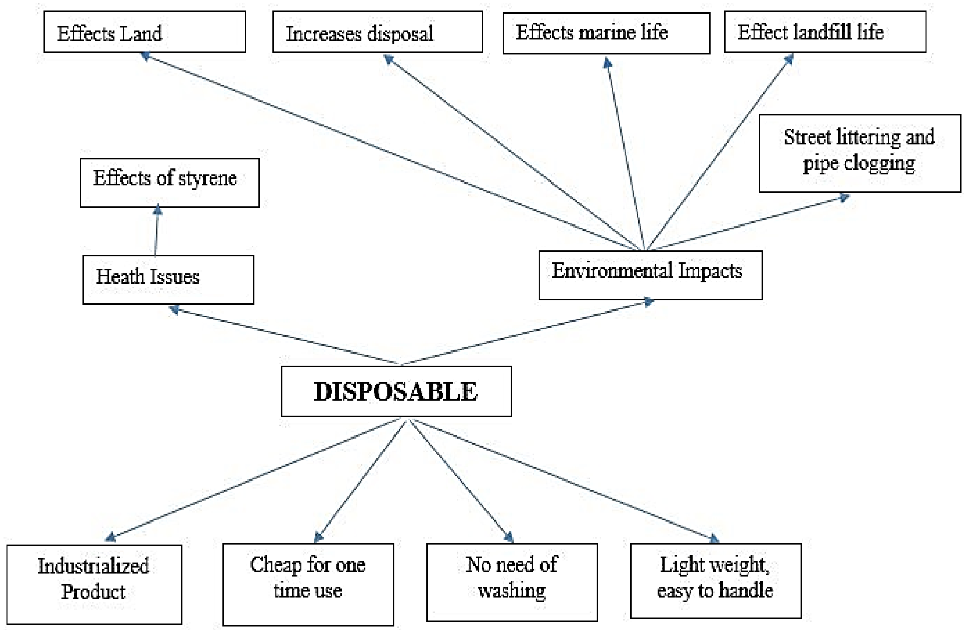
From economic perspective both disposable and reusable cups can be analyzed as follow:

Type	**Useful Life**	**Average Price**	**Hygienic**	**Washing Cost**	**Disposal Cost**
Disposable	1 time	7.5	No	No	Much More
Reusable	>60 times	85	Yes	Yes	Not considerable

According to above perimeters we can see the reusable cup is much better than disposable cup. There is only price not justifiable for a single time use. But for more than one time use reusable cups are even more economical than disposable cups. From this we can say that is much better to use reusable cups than disposable cups. Life time use of reusable cups is less influencing the environmental profile (Ligthart and Ansems, 2007).

4. CONCLUSIONS

Effects we are facing



Why are we using Disposable Cups?

Figure 3: Problem Tree Diagram

Since the benefits of using foam are not as much as we have face there negative impacts on human health and environments. Impacts of foam cups starts from there use till there break down. It cannot be recycled. Foam cups are also not degradable easily. Use of reusable or paper cup can overcome many problems causing by using foam cups. Figure 3 is the summery of the whole study which show why are we using disposable cups and what impacts we are facing.

5. RECOMMENDATIONS

- Use of these types of materials must be prohibited because of their health and environmental impacts.
- Use of ceramic or paper cups should be promoted.
- Use of reusable cups is more economical than disposable cups.
- Reusable (ceramic) are more hygienic than disposable cups.
- Use of disposable cups can be reduced by giving some discount to the customers on taking coffee or tea in reusable cups and more to those having their on cups.

REFERENCES

1. Kremer, A. (2003). *Cradle to grave: the life-cycle of styrofoam*. Urban Studies Program, San Francisco State University, Spring.
2. Ziada, H. (2009). *Disposable coffee cup waste reduction study*. Unpublished manuscript, McMaster University, Hamilton, Ontario. Retrieved from http://wbooth. mcmaster. ca/epp/publications/student/DisposableCoffeeCup.pdf.
3. Ligthart, T.N. and Ansems, A.M.M. (2007). *Single use cups or reusable (coffee) drinking systems: an environmental comparison*. Netherlands Organization for Applied and Scientific Research, TNO Report, (2006-A), R0246.
4. http://sciencing.com/long-styrofoam-break-down-5407877.html
5. http://www.webstaurantstore.com/guide/610/types-of-disposable-cups.html
6. http://www.independent.co.uk/environment
7. http://www8.nationalacademies.org/onpinews/newsitem.aspx?RecordID=18725
8. http://www.1millionwomen.com.au/blog/why-you-should-say-no-styrofoam/

Proc. 17th International Conference on Statistical Sciences
Lahore, Pakistan – January 21-23, 2019, Vol. 33, pp. 409-415

ON BAYESIAN ESTIMATION OF THE GENERALIZED INVERSE WEIBULL DISTRIBUTION USING INFORMATIVE PRIORS

Mahnoor Shahzad[1] **Wajiha Nasir**[1], **Murafa Hanif**[1] and **Muhammad Zubair**[4]
[1] Department of Statistics, Govt. College Women University,
Sialkot, Pakistan.
Email: mahnoors240@gmail.com
wajiha.nasir@gcwus.edu.pk
murafahanif@gmail.com
[2] Department of Statistics, University of Sargodha,
Sargodha, Pakistan. Email: zubair.bakhsh@uos.edu.pk

ABSTRACT

The generalized inverse Weibull distribution can be readily applied to a wide range of situations including applications in medicine, reliability and ecology. A three parameter generalized inverse Weibull distribution has been studied. Posterior distribution has been derived by using informative prior. Bayes estimators and their corresponding risks for posterior distribution has been derived by using different loss function. Performance of Bayes estimator and their corresponding risk has been studied by Monte Carlo Simulation for different sample size. We have also used real life example to compare our results. We also have been plotted posterior distribution for scale parameter.

1. INTRODUCTION

Bayesian statistics is a theory in the field of the statistics based on the Bayesian interpretation of the probability expresses a degree of belief in an event which can change as view information is gathered, rather than a fixed value based upon frequency. Bayesian statistics begins with an uneducated opinion called the prior. The prior is represented mathematically as a probability to an event or state of the world. Sindhu et al. has studied the Kumaraswamy distribution under failure censoring sampling scheme. Feroze and Yaseen (2015) has studied nakagami distribution under Bayesian Analysis using left censoring scheme. They have used informative as well as non-informative priors to study the distribution.

The generalized inverse Weibull distribution is widely used in business managements etc. In this chapter we use three parameters Generalized inverse Weibull distribution in which shape parameter is known and scale parameter is unknown. A random variable t is said to be generalized inverse distribution if its p.d.f has following form

$$f(t,\alpha,\gamma)=\gamma\beta\alpha^{\beta}t^{-(\beta+1)}e^{-\gamma\left(\frac{\alpha}{t}\right)^{\beta}} \tag{1}$$

where α and β is shape parameter and γ is scale parameter. In this paper we see the behavior of the scale parameter under Bayesian analysis.

2. BAYESIAN ANALYSIS USING COMPLETE SAMPLES

In this section the Bayesian analysis under complete samples technique has been utilized. Posterior distribution has been derived under information prior (Gamma and Exponential). The Bayes estimators and their corresponding risks are derived by using Square error loss function (SELF), Precautionary loss function (PLF) and Degroot error loss function (DeLF) different loss functions.

Let the random sample $x_1, x_2, ..., x_n$ is assumed to be taken from GIW distribution with known shape parameter ϕ and assuming $\phi = 1$ and with unknown parameter. Then likelihood function is

$$L(t,\alpha,\beta,\gamma) = \gamma^n \beta^n \alpha^{n\beta} \prod_{i=t}^{n} t^{-(\beta+1)} \exp - \left[\gamma \left(\frac{\alpha}{t} \right)^{\beta} \right] \tag{2}$$

The Exponential distribution is used as informative prior with parameter α, β, γ is

$$P(\gamma) = \frac{1}{\theta} e^{-\frac{\gamma}{\theta}} \qquad 0 < \gamma < \infty \tag{3}$$

Now the posterior distribution of γ using (2) and (3) is

$$p(\gamma \mid x) = \frac{\gamma^{(n+1)-1} e^{-\left(l + \sum_{i=1}^{n} \left(\frac{\alpha}{t} \right)^{\beta} \right)}}{\Gamma(n+1)} \tag{4}$$

Now the α_1 and β_1 from prior distribution of exponential distribution is as follows

$$a_1 = n+1 \qquad , \beta_1 = \sum_{i=1}^{n} \left(\frac{\alpha}{t} \right)^{\beta}$$

The Bayes estimator and posterior risks under SELF, PLF, DeLF by using exponential prior are

$$\hat{\gamma}_{(SELF)} = \frac{n+1}{\frac{1}{c} + \sum_{i=1}^{n} \left(\frac{\alpha}{t} \right)^{\beta}} \quad ; \qquad \rho_{(SELF)} = \frac{n+1}{\left\{ \frac{1}{c} + \sum_{i=1}^{n} \left(\frac{\alpha}{t} \right)^{\beta} \right\}^2}$$

$$\hat{\gamma}_{(SELF)} = \frac{\sqrt{(n+1)(n+2)}}{\frac{1}{c} + \sum_{i=1}^{n} \left(\frac{\alpha}{t} \right)^{\beta}} ; \qquad \rho_{(PLF)} = 2 \left(\frac{\sqrt{(n+1)(n+2)} - (n+1)}{\frac{1}{c} + \sum_{i=1}^{n} \left(\frac{\alpha}{t} \right)^{\beta}} \right)$$

$$\hat{\gamma}_{(DELF)} = \frac{n+2}{\frac{1}{c}+\sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}} \quad ; \qquad \rho_{(DELF)} = \frac{1}{n+a+1}$$

The Gamma distribution is used as informative prior with parameter a, b, γ is

$$\rho(\gamma) = \frac{b^{a}}{\Gamma(a)}\gamma^{a-1}e^{-\gamma b} \qquad 0 < \gamma < \infty \tag{5}$$

Now the posterior distribution of Gamma prior is as follows

$$p(\gamma \mid x) = \frac{\gamma^{(n+a)-1}e^{-\gamma\left(b+\sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}\right)}}{\Gamma(n+a)} \tag{6}$$

Now the α_2 and β_2 from prior distribution of Gamma distribution is as follows

$$\alpha_2 = n+a \qquad ; \quad \beta_2 = b+\sum_{i=t}^{n}\left(\frac{\alpha}{t}\right)^{\beta}$$

The Bayes estimators and posterior risks under SELF, PLF and DELF using Gamma distribution is as follow

$$\hat{\gamma}_{(SELF)} = \frac{n+a}{b+\sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}} \quad \rho_{(SELF)} = \frac{n+a}{\left\{b+\sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}\right\}^{2}}$$

$$\hat{\gamma}_{(PLF)} = \frac{\sqrt{(n+a)(n+a+1)}}{b+\sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}} \quad \rho_{(PLF)} = 2\left\{\frac{\sqrt{(n+a)((n+a+1)} - (n+a)}{b+\sum_{i=1}^{n}(\frac{\alpha}{t})^{\beta}}\right\}$$

$$\hat{\gamma}_{(DELF)} = \frac{n+a+1}{b+\sum_{i=1}^{n}\left(\frac{\alpha}{t}\right)^{\beta}} \quad ; \rho_{(DELF)} = \frac{1}{n+a+1}$$

3. ELICITATION OF HYPER PARAMETERS

In this section, we have elicited the hyper parameters by using the method introduced by Aslam (2003). For this we have derive prior predictive distribution by using following equation

$$p(y) = \int_{-\infty}^{\infty} p(\gamma)\, p(y|\gamma)\, d\gamma \tag{7}$$

By using eq.7 prior predictive distribution using gamma prior is

$$p(y)=\frac{ab^{a}\beta\alpha^{\beta}\left(y\right)^{-(\beta+1)}}{\left\{b+\left(\frac{\alpha}{y}\right)^{\beta}\right\}^{a+1}} \tag{8}$$

The values of hyper parameters are $a=2.5412,\ b=1.2546$ obtained by method of elicitation.

By using eq.7 prior predictive distribution using exponential prior is

$$p(y)=\frac{\beta\alpha^{\beta}\left(y\right)^{-(\beta+1)}}{c\left\{\frac{1}{c}+\left(\frac{\alpha}{y}\right)^{\beta}\right\}^{a+1}} \tag{9}$$

The values of hyper parameters are $c=0.8412$ obtained by method of elicitation.

4. SIMULATION STUDY

Simulation is a way to model random events, such that simulated outcomes closely match real world outcomes. By observing simulated outcomes, researchers gain insight on the real world. Simulation is used in many contexts, such as simulation of technology for performance optimization, safety engineering, testing, training, education etc. often computer experiments are used to study simulation models. Simulation is used advantageously in a number of simulations. This includes providing the empirical estimation of sampling distributions, studying the misspecification of assumptions in statistical procedures, determining the power in hypothesis tests, etc. the use of simulation study in business are varied and it is often utilized when conducting experiments on a real system is impossible or impractical. In this, the simulation study is carried out to obtain the BE's and PR's under different loss functions using different priors. The simulation process is repeated 10,000 times considering generation of random samples of size $n\in\{30,70,100,200\}$ by using inverse method assuming $\alpha\in\{2,5\}$, $\beta\in\{2,3\}$ and $\gamma\in\{2,3\}$. The results have then been averaged. These results are shown in the following tables.

Table 1
Expression of BEs and PRs using Gamma Prior

	$\alpha=2,\beta=2,\gamma=2$			$\alpha=2,\beta=2,\gamma=3$		
	SELF	**PLF**	**DeLF**	**SELF**	**PLF**	**DeLF**
30	2.1144 (0.1491)	2.1489 (0.0676)	2.1799 (0.0312)	3.1651 (0.3345)	3.2134 (0.1012)	3.2705 (0.0312)
70	2.0327 (0.0599)	2.0640 (0.0287)	2.0797 (0.01388)	3.0734 (0.1350)	3.0807 (0.0429)	3.1078 (0.0138)
100	2.0327 (0.0413)	2.0427 (0.02007)	2.0562 (0.0098)	3.0479 (0.0929)	3.0636 (0.0301)	3.0784 (0.0098)
200	2.0188 (0.02037)	2.0241 (0.0100)	2.02696 (0.0049)	3.0271 (0.0458)	3.0332 (0.015)	3.0381 (0.0049)
	$\alpha=2,\beta=3,\gamma=2$			$\alpha=2,\beta=3,\gamma=3$		
30	2.1172 (0.1497)	2.1495 (0.0677)	2.1876 (0.0312)	3.1630 (0.3339)	3.2181 (0.1013)	3.2541 (0.0312)
70	2.0496 (0.06003)	2.0655 (0.0287)	2.0771 (0.0138)	3.0741 (0.1350)	3.0873 (0.043)	3.1096 (0.0138)
100	2.0357 (0.0414)	2.0427 (0.0200)	2.05419 (0.0098)	3.0459 (0.0928)	3.0665 (0.0301)	3.0773 (0.0098)
200	2.0173 (0.0203)	2.0205 (0.01)	2.0287 (0.0049)	3.0258 (0.0457)	3.0326 (0.015)	3.0372 (0.0049)
	$\alpha=5,\beta=2,\gamma=2$			$\alpha=5,\beta=2,\gamma=3$		
30	2.1146 (0.1493)	2.1446 (0.0675)	2.1834 (0.0312)	3.1645 (0.3344)	3.2070 (0.101)	3.2515 (0.0312)
70	2.0479 (0.0599)	2.0615 (0.0287)	2.0833 (0.0138)	3.0662 (0.1343)	3.0874 (0.043)	3.1097 (0.0138)
100	2.0367 (0.0414)	2.0462 (0.0201)	2.0551 (0.0098)	3.0479 (0.0929)	3.0558 (0.03)	3.0826 (0.0098)
200	2.0202 (0.0204)	2.0234 (0.01)	2.0294 (0.0049)	3.0238 (0.0457)	3.0280 (0.015)	3.0377 (0.0049)
	$\alpha=5,\beta=3,\gamma=2$			$\alpha=5,\beta=3,\gamma=3$		
30	2.0470 (0.1501)	2.1471 (0.0676)	2.1876 (0.0312)	3.1641 (0.3341)	3.2153 (0.1012)	3.2634 (0.0312)
70	2.0470 (0.0598)	2.0637 (0.0287)	2.0804 (0.0138)	3.0702 (0.1347)	3.0889 (0.043)	3.1135 (0.0138)
100	2.034 (0.0413)	2.0444 (0.02)	2.0564 (0.0098)	3.0421 (0.0925)	3.0652 (0.0301)	3.0738 (0.0098)
200	2.0185 (0.0203)	2.0191 (0.01)	2.0297 (0.0049)	3.02387 (0.0457)	3.031 (0.015)	3.0364 (0.0049)

Table 2
Expression of BEs and PRs using Exponential Prior

	$\alpha=2,\beta=2,\gamma=2$			$\alpha=2,\beta=2,\gamma=3$		
	SELF	**PLF**	**DeLF**	**SELF**	**PLF**	**DeLF**
30	2.2513 (0.1407)	2.2915 (0.0607)	2.3214 (0.0263)	3.1960 (0.2827)	3.2457 (0.0859)	3.2860 (0.0263)
70	2.1197 (0.0591)	2.1326 (0.0274)	2.1457 (0.0128)	3.0959 (0.1260)	3.1129 (0.0400)	3.1319 (0.0128)
100	2.0827 (0.0409)	2.0940 (0.0194)	2.1062 (0.0092)	3.0716 (0.0889)	3.0857 (0.0286)	3.0986 (0.0092)
200	2.0423 (0.0202)	2.0474 (0.0098)	2.0513 (0.0048)	3.0348 (0.0447)	3.0456 (0.0146)	3.0485 (0.0048)
	$\alpha=2,\beta=3,\gamma=2$			$\alpha=2,\beta=3,\gamma=3$		
30	2.2516 (0.1407)	2.2797 (0.0603)	2.3128 (0.0263)	3.2044 (0.2843)	3.2418 (0.0858)	3.2831 (0.0263)
70	2.1184 (0.059)	2.1316 (0.0274)	2.1449 (0.0128)	3.0960 (0.126)	3.1141 (0.0400)	3.1297 (0.0128)
100	2.0839 (0.0409)	2.0905 (0.0194)	2.1001 (0.0092)	3.0729 (0.089)	3.0801 (0.0285)	3.1018 (0.0092)
200	2.0405 (0.0202)	2.0483 (0.0098)	2.0524 (0.0048)	3.0366 (0.0447)	3.0444 (0.0146)	3.0484 (0.0048)
	$\alpha=5,\beta=2,\gamma=2$			$\alpha=5,\beta=2,\gamma=3$		
30	2.2587 (0.1416)	2.2800 (0.0604)	2.3179 (0.0263)	3.1931 (0.282)	3.2336 (0.0856)	3.2768 (0.0263)
70	2.1164 (0.0589)	2.1327 (0.0274)	2.1455 (0.0128)	3.1005 (0.1264)	3.1232 (0.0401)	3.1381 (0.0128)
100	2.0862 (0.041)	2.0956 (0.0194)	2.1051 (0.0092)	3.0718 (0.0889)	3.0824 (0.0286)	3.0950 (0.0092)
200	2.0415 (0.0202)	2.0474 (0.0098)	2.05211 (0.0048)	3.0365 (0.0447)	3.0405 (0.0146)	3.0534 (0.0048)
	$\alpha=5,\beta=3,\gamma=2$			$\alpha=5,\beta=3,\gamma=3$		
30	2.2589 (0.1417)	2.2830 (0.0604)	2.3075 (0.0263)	3.2012 (0.283)	3.2539 (0.0862)	3.2795 (0.0263)
70	2.1159 (0.0588)	2.1342 (0.0274)	2.1423 (0.0128)	3.0994 (0.1263)	3.1147 (0.0400)	3.1329 (0.0128)
100	2.0831 (0.0409)	2.0928 (0.0194)	2.1020 (0.0092)	3.0687 (0.0888)	3.0826 (0.0286)	3.0943 (0.0092)
200	2.0469 (0.0203)	2.0472 (0.0098)	2.051 (0.0048)	3.0348 (0.0447)	3.0468 (0.0146)	3.0498 (0.0048)

5. COMMENTS AND CONCLUSION

We concluded from this study, by increasing sample size Bayes posterior risk decreases and Bayes estimator approaches to its true value of the parameter. By increasing the value of parameter Bayes posterior risk also increases. After the simulation and comparing different loss functions, the results shows that DELF is performing better because its Bayes posterior risk is smaller than other loss functions. Gamma Prior gives the best estimate or as compare to Exponential .Gamma prior with DELF is performing better among others .Gamma prior is found to be the best prior for the estimation. The work has been done by using only two informative priors with three loss functions others can extend their study by using other prior and loss functions. We have used only complete samples techniques others can used censored samples or other sampling schemes.

REFERENCES

1. De Gusmao, F.R., Ortega, E.M. and Cordeiro, G.M. (2011). The generalized inverse Weibull distribution. *Statistical Papers*, 52(3), 591-619.
2. Aslam, M. (2003). An application of prior predictive distribution to elicit the prior density. *Journal of Statistical Theory and Applications*, 2(1), 70-83.
3. Berger, J.O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.
4. Sindhu, T.N., Feroze, N. and Aslam, M. (2013). Bayesian analysis of the Kumaraswamy distribution under failure censoring sampling scheme. International *Journal of Advanced Science and Technology*, 51, 39-58.
5. Feroze, N. and Yaseen, W. (2015). Left censored nakagami distribution under bayesian framework. *Science International*, 27(2), 945-948.