



# BIOSTATISTICS

*For Health Students  
With Manual on Software Applications*

*(Second Edition)*

*Muhammad Hanif*

*Munir Ahmad*

*Ezz H. Abdelfattah*



An ISOSS Publication  
ISLAMIC COUNTRIES SOCIETY OF STATISTICAL SCIENCES  
Lahore, Pakistan.

# **BIOSTATISTICS**

**For Health Students**

**With Manual on Software Applications**

**Muhammad Hanif**

Ph.D.

**Munir Ahmad**

Ph.D.

**National College of Business Administration & Economics  
Lahore, Pakistan**

and

**Ezz H. Abdelfattah**

Ph.D.

**Statistics Department, Faculty of Science  
King Abdul Aziz University  
21589, Jeddah 80203, Saudi Arabia**



**An ISOSS Publication**

**ISLAMIC COUNTRIES SOCIETY OF STATISTICAL SCIENCES  
Lahore, Pakistan.**

© 2001, 2014 Islamic Countries Society of Statistical Sciences.

Pakistan Science Foundation, Islamabad, Pakistan has financed the publication of this book and as such this book or any part thereof must not be reproduced or retrieved in any form without prior permission of authors and the Pakistan Science Foundation.

ISBN: 969-8858-008

Muhammad Hanif and Munir Ahmad

BIOSTATISTICS:  
QA 574.015 HAN-B

Edition: First

Impression: 1000

Printed in Pakistan

Printers: Taya Sons Printer, Rattigon Road, Lahore, Pakistan

Publishers: Islamic Countries Society of Statistical Sciences

Email: [drmianhanif@gmail.com](mailto:drmianhanif@gmail.com); [drmunir@brain.net.pk](mailto:drmunir@brain.net.pk)

## بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ

وَلَا تَقُولَنَّ لِشَیْءٍ اِنِّیْ فَاعِلٌ ذٰلِکَ عَدَاۗءَ  
اِلٰہِ اَنْ یَّشَآءَ اللّٰهُ ۗ وَادْکُرْ رَبَّکَ اِذَا نَسِیْتَ وَقُلْ  
عَسَیْ اَنْ یَّہْدِیْنِ رَبِّیْۤیْ لَا قُرْبَ مِنْ هٰذَا اَرْشٰدًا ﴿۲۳﴾  
سُوْرَةُ الْکٰهِفِ (۲۳-۲۴)

اور نہ کہنا کسی کام کو کہ میں یہ کل کرونگا۔ مگر یہ کہ اللہ چاہے تو  
اور یاد کر لے اپنے رب کو جب بھول جائے اور کہہ دو کہ امید  
ہے میرا رب مجھ کو اس سے زیادہ نزدیک نیکی کی راہ دکھلائے۔  
سورة الكهف (23-24)

No say anything “I shall be sure to do so and so tomorrow”, except “if ALLAH so wills”⊗ And remember your Lord when you forget [it] and say, "Perhaps my Lord will guide me to what is nearer than this to right conduct.⊗

*Surat Al-Kahf (23-24)*

## FOREWORD

When I was a doctorate student at Johns Hopkins School of Public Health. I used to take Biostatistics as a course, which I have to accept and live with it. I did not have much of a problem with it, but I could have enjoyed it more if it were presented to me in more attractive way. I mean in relation to real life rather than abstracts of figures. With this innovative writing of Prof. Hanif and Prof. Ahmad, I can see that the science of numbers and ratios is being wisely integrated with epidemiology.

Through feedback from the learners, I am sure that more will be added to this healthy relation between Biostatistics and other medical and public health sciences.

**Prof. Zohair Sebai**  
**Saudi Arabia**

## مقدمه

عندما كن اطلب العلم فى كلية الصحة العامة بجامعة جونز هو بكنزبامريكا، كنت مضطر الى ان اتقبل علم الاحصاء كمادة تفيله تتحدث عن الارقام بلهجة جافة، ولم يكن لدى خيار فى ذلك. اى نعم لم اكن اجد صعوبة فى دراستها ولكن كم تمنيت ان تعطى لى بشكل افضل، اى ان يكون فيها حيلة اكثر مع هذه الكتابة المبدعة من البر وفسور حنيف والبر وفسور احمد اجدان مادة الاحصاء اكتسبت حياة بفضل و صلاحها بعلم آلابيد ميولوجى. وانلواثق من ان ربود الافعال من الدار سين سوف تضىف عليها حياة اكثر و تجعلها اكثر صلة بعلوم الطلب والصحة العامة.

الاستاذ زهير اسباعى

## **PREFACE TO SECONED EDITION**

In this Edition the analysis of statistical data have been done on the basis of IBM 22 SPSS Package. In logistic regression (Chapter 9) basic concept with analysis of ordinal logistic regression and multinomial logistic regression have been added. A new Chapter of survival analysis is included as Chapter 10. The previous Chapter 10 (Reliability Coefficient) from the old addition is now Chapter 11. We are thankful to Dr. Nadeem Shafique Butt of COMSATS Institute of Information Technology, Lahore for the addition of new material in this Edition. We are also thankful to Mr. M. Imtiaz and M. Iftikhar of Islamic Countries Society of Statistical Sciences (ISOSS) for excellent typesetting of this book.

Muhammad Hanif  
Munir Ahmad  
Ezz H. Abdelfattah

## PREFACE

The use of statistical techniques of data analysis has been observed to have dramatically increased recently, particularly for application in the biomedical and social sciences. This may be partially attributed to the developments during the last few decades of sophisticated methods for analyzing quantitative and categorical data. It also reflects the increasing methodological sophistication of scientists and applied statisticians. The Islamic Educational Scientific and Cultural Organization (ISESCO) realized that the knowledge of these statistical methods in health and medical research as well as in clinical practice was very important for dealing with uncertainty in diagnosis, treatment and prognosis. Moreover these methods are useful for health professionals, since they have to evaluate their day-to-day clinical data and research material. Such statistical analyses could improve their understanding and skills for treatment of patients, as well as planning, implementation and evaluation of health programs. Considering all these reasons, ISESCO formed a committee headed by Dr. Munir Ahmad in 1993 to develop a curriculum regarding Bio-statistics for medical colleges in the Islamic Countries. The senior author was also member of this committee. The curriculum was developed and circulated among the medical colleges of the Islamic Countries. Most of the Islamic Countries sent their comments and suggestions, which were incorporated in the curriculum before approval. Then we decided to write this manual for the medical, health and social sciences students. This is a self-reading manual written in a simple language, which can easily be comprehended and could be of use for health related and social studies, both at the undergraduate and postgraduate levels.

This manual consists of 10 chapters and presents the most important methods for analyzing quantitative and categorical data. It summarizes methods that have long played a prominent role, such as parametric and non-parametric tests; linear regression, chi square tests and measures of association including the tests of significance of relative risk, odds ratio and Mental-Haenszel odds ratio. A chapter on various types of sampling techniques and estimation of sample size has been added which is normally not included in common books on Bio-statistics. Various methods of reliability co-efficient with applications have been put together to facilitate the research workers. This manual puts special emphasis on logistic regression, a newly developed technique for qualitative data analysis. Another feature of this manual is that one can easily understand and use SPSS (Statistical Package for Social Sciences) software. Much emphasis has been given to the ability to select an appropriate test for the analysis of data with medical interpretation in the context of the problem.

The technical components of the manual have been explained in a way that does not require familiarity with mathematics such as calculus and matrix algebra. Examples relating to health problems have been solved using SPSS software. Permission has been taken for the examples and tables included in this manual.

In general most statistical methods require extensive computations. We have tried to avoid details of complex calculations, since software for data analyses are available. It is recommended for the users of this manual to use software, where possible, in solving the

problems. The data entry system has been explained either in the text or at the end of each chapter. However, for those who wish to solve problems manually, all the steps have been clearly demonstrated. At the end of each chapter the applications of SPSS software have been demonstrated in details.

We are deeply grateful to Prof. Zohair Al-Sebai Ex-Professor of Family and Community Medicine King Faisal University Dammam for providing full facilities to write this manual. We are also thankful to Dr. Nabil Yasin Kurashi, Dr. Adnan Al-Bar, Dr. Abdullah Mangood, Dr. Kasim Al-Dwood, Dr. Sameeh Al-Maie and Post-Graduates students of the Department of Family and Community Medicine, King Faisal University, Dammam, Saudi Arabia for encouraging us to write this manual. In this respect we also appreciate with gratitude to the National College of Business Administration and Economics for providing for administrative work.

We particularly appreciate the efforts of Dr. M. Samiuddin, Ex. Professor of King Abdul Aziz University, Jeddah, who read the manuscript critically and suggested useful changes to improve the text of the manual. We express our gratitude to Prof. Akhlaq Ahmad of Islamic Countries Society of Statistical Sciences (ISOSS), Lahore for reading the first and final draft of the manuscript and suggesting useful changes in the text and to Prof. M. Afzal, Ex-Joint Director, PIDE, Islamabad for critically reviewing the book.

Last but not the least, we are indebted to Mr. Mohammad Junaid, of King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, for composing the manuscript.

We would like to thank Mr. Muhammad Iftikhar and Mr. Muhammad Imtiaz of Islamic Countries Society of Statistical Sciences (ISOSS) for assistance in adjusting the corrections in the manuscript.

**Muhammad Hanif  
Munir Ahmad**



# *Contents*

## **Foreword**

## **Preface**

|  |          |
|--|----------|
| <b>Chapter 1: Basic Concept and Data Presentation</b>          | <b>1</b> |
| 1.1 Introduction   | 1        |
| 1.1.1 Population versus Sample                                 | 2        |
| 1.1.2 Parameter versus Statistic                               | 2        |
| 1.1.3 Descriptive versus Inferential Statistics                | 3        |
| 1.1.4 Descriptive versus Analytic Studies                      | 3        |
| 1.1.5 Cohort Study   | 3        |
| 1.1.6 Case versus Control study                                | 4        |
| 1.1.7 Experimental study                                       | 5        |
| 1.1.8 Intervention Studies                                     | 5        |
| 1.2 Variable   | 5        |
| 1.2.1 Categorical Variable                                     | 5        |
| 1.2.2 Numerical Variable                                       | 6        |
| 1.2.3 Dependent and Independent Variables                      | 6        |
| 1.3 Measurement Scales   | 7        |
| 1.3.1 Qualitative Scale  | 7        |
| 1.3.2 Numerical Scale  | 7        |
| 1.4 Types of Statistical Data                                  | 8        |
| 1.4.1 Qualitative Data   | 8        |
| 1.4.2 Quantitative Data  | 9        |
| 1.5 Graphical Presentation of Qualitative Data                 | 16       |
| 1.5.1 Bar Charts   | 16       |
| 1.5.2 Subdivided and Multiple Bar Charts                       | 19       |
| 1.5.3 Pie Chart  | 22       |
| 1.6 Summarization of Quantitative Data                         | 25       |
| 1.6.1 Frequency Table and Frequency Distribution               | 25       |
| 1.6.2 Relative Frequency                                       | 27       |
| 1.6.3 Cumulative Frequency                                     | 27       |
| 1.6.4 Relative Cumulative Frequency                            | 28       |
| 1.7 Graphical Presentation of Quantitative Data                | 28       |
| 1.7.1 Histogram  | 31       |
| 1.7.2 Frequency Polygon and Frequency Curve                    | 31       |
| 1.7.3 Types of Frequency Curve                                 | 35       |
| 1.7.4 Cumulative Frequency Curve                               | 35       |
| 1.8 Histogram: Graphical Presentation of Data Relating to Time | 36       |
| 1.9 Descriptive Statistics                                     | 40       |
| 1.9.1 Rates  | 40       |
| 1.9.2 Ratios   | 42       |
| 1.9.3 Odds Ratio   | 43       |
| 1.9.4 Measures of Central Tendency                             | 48       |

|   |            |
|---|------------|
| 1.9.5 Measures of Dispersion  | 50         |
| 1.9.6 Relative Measure  | 53         |
| 1.10 Mean $\pm k \times$ Standard Deviation                                   | 60         |
| <b>Chapter 2: Basic Concepts of Probability and Probability Distributions</b> | <b>61</b>  |
| 2.1 Introduction  | 61         |
| 2.2 Definition and rules of probability                                       | 62         |
| 2.2.1 Additive Rule of Probability for Mutually Exclusive Events              | 63         |
| 2.2.2 Independent Events and Multiplicative Rule of Probability               | 63         |
| 2.2.3 Additive Rule for non Mutually Exclusive Events                         | 64         |
| 2.2.4 Conditional Probability   | 64         |
| 2.2.5 Rule of multiplication for non-independent events                       | 65         |
| 2.2.6 Properties of Probability   | 65         |
| 2.3 Probability distribution  | 67         |
| 2.3.1 The Binomial Probability Distribution                                   | 67         |
| 2.3.2 The Poisson Probability Distribution                                    | 71         |
| 2.3.3 The Normal Probability Distribution                                     | 72         |
| <b>Chapter 3: Sampling Procedures and Sample Size Estimation</b>              | <b>83</b>  |
| 3.1 Introduction  | 83         |
| 3.2 Types of Sampling   | 84         |
| 3.2.1 Probability Sampling  | 84         |
| 3.2.2 Non-Probability Sampling  | 84         |
| 3.3 Some Commonly Used Selection Procedures                                   | 84         |
| 3.3.1 Simple Random Sampling  | 84         |
| 3.3.2 Estimation of mean and variance for sample mean and sample proportion   | 86         |
| 3.3.3 Estimation of Sample Size   | 88         |
| 3.3.4 Standard Deviation and Standard Error                                   | 92         |
| 3.3.5 Confidence Limits   | 93         |
| 3.3.6 Stratified Random Sampling  | 102        |
| 3.3.7 Sytematic Sampling  | 104        |
| 3.3.8 Single Stage Cluster Sampling   | 105        |
| 3.3.9 Probability Proportional to Size Sampling Procedure                     | 106        |
| 3.3.10 Random Systematic Selection Procedure                                  | 108        |
| 3.3.11 Multistage Sampling  | 109        |
| <b>Chapter 4: Hypothesis Testing Procedures</b>                               | <b>115</b> |
| 4.1 Introduction  | 115        |
| 4.1.1 Hypothesis or a Statistical Hypothesis                                  | 116        |
| 4.1.2 One-tail and Two-tail Test  | 117        |
| 4.1.3 Level of Significance ( $\alpha$ )                                      | 118        |
| 4.1.4 Confidence Level ( $1 - \alpha$ )                                       | 118        |
| 4.1.5 A Critical Value  | 118        |
| 4.1.6 Test Statistic  | 119        |
| 4.1.7 Type I and Type II Errors   | 119        |

|  |     |
|--|-----|
| 4.2 Estimation of Sample size when Probability of Type I Error and Power of the test are known | 125 |
| 4.2.1 Sample size for comparing proportions  | 125 |
| 4.2.2 Sample size for a single mean  | 127 |
| 4.2.3 Sample size for Comparing of two proportions   | 128 |
| 4.3 Diagnosing a Test-Statistic for Testing of Hypotheses and p-Value                          | 128 |
| 4.3.1 Diagnosing a Test-Statistic  | 128 |
| 4.3.2 p -Value   | 129 |
| 4.4 General Procedure of Testing of Hypothesis   | 131 |
| 4.5 Tests of Significance  | 132 |
| 4.5.1 Z-Test for one and two samples for means and proportions                                 | 136 |
| 4.5.2 t-test for single and two samples  | 142 |
| 4.5.3 Application of SPSS package  | 151 |
| 4.5.4 t-test for Paired Observations   | 156 |
| 4.6 Testing a Population Variance for Single Samples   | 164 |
| 4.7 Testing the Ratio of Two Population Variances  | 167 |
| <br>   |     |
| <b>Chapter 5: Analysis of Variance</b>   | 177 |
| 5.1 Introduction   | 177 |
| 5.2 Analysis of Variance With One- Way classification  | 177 |
| 5.3 Analysis of variance for two-Way classification  | 190 |
| 5.4 Repeated Measure Design or Repeated Measure Analysis of Variance                           | 198 |
| 5.5 Multivariate Analysis of Variance (MANOVA)   | 206 |
| 5.6 Simple Factorial Experiment  | 213 |
| 5.7 “n of 1 Trials”: Controlled Trials in Single Subjects                                      | 220 |
| 5.7.1 Statistics in “n of 1 trials”  | 221 |
| 5.7.2 Use of Analysis of Variance for “n of 1 trials”  | 223 |
| <br>   |     |
| <b>Chapter 6: Regression and Correlation</b>   | 227 |
| 6.1 Introduction   | 227 |
| 6.2 Simple Linear Regression Analysis  | 227 |
| 6.2.1 Method of Least Squares  | 230 |
| 6.2.2 Some Applications of Simple Regression   | 231 |
| 6.3 The Coefficient of Correlation   | 243 |
| 6.4 Regression Model for Prediction  | 247 |
| 6.5 Multiple Regression Analysis   | 250 |
| 6.5.1 Applications of multiple-regression  | 250 |
| 6.5.2 Fitting the model and interpretation of coefficients                                     | 251 |
| 6.6 Partial Correlation  | 277 |
| 6.7 Intra-Class Correlation Coefficient  | 279 |
| <br>   |     |
| <b>Chapter 7: Analysis of Categorical Data</b>   | 287 |
| 7.1 Introduction   | 287 |
| 7.2 Assumptions  | 288 |
| 7.3 Uses of Chi-Square Test  | 289 |
| 7.4 Independence and Homogeneity   | 290 |

|  |            |
|--|------------|
| 7.4.1 2x2 Contingency Table  | 290        |
| 7.4.2 Phi Coefficient  | 292        |
| 7.4.3 Contingency coefficient (C)  | 292        |
| 7.4.4 Cramer's-V (V)   | 293        |
| 7.4.5 Adjusted Chi-square (Yates' Correction)                            | 293        |
| 7.4.6 Fisher's exact test  | 298        |
| 7.4.7 R x C contingency table  | 300        |
| 7.4.8 Application of Kendall's Tau $b(\tau_b)$                           | 303        |
| 7.4.9 2 x 2 x K Tables (Meta Analysis)                                   | 308        |
| 7.5 Matched Samples (McNemar test)                                       | 316        |
| 7.5.1 Layout of Tests of Significance                                    | 317        |
| 7.6 Mantel-Haenszel Test for Linear Association                          | 322        |
| 7.7 Testing the Statistical Significance of Relative Risk and Odds Ratio | 327        |
| 7.7.1 Relative Risk (RR) Estimate  | 327        |
| 7.7.2 Odds ratio   | 328        |
| 7.7.3 Attributable risk (Risk difference, Rate difference)               | 328        |
| 7.7.4 Relative risk of matched-pairs                                     | 334        |
| 7.7.5 Odds ratio and tests of significance                               | 335        |
| 7.7.6 Matched analysis in case-control study                             | 338        |
| 7.8 Relation between odds ratio and relative risk                        | 339        |
| 7.9 Mantel-Haenszel Procedure for Relative Risk and Odds Ratio           | 341        |
| 7.9.1 Mantel-Haenszel relative risk                                      | 346        |
| 7.9.2 Mantel-Haenszel chi-square   | 346        |
| 7.9.2 Mantel-Haenszel chi-square   | 348        |
| 7.10 Sensitivity, Specificity and Kappa-Statistic                        | 354        |
| 7.10.1 Screening test  | 354        |
| 7.10.2 Validity of a screening test                                      | 354        |
| 7.10.3 Diagnostic Tests (Sensitivity and Specificity)                    | 358        |
| 7.10.4 Kappa (Cohen's Kappa)-Statistic                                   | 359        |
| <b>Chapter 8: Non-Parametric Tests</b>                                   | <b>369</b> |
| 8.1 Introduction   | 369        |
| 8.2 The Sign Test  | 371        |
| 8.2.1 The Sign test for a single sample                                  | 371        |
| 8.2.2 The Sign test for samples of paired observation                    | 377        |
| 8.3 The Wilcoxon Signed-rank test  | 382        |
| 8.4 Test for Two Independent Samples                                     | 386        |
| 8.4.1 The median test  | 386        |
| 8.4.2 The Mann-Whitney and Wilcoxon Rank sum-W tests                     | 391        |
| 8.5 Test for K-Independent Samples                                       | 397        |
| 8.5.1 The Kruskal-Wallis test (or H-test)                                | 397        |
| 8.6 K-Related Samples  | 405        |
| 8.6.1 The Friedman test  | 405        |
| 8.6.2 Kendall's coefficient of concordance or W-statistic                | 409        |
| 8.6.3 Cochran's Q test   | 412        |
| 8.7 Measures of Rank Correlation   | 416        |

|  |     |
|--|-----|
| <b>Chapter 9: Logistic Regression</b>                      | 421 |
| 9.1 Introduction   | 421 |
| 9.2 Fitting of Simple Logistic Model                       | 423 |
| 9.2.1 Application of simple logistic model for prediction  | 427 |
| 9.2.2 Confidence limits for odds ratio                     | 429 |
| 9.3 The Multiple Logistic Model                            | 431 |
| 9.4 The Ordinal Regression                                 | 445 |
| 9.5 The Multinomial Logistic Regression                    | 451 |
| <br>   |     |
| <b>Chapter 10: Survival Analysis</b>                       | 473 |
| 10.1 Introduction  | 473 |
| 10.2 Survival analyses                                     | 474 |
| 10.3 Kaplan Meier  | 484 |
| 10.4 Cox – Regression                                      | 494 |
| <br>   |     |
| <b>Chapter 11: Reliability Coefficient</b>                 | 501 |
| 11.1 Introduction  | 501 |
| 11.2 Reliability of a Test                                 | 502 |
| 11.3 Different Forms of Measuring Reliability Coefficients | 504 |
| 11.3.1 Test-Retest Method                                  | 505 |
| 11.3.2 Split-half Method                                   | 509 |
| 11.3.3 Kuder-Richardson Formula-20                         | 513 |
| 11.3.4 Cronbach's Alpha ( $\alpha$ )                       | 516 |
| <br>   |     |
| Bibliography   | 519 |
| Index  | 522 |