

Second
Edition

Basics of Biostatistics



AP Kulkarni



CBS Dedicated to Education

CBS Publishers & Distributors Pvt Ltd



Basics of Biostatistics

Second Edition

Basics of Biostatistics

Second Edition

AP Kulkarni

BSc, MBBS, DPH, MD, PhD, FIAPSM

Former

Professor and Head
Department of Community Medicine
Government Medical College
Aurangabad, Maharashtra
Pro-Vice Chancellor
Maharashtra University of Health Sciences
Nashik, Maharashtra



CBS Publishers & Distributors Pvt Ltd

New Delhi • Bengaluru • Chennai • Kochi • Kolkata • Mumbai
Bhopal • Bhubaneswar • Hyderabad • Jharkhand • Nagpur • Patna • Pune • Uttarakhand • Dhaka (Bangladesh)

Disclaimer

Science and technology are constantly changing fields. New research and experience broaden the scope of information and knowledge. The authors have tried their best in giving information available to them while preparing the material for this book. Although, all efforts have been made to ensure optimum accuracy of the material, yet it is quite possible some errors might have been left uncorrected. The publisher, the printer and the authors will not be held responsible for any inadvertent errors, omissions or inaccuracies.

eISBN: 978-93-896-8818-4

Copyright © Authors and Publisher

Second eBook Edition: 2020

All rights reserved. No part of this eBook may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system without permission, in writing, from the authors and the publisher.

Published by Satish Kumar Jain and produced by Varun Jain for
CBS Publishers & Distributors Pvt. Ltd.

Corporate Office: 204 FIE, Industrial Area, Patparganj, New Delhi-110092

Ph: +91-11-49344934; Fax: +91-11-49344935; Website: www.cbspd.com; www.eduport-global.com;

E-mail: eresources@cbspd.com; marketing@eduport-global.com

Head Office: CBS PLAZA, 4819/XI Prahlad Street, 24 Ansari Road, Daryaganj, New Delhi-110002, India.

Ph: +91-11-23289259, 23266861, 23266867; Fax: 011-23243014; Website: www.cbspd.com;

E-mail: publishing@cbspd.com; eduportglobal@gmail.com.

Branches

-
- **Bengaluru:** Seema House 2975, 17th Cross, K.R. Road, Banasankari 2nd Stage, Bengaluru - 560070, Karnataka Ph: +91-80-26771678/79; Fax: +91-80-26771680; E-mail: bangalore@cbspd.com
 - **Chennai:** No.7, Subbaraya Street Shenoy Nagar Chennai - 600030, Tamil Nadu Ph: +91-44-26680620, 26681266; E-mail: chennai@cbspd.com
 - **Kochi:** 36/14 Kalluvilakam, Lissie Hospital Road, Kochi - 682018, Kerala Ph: +91-484-4059061-65; Fax: +91-484-4059065; E-mail: kochi@cbspd.com
 - **Mumbai:** 83-C, 1st floor, Dr. E. Moses Road, Worli, Mumbai - 400018, Maharashtra Ph: +91-22-24902340 - 41; Fax: +91-22-24902342; E-mail: mumbai@cbspd.com
 - **Kolkata:** No. 6/B, Ground Floor, Rameswar Shaw Road, Kolkata - 700014 Ph: +91-33-22891126 - 28; E-mail: kolkata@cbspd.com

Representatives

-
- **Hyderabad**
 - **Pune**
 - **Nagpur**
 - **Manipal**
 - **Vijayawada**
 - **Patna**

to

*my students
who have made
teaching of biostatistics
an enjoyable affair*

Basics of Biostatistics

This Page is Intentionally Left Blank

Preface to the Second Edition

It is a matter of immense satisfaction to come up with the second edition of this book. It is probably because of its uniqueness in ample use of MS-Excel[®]. I think that the readers would continue their support to this book.

AP Kulkarni

e-mail: drapkulkarni@gmail.com

Preface to the First Edition

It gives me an immense pleasure in bringing to you this book *Basics of Biostatistics*. I have been teaching biostatistics to undergraduate and postgraduate medical students for the last three-and-a-half decades. I have witnessed a sort of phobia about biostatistics not only in the students but also in the teaching faculty and hope to reduce this phobia through the introduction of this book.

A lot of changes have occurred in teaching methods in medicine in general, and biostatistics in particular, during the last few years. Introduction of computer software is one among them. Readers will find an ample use of MS Excel® in this book in getting solutions to complex statistical procedures. In fact, this is a unique feature of this book. The part on tests of significance is arranged in such a way that readers can relate these to their own research problems immediately. Obviously, it is for medical undergraduate and postgraduate students who are taking their initial steps in the subject.

I thank CBS Publishers & Distributors for taking initiative in publishing this book. Feedback from the readers would be highly appreciated.

AP Kulkarni

e-mail: drapkulkarni@gmail.com

Contents

<i>Preface to the Second Edition</i>	<i>vii</i>
<i>Preface to the First Edition</i>	<i>viii</i>
<i>Abbreviations</i>	<i>xii</i>
1. Introduction	1
1.1 Some Basic Terminologies Used	1
1.2 Types of Data	4
2. Data Collection	5
2.1 Mistakes and Errors	5
2.2 Recording of the Information	5
2.3 Storage	6
2.4 Pretesting	6
2.5 Data Collection Methods	6
3. Data Compilation	7
3.1 Data Editing	7
4. Introduction to MS Excel®	9
4.1 Preparing Customized Database with MS Excel	10
4.2 Data Validation with MS Excel	11
4.3 Statistical Functions in MS Excel	11
5. Data Presentation	13
5.1 Tabular Presentation	13
5.2 Graphical Presentation	14
6. Common Methods in MS Excel	21
7. Centering Constants	22
7.1 Arithmetic Mean (M)	22
7.2 Geometric Mean (GM)	23
7.3 Harmonic Mean	24
7.4 Median	24
7.5 Mode	25
8. Measures of Variation	26
8.1 Range	26
8.2 Interquartile Range	27
8.3 Average Deviation	28

8.4 Standard Deviation (SD)	29
8.5 Variance	30
8.6 Coefficient of Variation (CV)	30
8.7 Percentile	31
8.8 Box Plot	32
8.9 Skewness	32
9. Normal Distribution	34
9.1 The Standard Normal Curve	34
9.2 Applications of the Concept of Normal Curve	35
9.3 Kurtosis	37
10. Sampling Variation	38
10.1 Standard Error of Mean (SEM)	38
10.2 Standard Error of Proportion (SEP)	40
11. Probability	42
11.1 Addition Rule of Probability	42
11.2 Multiplication Rule of Probability	43
11.3 Conditional Rule of Probability	43
12. Null Hypothesis	45
12.1 Errors in Test of Significance	45
13. Introduction to Tests of Significance	47
14. Tests of Significance: Single Sample Situations	50
14.1 Chi-Square Test for Single Sample	50
14.2 One Sample T-Test	52
14.3 One Sample Z-Test	53
14.4 Wilcoxon Signed Rank Test for Single Sample	53
15. Tests of Significance: Two Unmatched Samples	55
15.1 Chi-Square as Test of Homogeneity of Two Unmatched Samples	55
15.2 Fischer's Exact Test	57
15.3 Test of Significance of Difference between Two Proportions (Z-Test)	58
15.4 Mann-Whitney Test	59
15.5 Test of Significance of Difference between Two Means (Z-Test)	61
15.6 Unpaired T-Test	62
16. Tests of Significance: Two Matched/Paired Samples	65
16.1 McNemar's Test	65
16.2 Wilcoxon Test for Paired Samples	66
16.3 Paired T-Test	67

17. Tests of Significance: Three or More Unmatched Samples	70
17.1 Chi-Square Test for three or More Groups	70
17.2 Kruskal-Wallis Test	72
17.3 One-way ANOVA	73
18. Tests of Significance: Three or More Matched Samples	76
18.1 Cochran's Q-Test	76
18.2 Repeated Measure ANOVA	77
18.3 Friedman's Test	80
19. Measuring Relationship between Two Variables	82
19.1 Phi (ϕ)	82
19.2 Contingency Coefficient (C)	83
19.3 Cramér's V	83
19.4 Pearson's Correlation Coefficient (r)	84
19.5 Spearman's Rank Order Correlation Coefficient (Rho, ρ)	86
20. Prediction of One Variable From Other	89
20.1 Simple Linear Regression	89
20.2 Some Alternatives to Simple Linear Regression	91
21. Prediction of One Variable from More Than One Variable	93
21.1 Multiple Linear Regression	93
21.2 Multiple Binary Logistic Regression	93
21.3 Multinomial Logistic Regression (MLR)	95
21.4 Ordinal Regression	95
22. Sample Size Determination for Common Situations	97
22.1 Introduction	97
22.2 Sample Size for Estimation of Population Proportion	99
22.3 Sample Size for Estimation of Population Mean	99
22.4 Sample Size for Unmatched Case Control Study	99
22.5 Sample Size for Clinical Trial, Superiority Design, Outcome: Nominal Variable	100
22.6 Sample Size for Difference in Proportions of Two Independent Samples	101
22.7 Sample Size for Clinical Trial, Superiority Design, Outcome Variable: Ratio or Interval Scale	102
22.8 Sample Size, Two Independent Samples, Variable Ratio or Interval Scale	102
22.9 Sample Size, Evaluation of New Diagnostic Test	103
22.10 Sample Size for Pearson's Correlation Coefficient (r)	103
22.11 Sample Size for Cohort Study	104
23. Sampling Methods	105
24. Supplement to Tests of Significance	110
<i>Appendix</i>	113
<i>Index</i>	123

Abbreviations

<i>Description</i>	<i>Abbreviation/Symbol</i>
Absolute precision	d
Alternate hypothesis, one-sided	AH ₁
Alternate hypothesis, two-sided	AH ₂
Arithmetic mean of population	M
Arithmetic mean of sample	m
Geometric mean	GM
Harmonic mean	x_h
Null hypothesis	NH
Population size	N
Relative precision	ϵ
Sample size	n
Standard deviation, population	S
Standard deviation, sample	s
Summation	Σ
Z-alpha	Z ₁
Z-beta	Z ₂