

EXHIBIT
6-1817
5-1-83 RATH
PENICAD Bayonne, N J

FUNDAMENTALS of

Biostatistics



Bernard Rosner

FOURTH
EDITION

2000N-1571

3 1/2" DOS data disk
enclosed
EXB 2

FOURTH EDITION

FUNDAMENTALS OF BIostatISTICS

BERNARD ROSNER
HARVARD UNIVERSITY



Duxbury Press

An Imprint of Wadsworth Publishing Company

ITP™

An International Thomson Publishing Company

Belmont • Albany • Bonn • Boston • Cincinnati • Detroit • London • Madrid • Melbourne
Mexico City • New York • Paris • San Francisco • Singapore • Tokyo • Toronto • Washington

**This book is dedicated to my wife Cynthia
and my children Sarah, David, and Laura.**

Statistics Editor: Alexander Kugushev
Editorial Assistant: Jennifer Burger
Production Editor: The Book Company
Print Buyer: Karen Hunt
Copy Editor: Linda Parrington
Technical Illustrator: TUCKERdesign
Cover: William Reuter
Compositor: Jonathan Peck Typographers
Printer: Banta Co., Harrisonburg, VA



*This book is printed on
acid-free recycled paper.*

COPYRIGHT © 1995
By Wadsworth Publishing Company
A Division of International Thomson Publishing Inc.

ITP™
The trademark ITP is used under license.

Printed in the United States of America

For more information, contact:

Wadsworth Publishing
Company
10 Davis Drive
Belmont, California 94002

International Thomson
Publishing
Berkshire House 168-173
High Holborn
London, WC1V7AA
England

Thomas Nelson Australia
102 Dodds Street
South Melbourne 3205
Victoria, Australia

Nelson Canada
1120 Birchmount Road
Scarborough, Ontario
Canada M1K 5G4

International Thomson
Publishing GmbH
Königwinterer Strasse 418
53227 Bonn
Germany

International Thomson
Publishing Asia
221 Henderson Road
#05-10
Singapore 0315

International Thomson
Publishing-Japan
Hirakawacho-cho Kyowa
Building, 3F
2-2-1 Hirakawacho-cho
Chiyoda-ku, 102 Tokyo
Japan

All rights reserved. No part of this work covered by the copyright hereon may be reproduced or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without the written permission of the publisher.

1 2 3 4 5 6 7 8 9 10—01 00 99 98 97 96 95

Library of Congress Cataloging-in-Publication Data

Rosner, Bernard (Bernard A.)

Fundamentals of biostatistics / Bernard Rosner,—4th ed.

p. cm.

Includes bibliographical references and index.

ISBN 0-531-20840-8

1. Biometry. 2. Medical statistics. I. Title

QH323.5.R674 1994

574'.01'5185—dc20

94-13154

CIP

CONTENTS

CHAPTER 1 General Overview 1

- Reference, 4

CHAPTER 2 Descriptive Statistics 5

- 2.1 Introduction, 5
- 2.2 Measures of Central Location, 6
- 2.3 Some Properties of the Arithmetic Mean, 14
- 2.4 Measures of Spread, 15
- 2.5 Some Properties of the Variance and Standard Deviation, 21
- 2.6 The Coefficient of Variation, 23
- 2.7 Grouped Data, 24
- 2.8 Graphic Methods for Grouped Data, 29
- *2.9 Case Study: Effects of Lead Exposure on Neurological and Psychological Function in Children, 35
- 2.10 Summary, 37
 - Problems, 38
 - References, 42

CHAPTER 3 Probability 43

- 3.1 Introduction, 43
- 3.2 Definition of Probability, 43
- 3.3 Some Useful Probabilistic Notation, 45
- *3.4 The Multiplication Law of Probability, 47
- 3.5 The Addition Law of Probability, 49
- 3.6 Conditional Probability, 52
- 3.7 Bayes' Rule and Screening Tests, 56
- 3.8 Prevalence and Incidence, 61
- 3.9 Summary, 61
 - Problems, 62
 - References, 69

CHAPTER 4 Discrete Probability Distributions 71

- 4.1 Introduction, 71
- 4.2 Random Variables, 72
- 4.3 The Probability Mass Function for a Discrete Random Variable, 72
- 4.4 The Expected Value of a Discrete Random Variable, 74
- 4.5 The Variance of a Discrete Random Variable, 76
- 4.6 The Cumulative-Distribution Function of a Discrete Random Variable, 78
- 4.7 Permutations and Combinations, 79
- 4.8 The Binomial Distribution, 82

*asterisks indicate new section or subsection for the Fourth Edition

viii CONTENTS

- 4.9 Expected Value and Variance of the Binomial Distribution, 87
- 4.10 The Poisson Distribution, 88
- 4.11 Computation of Poisson Probabilities, 92
- 4.12 Expected Value and Variance of the Poisson Distribution, 94
- 4.13 Poisson Approximation to the Binomial Distribution, 95
- 4.14 Summary, 97
 - Problems, 97
 - References, 103

CHAPTER 5 Continuous Probability Distributions 105

- 5.1 Introduction, 105
- 5.2 General Concepts, 105
- 5.3 The Normal Distribution, 107
- 5.4 Properties of the Standard Normal Distribution, 110
- 5.5 Conversion from an $N(\mu, \sigma^2)$ Distribution to an $N(0, 1)$ Distribution, 115
- 5.6 Linear Combinations of Random Variables, 119
- 5.7 Normal Approximation to the Binomial Distribution, 121
- 5.8 Normal Approximation to the Poisson Distribution, 125
- 5.9 Summary, 130
 - Problems, 133
 - References, 140

CHAPTER 6 Estimation 141

- 6.1 Introduction, 141
- 6.2 The Relationship Between Population and Sample, 142
- 6.3 Random-Number Tables, 143
- *6.4 Randomized Clinical Trials, 147
- 6.5 Estimation of the Mean of a Distribution, 151
- 6.6 Estimation of the Variance of a Distribution, 168
- 6.7 Estimation for the Binomial Distribution, 173
- *6.8 Estimation for the Poisson Distribution, 178
- 6.9 One-Sided Confidence Intervals, 182
- 6.10 Summary, 184
 - Problems, 185
 - References, 190

CHAPTER 7 Hypothesis Testing: One-Sample Inference 191

- 7.1 Introduction, 191
- 7.2 General Concepts, 192
- 7.3 One-Sample Test for the Mean of a Normal Distribution with Known Variance: One-Sided Alternatives, 194
- 7.4 One-Sample Test for the Mean of a Normal Distribution with Known Variance: Two-Sided Alternatives, 203
- 7.5 One-Sample t Test, 207
- 7.6 The Power of a Test, 212
- 7.7 Sample-Size Determination, 219
- 7.8 The Relationship Between Hypothesis Testing and Confidence Intervals, 225
- 7.9 One-Sample χ^2 Test for the Variance of a Normal Distribution, 228

- 7.10 One-Sample Test for a Binomial Proportion, 231
- *7.11 One-Sample Inference for the Poisson Distribution, 237
- 7.12 Summary, 243
- Problems, 245
 - References, 250

CHAPTER 8 **Hypothesis Testing: Two-Sample Inference 251**

- 8.1 Introduction, 251
- 8.2 The Paired t Test, 253
- 8.3 Interval Estimation for the Comparison of Means from Two Paired Samples, 256
- 8.4 Two-Sample t Test for Independent Samples with Equal Variances, 257
- 8.5 Interval Estimation for the Comparison of Means from Two Independent Samples (Equal Variance Case), 261
- 8.6 Testing for the Equality of Two Variances, 263
- 8.7 Two-Sample t Test for Independent Samples with Unequal Variances, 270
- *8.8 Case Study: Effects of Lead Exposure on Neurological and Psychological Function in Children, 276
- *8.9 The Treatment of Outliers, 277
- 8.10 Estimation of Sample Size and Power for Comparing Two Means, 283
- 8.11 Summary, 285
- Problems, 286
 - References, 297

CHAPTER 9 **Multisample Inference 299**

- 9.1 Introduction to the One-Way Analysis of Variance, 299
- 9.2 One-Way Analysis of Variance—Fixed-Effects Model, 299
- 9.3 Hypothesis Testing in One-Way ANOVA—Fixed-Effects Model, 301
- 9.4 Comparisons of Specific Groups in One-Way ANOVA, 306
- *9.5 Case Study: Effects of Lead Exposure on Neurological and Psychological Function in Children, 319
- *9.6 One-Way ANOVA—The Random-Effects Model, 322
- *9.7 The Cross-Over Design, 329
- 9.8 Summary, 337
- Problems, 337
 - References, 343

CHAPTER 10 **Hypothesis Testing: Categorical Data 345**

- 10.1 Introduction, 345
- 10.2 Two-Sample Test for Binomial Proportions, 346
- *10.3 Study Design, 359
- *10.4 Measures of Effect for Categorical Data, 361
- 10.5 Fisher's Exact Test, 370
- 10.6 Two-Sample Test for Binomial Proportions for Matched-Pair Data (McNemar's Test), 377

X CONTENTS

- 10.7** Estimation of Sample Size and Power for Comparing Two Binomial Proportions, 383
- 10.8** $R \times C$ Contingency Tables, 392
- *10.9** Confounding and Standardization, 399
- 10.10** Methods of Inference for Stratified Categorical Data—The Mantel-Haenszel Test, 404
- *10.11** Power and Sample-Size Estimation for Stratified Categorical Data, 417
- 10.12** Chi-Square Goodness-of-Fit Test, 419
- 10.13** The Kappa Statistic, 423
- 10.14** Summary, 427
 - Problems, 428
 - References, 440

CHAPTER 11 Regression and Correlation Methods 443

- 11.1** Introduction, 443
- 11.2** General Concepts, 444
- 11.3** Fitting Regression Lines—The Method of Least Squares, 447
- 11.4** Inferences About Parameters from Regression Lines, 452
- 11.5** Interval Estimation for Linear Regression, 462
- *11.6** Assessing the Goodness of Fit of Regression Lines, 466
- 11.7** Multiple Regression, 469
- *11.8** Case Study: Effects of Lead Exposure on Neurological and Psychological Function in Children, 483
- 11.9** Two-Way Analysis of Variance, 496
- 11.10** The Correlation Coefficient, 503
- 11.11** Statistical Inference for Correlation Coefficients, 506
- *11.12** Partial and Multiple Correlation, 516
- *11.13** The Intraclass Correlation Coefficient, 517
- 11.14** Multiple Logistic Regression, 521
- 11.15** Summary, 540
 - Problems, 541
 - References, 549

CHAPTER 12 Nonparametric Methods 551

- 12.1** Introduction, 551
- 12.2** The Sign Test, 553
- 12.3** The Wilcoxon Signed-Rank Test, 558
- 12.4** The Wilcoxon Rank-Sum Test, 562
- 12.5** The Kruskal-Wallis Test, 569
- 12.6** Rank Correlation, 575
- 12.7** Summary, 579
 - Problems, 580
 - References, 584

CHAPTER 13 Hypothesis Testing: Person-Time Data 585

- *13.1** Measures of Effect for Person-Time Data, 585
- 13.2** Two-Sample Inference for Incidence-Rate Data, 587
- *13.3** Inference for Stratified Person-Time Data, 594
- *13.4** Power and Sample-Size Estimation for Person-Time Data, 601

- *13.5 Testing for Trend-Incidence-Rate Data, 604
 13.6 Introduction to Survival Analysis, 607
 *13.7 Estimation of Survival Curves: The Kaplan-Meier Estimator, 609
 13.8 The Log-Rank Test, 614
 13.9 The Proportional-Hazards Model, 621
 13.10 Summary, 628
 ■ Problems, 628
 ■ References, 631

APPENDIX **Tables 633**

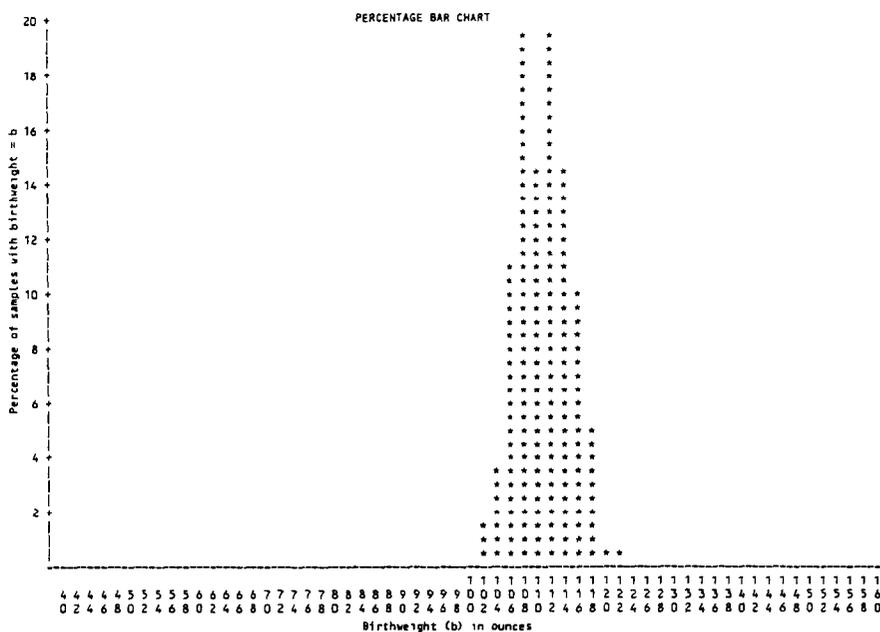
- 1** Exact Binomial Probabilities
 $Pr(X = k) = \binom{n}{k} p^k q^{n-k}$, 635
2 Exact Poisson Probabilities
 $Pr(X = k) = \frac{e^{-\mu} \mu^k}{k!}$, 640
3 The Normal Distribution, 643
4 Table of 1000 Random Digits, 648
5 Percentage Points of the t Distribution ($t_{d,u}$), 649
6 Percentage Points of the Chi-Square Distribution ($\chi^2_{d,u}$), 650
7a Exact Two-Sided 100% ($1 - \alpha$) Confidence Limits for Binomial Proportions ($\alpha = .05$), 651
7b Exact Two-Sided 100% ($1 - \alpha$) Confidence Limits for Binomial Proportions ($\alpha = .01$), 652
***8** Confidence Limits for the Expectation of a Poisson Random Variable (μ), 653
9 Percentage Points of the F Distribution ($F_{d_1, d_2, p}$), 654
***10** Critical Values for the ESD (Extreme Studentized Deviate) Outlier Statistic ($ESD_{1-\alpha}$, $\alpha = .05, .01$), 656
11 Fisher's z Transformation, 657
12 Two-Tailed Critical Values for the Wilcoxon Signed-Rank Test, 657
13 Two-Tailed Critical Values for the Wilcoxon Rank-Sum Test, 658
14 Critical Values for the Kruskal-Wallis Test Statistic (H) for Selected Sample Sizes for $k = 3$, 660
15 Two-Tailed Upper Critical Values for the Spearman Rank-Correlation Coefficient (r_s), 661

Answers to Selected Problems 663

Flowchart for Appropriate Methods of Statistical Inference 669

Index of Data Sets 676

Index 677



(c) $n = 30$

To use this method, a good estimate of basal body temperature during a period when ovulation is definitely not occurring is needed. Suppose that for this purpose a woman measures her body temperature on awakening on the first 10 days after menstruation and obtains the following data: 97.2°, 96.8°, 97.4°, 97.4°, 97.3°, 97.0°, 97.1°, 97.3°, 97.2°, 97.3°. What is the best estimate of her underlying basal body temperature (μ)? How accurate is this estimate?

SOLUTION The best estimate of her underlying body temperature during the nonovulation period (μ) is given by

$$\bar{x} = (97.2 + 96.8 + \dots + 97.3)/10 = 97.20^\circ$$

The standard error of this estimate is given by

$$s/\sqrt{10} = 0.189/\sqrt{10} = 0.06^\circ$$

In our work on confidence intervals in Section 6.5.6 we will show that for many underlying distributions, we can be fairly certain that the true mean μ is approximately within two standard errors of \bar{x} . In this case, true mean basal body temperature (μ) is within $97.20^\circ \pm 2(0.06)^\circ \approx (97.1^\circ - 97.3^\circ)$. Thus, if the temperature is elevated by at least 0.5° above this range on a given day, then it might indicate that the woman was ovulating, and for contraceptive purposes, intercourse should not be attempted on that day. ■■■

65.3 **Central-Limit Theorem**

If the underlying distribution is normal, then it can be shown that the sample mean will itself be normally distributed with mean μ and variance σ^2/n (see Section 5.6). In other words, $\bar{x} \sim N(\mu, \sigma^2/n)$. If the underlying distribution is *not* normal, we would still like to make some statement about the sampling distribution of the sample mean. This statement is given by the following theorem:

6.3 Central-Limit Theorem

Let x_1, \dots, x_n be a random sample from some population with mean μ and variance σ^2 . Then for large n , $\bar{x} \sim N(\mu, \sigma^2/n)$ even if the underlying distribution of individual observations in the population is not normal. (The symbol \sim is used to represent "approximately distributed.")

This theorem is very important because many of the distributions encountered in practice are not normal. In such cases the central-limit theorem can often be applied; this will allow us to perform statistical inference based on the approximate normality of the sample mean, despite the nonnormality of the distribution of individual observations.

EXAMPLE 6.25

Obstetrics The central-limit theorem is illustrated by plotting, in Figure 6.4(a), the sampling distribution of mean birthweights obtained by drawing 200 random samples of size 1 from the collection of birthweights in Table 6.2. Similar sampling distributions of sample means are plotted from samples of size 5, in Figure 6.4(b), and samples of size 10, in Figure 6.4(c). Notice that the distribution of individual birthweights (i.e., sample means from samples of size 1) is slightly skewed to the left. However, the distribution of sample means becomes increasingly closer to bell-shaped as the sample size increases to 5, in Figure 6.4(b), and 10, in Figure 6.4(c). ■■■

EXAMPLE 6.26

Cardiovascular Disease Serum triglycerides are an important risk factor for certain types of coronary disease. Their distribution tends to be positively skewed or skewed to the right, with a few people with very high values, as is shown in Figure 6.5. However, hypothesis tests can be performed based on mean serum triglycerides over moderate samples of people, since from the central-limit theorem the distribution of means will be approximately normal, even if the underlying distribution of individual measurements is not. To further ensure normality, the data can also be transformed onto a different scale. For example, if a log transformation is used, then the skewness of the distribution will be reduced and the central-limit theorem will be applicable for smaller sample sizes than if the data are kept in the original scale. We discuss data transformations in more detail in Chapter 11. ■■■

EXAMPLE 6.27

Obstetrics Compute the probability that the mean birthweight from a sample of 10 infants drawn from the Boston City Hospital population in Table 6.2 will fall between 98.0 and 126.0 oz (i.e., $98 \leq \bar{x} < 126$) if the mean birthweight for the 1000 birthweights from the Boston City Hospital population is 112.0 oz with a standard deviation of 20.6 oz.

SOLUTION

The central-limit theorem is applied and it is assumed that \bar{x} follows a normal distribution with mean $\mu = 112.0$ oz and standard deviation $\sigma/\sqrt{n} = 20.6/\sqrt{10} = 6.51$ oz. It follows that

$$\begin{aligned} Pr(98.0 \leq \bar{x} < 126.0) &= \Phi\left(\frac{126.0 - 112.0}{6.51}\right) - \Phi\left(\frac{98.0 - 112.0}{6.51}\right) \\ &= \Phi(2.15) - \Phi(-2.15) \\ &= \Phi(2.15) - [1 - \Phi(2.15)] = 2\Phi(2.15) - 1 \end{aligned}$$

Refer to Table 3 in the Appendix and obtain

$$Pr(98.0 \leq \bar{x} < 126.0) = 2(.9842) - 1.0 = .968$$

FIGURE
1
centric
(100 = 1)

