FUNDAMENTALS of

Biostatistics

Bernard Rosner

3 1/2" DOS data disk enclosed

FOURTH EDITION

FOURTH EDITION

FUNDAMENTALS OF BIOSTATISTICS

BERNARD ROSNER
HARVARD UNIVERSITY

Duxbury Press
An Imprint of Wadsworth Publishing Company
I(T)P™ 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 Purrington
Technical Illustrator: TUCKERdesign
Cover: William Reuter
Compositor: Jonathan Peck Typographers
Printer: Banta Co., Harrisonburg, VA

This book is printed on acid-free recycled paper.

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

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

*asterisks indicate new section or subsection for the Fourth Edition

CHAPTER 8 Hypothesis Testing: Two-Sample Inference 251

CHAPTER 9 Multisample Inference 299

CHAPTER 10 Hypothesis Testing: Categorical Data 345

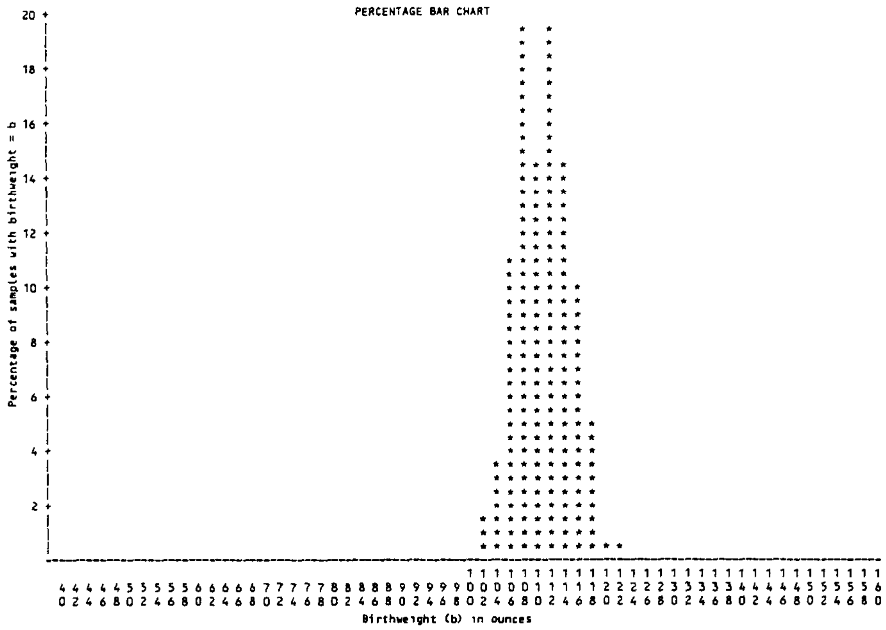
APPENDIX Tables 633

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

$$\bar{x} = (97.2 + 96.8 + \cdots + 97.3)/10 = 97.20°$$

The standard error of this estimate is given by

$$s/\sqrt{10} = 0.189/\sqrt{10} = 0.06°$$

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 $97.20° \pm 2(0.06)° \approx (97.1°–97.3°)$. 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. ■■■

6.5.3 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 $x_1, \ldots, x_n$ be a random sample from some population with mean μ and variance σ^2. Then for large n, $\bar{x} \dot{\sim} N(\mu, \sigma^2/n)$ even if the underlying distribution of individual observations in the population is not normal. (The symbol $\dot{\sim}$ 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., $98 \le \bar{x} < 126$) 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.

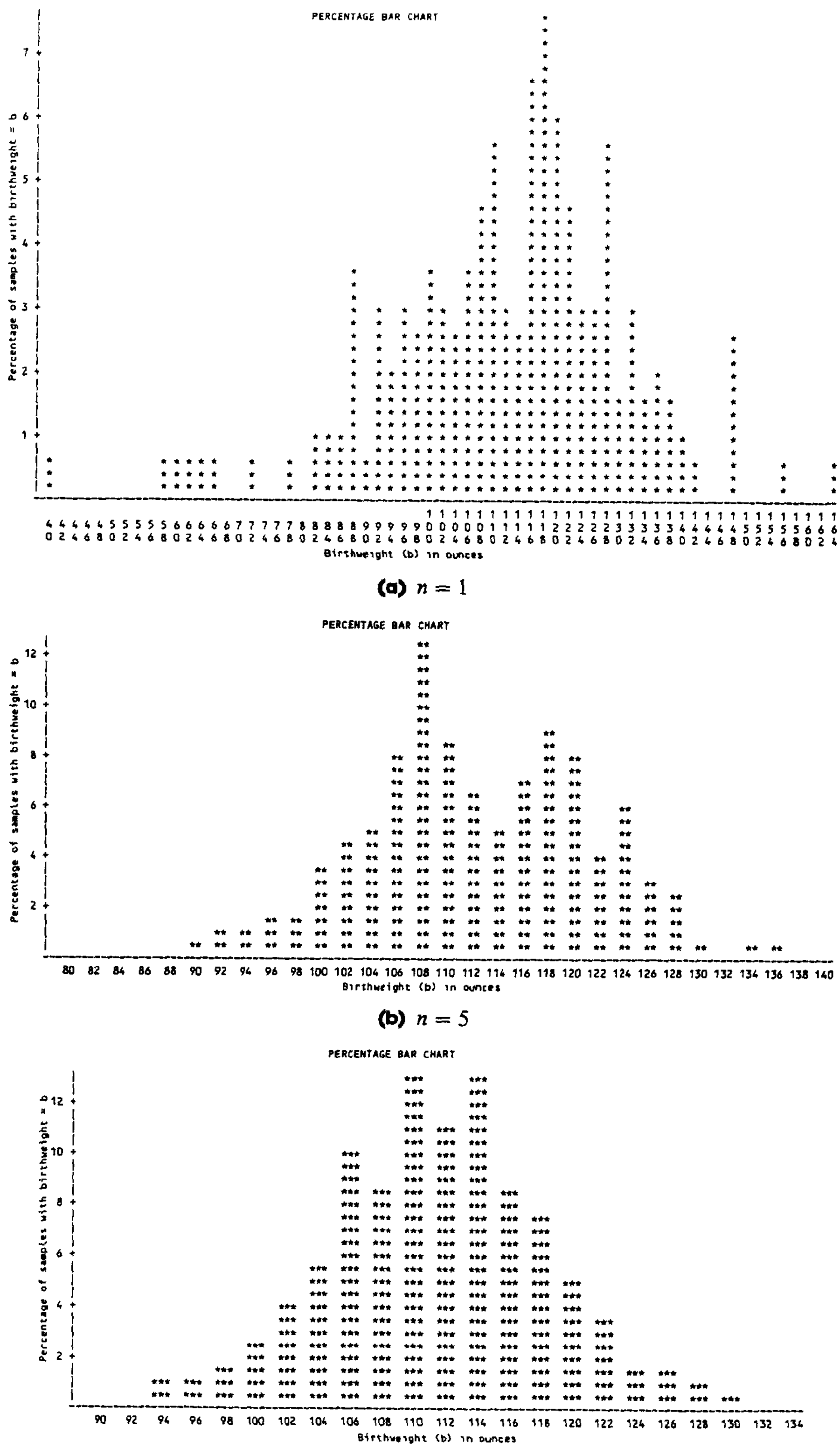
SOLUTION The central-limit theorem is applied and it is assumed that $\bar{x}$ follows a normal distribution with mean $\mu = 112.0$ oz and standard deviation $\sigma/\sqrt{n} = 20.6/\sqrt{10} = 6.51$ oz. It follows that

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

Refer to Table 3 in the Appendix and obtain

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

FIGURE 6.4(a)(b)(c) Illustration of the central-limit theorem (100 = 100.0–101.9, etc.)



(a) $n = 1$

(b) $n = 5$

(c) $n = 10$