STATISTICAL APPLICATIONS *for* HEALTH INFORMATION MANAGEMENT

Second Edition

CAROL E. OSBORN

STATISTICAL APPLICATIONS *for* HEALTH INFORMATION MANAGEMENT

Second Edition

CAROL E. OSBORN, PhD, RHIA
The Ohio State University Health System
Assistant Director
Documentation and Compliance
Department of Medical Information Management

AHIMA
American Health Information Management Association®

JONES AND BARTLETT PUBLISHERS
Sudbury, Massachusetts
BOSTON TORONTO LONDON SINGAPORE

World Headquarters
Jones and Bartlett Publishers
40 Tall Pine Drive
Sudbury, MA 01776
978-443-5000
info@jbpub.com
www.jbpub.com

Jones and Bartlett Publishers
Canada
6339 Ormindale Way
Mississauga, ON L5V 1J2
CANADA

Jones and Bartlett Publishers
International
Barb House, Barb Mews
London W6 7PA
UK

Jones and Bartlett's books and products are available through most bookstores and online booksellers. To contact Jones and Bartlett Publishers directly, call 800-832-0034, fax 978-443-8000, or visit our website at www.jbpub.com.

Substantial discounts on bulk quantities of Jones and Bartlett's publications are available to corporations, professional associations, and other qualified organizations. For details and specific discount information, contact the special sales department at Jones and Bartlett via the above contact information or send an email to specialsales@jbpub.com.

ISBN-13: 978-0-7637-2842-7
ISBN-10: 0-7637-2842-X

Library of Congress Cataloging-in-Publication Data

Osborn, Carol E.
Statistical applications for health information management / Carol E. Osborn.-- 2nd ed.
p. ; cm.
Includes bibliographical references and index.
ISBN 0-7637-2842-X (pbk.)
1. Medical statistics. 2. Health services administration--Statistical methods.
[DNLM: 1. Biometry--methods. 2. Data Interpretation, Statistical. 3. Health Services Administration. 4. Research--methods. WA 950 O815s 2005] I. Title.
RA409.O83 2005
610'.21--dc22

2004029263

6048

Production Credits
Publisher: Michael Brown
Production Director: Amy Rose
Associate Production Editor: Renée Sekerak
Editorial Assistant: Kylah Goodfellow McNeill
Senior Marketing Manager: Ed McKenna
Associate Marketing Manager: Marissa Henderson
Manufacturing Buyer: Therese Connell
Composition: ATLIS Graphics
Cover Design: Timothy Dziewit
Cover Image: © Photos.com
Printing and Binding: Malloy Incorporated
Cover Printing: Malloy Incorporated

Printed in the United States of America
10 09 08 07 06 10 9 8 7 6 5 4 3 2

About the Author

Carol E. Osborn, PhD, RHIA, earned an undergraduate degree in health information management from Mercy College of Detroit, a master's degree in health sciences education and evaluation from the State University of New York at Buffalo, and a PhD in research and evaluation from The Ohio State University.

Dr. Osborn has spent most of her career in the academic setting, teaching at both the University of Illinois at Chicago and at The Ohio State University. She is currently Assistant Director, Medical Information Management, The Ohio State University Health System. She has also consulted in a variety of health care settings including acute care, specialty care, and healthcare-related organizations. She has served on The Ohio State University's Institutional Review Board for over 10 years. She was part of the allied health education consulting team to the Ministry of Health, Saudi Arabia. She has also served as a consultant to HIM baccalaureate degree programs.

Dr. Osborn has been active in professional associations at the state and national levels. She served as secretary, treasurer, and president of the Illinois Medical Record Association. Nationally she has served on the Council of Education, Sub-Panel for Accreditation of Academic Programs, and the Joint Committee on Education, and for 10 years, as a member of the Panel of Accreditation Surveyors. She has authored numerous articles and book chapters for professional journals and textbooks, and she is currently on the editorial review board for *Perspectives in Health Information Management*.

Dedication

To my husband Richards, and sons Rich and Tom.

Preface

This text was written specifically for health information management students enrolled in baccalaureate degree programs and for practicing health information management professionals. This text focuses on applying statistical techniques to problems in health care. Because the focus here is on application, it is assumed that the student has had a previous course in probability theory and the normal distribution. This text is set up so that students can either use the Jones and Bartlett Publishers website that supports this book or input the data for each problem using their own statistical software. It is not the intent of this book to teach the student how to use SPSS, Microsoft Excel, or any other type of statistical package or electronic spreadsheet. These programs are included in this text as examples only; I am not endorsing any of these products. My goal in writing this book was to introduce students and professionals to how statistical techniques can be used to describe and make inferences from health care data. There are many statistical books available on the market, but none is directed specifically to the health information management profession. Also, because there are other texts that introduce the student to traditional hospital statistics such as average length of stay and total inpatient service days, they are not covered in this text.

Contents

CHAPTER 1

Commonly Used Frequency Measures in Health Care

KEY TERMS

- Variable
- Frequency distribution
- Rate
- Ratio
- Proportion
- Dichotomous variables
- Confounding factor
- Confounding variable
- Mortality rates
  - Crude death rate
  - Age-specific death rate (ASDR)
  - Age-adjusted death rate
  - Standard mortality ratio (SMR)
  - Race-specific death rate
  - Sex-specific death rate
  - Cause-specific death rate
  - Case fatality rate
  - Proportionate mortality ratio (PMR)
  - Maternal mortality rate
  - Neonatal mortality rate
  - Postneonatal mortality rate
  - Infant mortality rate
- Morbidity rates
  - Incidence rate
  - Prevalence rate
  - Point prevalence rate
  - Risk ratios
  - Relative risk
  - Odds ratio
  - Attributable risk
  - Kaplan Meier method
  - Kaplan-Meier survival analysis

LEARNING OBJECTIVES

At the conclusion of this chapter, you should be able to:

1. Define key terms.
2. Calculate measures of morbidity, mortality, and risk of disease for health care facilities and communities.
3. Identify variables that affect morbidity and mortality rates over time.

4. Adjust measures of morbidity and mortality by both the direct and indirect methods of standardization.
5. After adjustment, compare health care facility mortality/morbidity rates with community, state, and/or national rates.
6. Calculate risk of disease between groups.
7. Conduct survival analysis for tumor registries and clinical trials.

It is often said that hospitals and other types of health care facilities are data rich but information poor. There are many types of databases within the facility, many contained within the organization's information warehouse. Information warehouses contain both clinical and financial information. It is the job of the health information management professional to turn the data contained in these databases into information that can be used by physicians, administrators, and other interested parties. The health information management professional can become an invaluable member of the health care team by providing data that are presented in a meaningful way and by presenting data that have been analyzed to serve a specific medical or clinical need. Some typical questions might be:

- What are the top 25 medical and top 10 surgical diagnosis-related groups (DRGs) for inpatient discharges from our facility?
- Which medical/surgical services admit the most patients?
- Is the average length of stay (ALOS) for these DRGs significantly different from the national ALOS for these DRGs?
- How do our charges compare with national charges? How does our reimbursement compare with our costs?
- What geographical area does the health care facility serve?
- How many patients were admitted to the facility by payer? What is the number of inpatient service days by payer? What are the average charges by payer?
- How do lengths of stay (LOSs) compare by physician?
- How many patients acquired nosocomial infections?

In the course of this text we will answer these questions. We will learn how to use descriptive statistics to describe patient populations, how to analyze clinical data for significant differences and relationships, and how to present data in graphic form. Our goal is to collect, analyze, and interpret clinical information for both clinical and administrative health care decision makers. We will begin our discussion of clinical data analysis by reviewing morbidity and mortality measures that are often used to describe patient and community populations.

INTRODUCTION TO FREQUENCY DISTRIBUTIONS

In health care, we deal with vast quantities of clinical data. Since it is very difficult to look at data in raw form, data are summarized into frequency distributions. A **frequency distri-**

bution shows the values that a variable can take and the number of observations associated with each value. A **variable** is a characteristic or property that may take on different values. Height, weight, sex, and third-party payer are examples of variables.

For example, suppose we are studying the variable patient LOS in the pediatric unit. To construct a frequency distribution, we first list all the values that LOS can take, from the lowest observed value to the highest. We then enter the number of observations (frequencies) corresponding to a given LOS. Table 1–1 illustrates what the resulting frequency distribution looks like. Note that all values for LOS between the lowest and highest are listed, even though there may not be any observations for some of the values. Each column of the distribution is properly labeled; the total is given in the bottom row. We can also display a frequency distribution by categories into which a variable may fall. Table 1–2 shows a frequency distribution for the number of patients discharged from Critical Care Hospital by religion, a variable composed of categories. The proportion for each category is also displayed in the table. The sum of the proportions for each category is equal to 1.0. We will examine frequency distributions in greater detail in Chapter 4.

Table 1–1 Frequency Distribution for Patient Length of Stay (LOS), Pediatric Unit

LOS in Days	*No. of Patients*
1	2
2	2
3	0
4	6
5	6
6	11
7	6
8	5
9	3
10	1
Total	42

Table 1–2 Frequency Distribution of Number of Patients Discharged from Critical Care Hospital by Religion, July 20xx

Religion	*Number of Discharges*	*Proportion*
Protestant	422	0.48
Catholic	315	0.36
Jewish	20	0.02
Other	127	0.14
Total	884	1.00

RATIOS, PROPORTIONS, AND RATES

Variables often have only two possible categories, such as alive or dead, or male or female. Variables having only two possible categories are called dichotomous. The frequency measures used with **dichotomous variables** are ratios, proportions, and rates. All three measures are based on the same formula:

$$\text{ratio, proportion, rate} = x/y \times 10^n$$

In this formula, x and y are the two quantities being compared, and x is divided by y. 10^n is read as "10 to the nth power." The size of 10^n may equal, for example, 1, 10, 100, or 1,000, depending on the value of n:

$$10^0 = 1$$
$$10^1 = 10$$
$$10^2 = 10 \times 10 = 100$$
$$10^3 = 10 \times 10 \times 10 = 1{,}000$$

Ratios

In a **ratio**, the values of a variable, such as sex (x = female, y = male), may be expressed so that x and y are completely independent of each other, or x may be included in y. For example, the sex of patients discharged from a hospital could be compared in either of two ways:

Female/male or x/y
Female/(male + female) or $x/(x + y)$

In the first option, x is completely independent of y, and the ratio represents the number of female discharges compared to the number of male discharges. In the second option, x is a proportion of the whole, $x + y$. The ratio represents the number of female discharges compared to the total number of discharges. Both expressions are considered ratios.

How, then, would you calculate the female-to-male ratio for a hospital that discharged 457 women and 395 men during the month of July? The procedure for calculating a ratio is outlined in Exhibit 1–1.

Proportions

A **proportion** is a particular type of ratio. A proportion is a ratio in which x is a portion of the whole, $x + y$. In a proportion, the numerator is always included in the denominator. Exhibit 1–2 outlines the procedure for determining the proportion of hospital discharges for the month of July that were female.

Exhibit 1–1 Calculation of a Ratio: Discharges for July 20xx

1. Define *x* and *y*.
   x = number of female discharges
   y = number of male discharges
2. Identify *x* and *y*.
   x = 457
   y = 395
3. Set up the ratio *x*/*y*.
   457/395
4. Reduce the fraction so that either *x* or *y* equals 1.
   1.16/1

There were 1.16 female discharges for every male discharge.

Exhibit 1–2 Calculation of a Proportion: Discharges for July 20xx

1. Define *x* and *y*.
   x = number of female discharges
   y = number of male discharges
2. Identify *x* and *y*.
   x = 457
   y = 395
3. Set up the proportion
   $x/(x + y)$ $457/(457 + 395) = 457/852$
4. Reduce the fraction so that either *x* or $x = y$ equals 1.
   0.54/1.00

The proportion of discharges that were female is 0.54.

Rates

Rates are a third type of frequency measure. In health care, rates are often used to measure an event over time and are sometimes used as performance improvement measures. The basic formula for a rate is:

$$\frac{\text{No. of cases or events occurring during a given time period} \times 10^n}{\text{No. of cases or population at risk during same time period}}$$

or

$$\frac{\text{Total number of times something did happen} \times 10^n}{\text{Total number of times something could happen}}$$

In inpatient facilities, there are many commonly computed rates. In computing the Caesarean section rate, we count the number of Caesarean sections (C-sections) performed during a given period of time; this value is placed in the numerator. The number of cases or population at risk is the number of women who delivered during the same time period; this number is placed in the denominator. By convention, inpatient hospital rates are calculated as the rate per 100 cases ($10^n = 10^2 = 10 \times 10 = 100$) and are expressed as a percentage. The method for calculating the hospital C-section rate is presented in Exhibit 1–3.

Exhibit 1–3 Calculation of C-Section Rate for July 20xx

For the month of July, 23 C-sections were performed; during the same time period, 149 women delivered. What is the C-section rate for the month of July?

1. Define the variable of interest (numerator) and population or number of cases at risk (denominator).
   Numerator: total number of C-sections performed in July
   Denominator: total number of women who delivered in July, including C-sections
2. Identify the numerator and denominator.
   Numerator: 23
   Denominator: 149
3. Set up the rate.
   23/149
4. Divide the numerator by the denominator, and multiply by 100 ($10^n = 10^2$).
   $(23/149) \times 100 = 15.4\%$.

The C-section rate for the month of July is 15.4%.

POPULATION-BASED MORTALITY MEASURES

As the profession of health information management moves into integrated health care delivery systems and assumes more prominence in managed care organizations, it becomes more important to be familiar with community-based mortality and morbidity data. This type of information is often used in planning health services, such as number of inpatient facilities, type of outpatient facilities, and number or size of managed care plans for a given community, as well as for developing managed care contracts with hospitals and physicians.

Crude Death Rate

The **crude death rate** is a measure of the actual or observed mortality in a given population. Crude rates apply to a population without regard to characteristics of the population, such as the distribution of age or sex. The crude death rate is the starting point for further development of adjusted rates. It measures the proportion of a population that has died during a specific period of time, usually one year, or the number of deaths per 1,000 in a community for a given period of time. The crude death rate is calculated as follows (the midinterval pop-

ulation is the estimated population of a given community at the midpoint of the time frame under study):

$$\frac{\text{Total deaths during a given time interval} \times 10^n = \text{deaths per } 10^n}{\text{Estimated midinterval population}}$$

In calculating the crude death rate, the power of *n* is usually equal to the value that will result in a value greater than 1. This allows for easier interpretation of the rate—a death rate of less than 1 per 100 is not very meaningful. For example, the 2004 midyear population of Anytown, USA, is 1,996,355; 275 deaths occurred in 2004. The power of *n* that will result in a whole number is 4; $10^4 = 10 \times 10 \times 10 \times 10 = 10{,}000$. The crude death rate is calculated as follows:

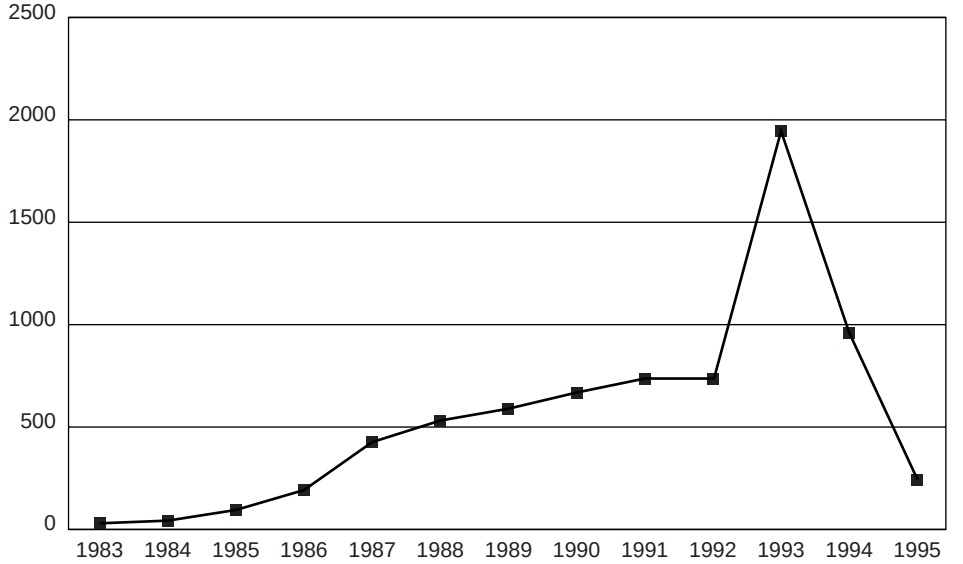
$$(275 \times 10{,}000)/1{,}996{,}355 = 2{,}750{,}000/1{,}996{,}355 = 1.38 \text{ deaths per } 10{,}000$$

When analyzing crude death rates, or any type of rate, it is important to remember that these events do not occur in a vacuum. When analyzing any data set, we need to remember that the data do not stand alone, but reflect trends in the environment. Trends in death rates can be influenced by three variables: time, place, and person. Examples of time, place, and person variables are outlined in Exhibit 1–4. An example of how trended data may be affected by time, place, and person variables is presented in Figure 1–1. The line graph shows that the number of newly diagnosed acquired immune deficiency syndrome (AIDS) cases steadily increased from 1983 to 1992; then a rather dramatic increase occurred in 1993,

Exhibit 1–4 Variables Affecting Trends in Community Morbidity and Mortality

- Time
  - Transition from International Classification of Diseases, 9th Revision (ICD-9) to ICD-10 in coding of death certificates
  - Improvements in medical technology
  - Earlier detection and diagnosis of disease
- Place
  - Changes in environments
  - International and intranational differences in medical technology and the use of medical technology
  - Diagnostic practices of physicians
  - Variation in physician practice patterns by region
- Person
  - Age
  - Sex
  - Ethnicity
  - Social habits (smoking, diet, alcohol)
  - Genetic background
  - Emotional and mental characteristics

Figure 1–1 AIDS Cases Diagnosed in Ohio by Year, 1983–1995.
Source: Reprinted from *Prevention Monthly,* Vol. 19, No. 3, p. 6, 1996, Ohio Department of Health.



which was then followed by a return to previous levels in 1994 and 1995. What happened in 1993 that resulted in such a large increase in the number of newly diagnosed AIDS cases?

This is an example of how the time variable can affect the number of cases diagnosed. In 1993, the case definition of AIDS changed so that individuals who were human immunodeficiency virus (HIV) positive were designated as having full-blown AIDS at an earlier point in the progression of their disease. In 1993, the case definition was expanded to include HIV-positive cases with low CD4 counts, pulmonary tuberculosis, and recurrent pneumonia as AIDS qualifying conditions. The result was that a large number of HIV-positive individuals who already had one of these conditions suddenly qualified as AIDS cases.

Now let's return to our discussion of the crude death rate. Crude rates do not allow for valid comparisons across populations because of differences in the populations—primarily age. This is because age is the most important variable that influences mortality. To illustrate, let's compare two hypothetical crude mortality rates for the states of Arizona (10.9/1,000) and Alaska (4.4/1,000). The conclusion drawn from a comparison of the crude mortality rates is that the death rate is 148% higher in Arizona than in Alaska: $(10.9 - 4.4)/4.4$. However, the discrepancy is due largely to the age differences in the populations of Arizona and Alaska. In general, the population in Arizona is older than the population in Alaska. Without adjusting the rate, one might erroneously conclude that the Alaskan population was healthier than the population of Arizona. In this example, the comparison is confounded by age. **Confounding factor** is a general term used to describe the effect of a third variable on the estimate of risk of a health outcome.

Confounding occurs when a third factor related to outcome is differentially distributed across the levels (or categories) of a variable of interest. When this happens, we must take measures to separate the effect of the **confounding variable**—in this case, age—from the effect of the variable of interest. We can accomplish this by selecting subjects to be compared so that they are matched with respect to the confounding variables, or by using statistical adjustments during analysis to remove the effect of the confounding variable. For example, review the data in Table 1–3. Analysis of the data reveals that the overall crude rate is less for blacks than for whites but that the age-specific death rate for blacks is higher than the rates for whites in every age group. Why is there such a contradiction? It is because the 2001 population of the state of Georgia consisted of old whites and young blacks—33.7% of the white population was 24 years old or younger, and 43.1% of the black population was 24 years old or younger.

Table 1–3 Age-Specific Death Rates per 1,000 Population, State of Georgia, 2001

Race	*Crude Rate*	*< 1 Yr.*	*1–4 Yrs.*	*5–14 Yrs.*	*15–24 Yrs.*	*25–44 Yrs.*	*45–64 Yrs.*	*≥ 65 Yrs.*
White	8.15	6.25	0.42	0.18	0.92	1.49	6.53	51.25
Black	7.04	13.33	0.51	0.24	1.04	2.54	10.68	59.02

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-line Database, wonder.cdc.gov.

Age-Specific Death Rates

In Table 1–3, we see the age-specific death rates (ASDR) for both whites and blacks. The ASDR is calculated as follows:

$$\frac{\text{No. of deaths in the age group of interest} \times 10^n}{\text{Estimated mid-period population in the age group of interest}}$$

Age-Adjusted Death Rates

Age-adjusted death rates are used when there are differences in the age distribution for the populations that are being compared. In Table 1–4, you can see that the population proportions for each age group vary slightly by race. For example, the proportion of whites that are older than age 65 is 0.115 (11.5%) and the proportion of blacks that are older than 65 is 0.064 (6.4%). When we adjust the crude rate for age, we are constructing a summary rate that is free of age bias. In Table 1–4, the ASDR for each age group is expressed as a percentage. There are two methods for adjusting the crude death rate—direct and indirect. We will first discuss the direct method of standardization.

Table 1–4 Calculation of Crude Death Rate, State of Georgia, 2001

Age	*(a) White Population*	*(b) Pop. Prop.*	*(c) Deaths*	*(d) ASDR (c/a) × 100*	*(e) Black Population*	*(f) Pop. Prop.*	*(g) Deaths*	*(h) ASDR (g/f) × 100*
<1	85,648	0.015	535	0.62%	43,727	0.018	583	1.33%
1–4	309,451	0.054	129	0.04%	163,909	0.067	83	0.05%
5–14	768,143	0.134	137	0.02%	444,244	0.181	108	0.02%
15–24	770,501	0.134	706	0.09%	404,438	0.165	420	0.10%
25–44	1,811,149	0.315	2,698	0.15%	793,495	0.324	2,015	0.25%
45–64	1,338,338	0.233	8,746	0.65%	442,005	0.180	4,719	1.07%
65+	660,428	0.115	33,847	5.13%	157,770	0.064	9,312	5.90%
Total	5,743,658	1.000	46,798	0.81%	2,449,588	1.000	17,240	0.70%
	Crude Death Rate = 0.81/100				Crude Death Rate = 0.70/100			

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-line Database, wonder.cdc.gov.

Direct Standardization

To age-adjust the crude death rates, we compare the two groups being studied to a standard population. We then apply the ASDRs for each group to this standard population. As an example, we will use the data in Table 1–4 to standardize the crude death rates for whites and blacks in the state of Georgia. The crude death rate for whites is 0.81 per 100, and the crude death rate for blacks is 0.70 per 100. To calculate the standardized rate, we first calculate the ASDR for each age group in the two populations. We then combine the populations for each age group. By multiplying ASDR for each group by the combined population, we can obtain the expected number of deaths for each group as if the population for each age group were the same. For example, for the age group from 1 to 4 years, we add 309,451 and 163,909 to obtain a total of 473,360. We then multiply the combined population total for each age group by the ASDR to obtain the expected number of deaths for each age group in each of the populations being compared. Thus, the groups are compared on an equal basis. The expected death rate for each population is calculated as follows:

Group	*Age Group*	*Total Population*	*ASDR*	*Expected No. of Deaths*
White	1–4	473,360	0.0004	189.3
Black	1–4	473,360	0.0005	236.7

After we have calculated the expected number of deaths for each age group in each population, we sum the expected number of deaths in each population group, as in Table 1–5. For whites the total number of expected deaths is 59,744.1, and for blacks the total is 77,209.7. The expected number of deaths for each population group is then divided by the

Table 1–5 Calculation of Adjusted Death Rate, Direct Standardization, State of Georgia, 2001

Age	*(a) Total Population*	*(b) ASDR Whites*	*(c) Expected No. Deaths (a × b)*	*(d) ASDR Blacks*	*(e) Expected No. Deaths (a × d)*
< 1	129,375	0.62%	802.1	1.33%	1,720.7
1–4	473,360	0.04%	189.3	0.05%	236.7
5–14	1,212,387	0.02%	242.5	0.02%	242.5
15–24	1,174,939	0.09%	1,057.4	0.10%	1,174.9
25–44	2,604,644	0.15%	3,907.0	0.25%	6,511.6
45–64	1,780,343	0.65%	11,572.2	1.07%	19,049.7
65+	818,198	5.13%	41,973.6	5.90%	48,273.7
Total	8,193,246	0.81%	59,744.1	0.70%	77,209.7
			0.73%		0.94%
Standardized Age-Adjusted Rate = 0.73% 59,744.1/8,193,246				Standardized Age-Adjusted Rate = 0.94% 77,209.7/8,193,246	

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-line Database, wonder.cdc.gov.

combined population. The result is that the standardized age-adjusted death rate for blacks is slightly higher (0.94%) than that for whites (0.73%).

Even though the standardized adjusted rate is not "real," it allows researchers to make better comparisons between groups. The crude rates indicate that the mortality rate is slightly higher for whites than for blacks, but the adjusted rates indicate that mortality among blacks is slightly higher than that for whites. Without adjustment, we would make the assumption that mortality was slightly higher in the white population. An adjusted rate informs us that this may not necessarily be the case.

Indirect Standardization

The indirect method of standardization is used when ASDRs are not available, or when the population that we wish to compare is small, as when we are comparing hospital inpatients to much larger populations. When using this method, we use standard rates obtained from some population and apply them to our population of interest. The basic steps for indirect standardization appear in Exhibit 1–5. In our hypothetical example, we compare 2002 Utah hospital discharges that resulted in death due to pneumonia to the number of hospital discharges that resulted in death due to pneumonia in Salt Lake County, Utah.

In our calculations in Table 1–6, we see that the overall mortality rate due to pneumonia in the state of Utah is 5.08% and that the mortality rate in Salt Lake County is 5.12% [(100 × 100)/1,953]. Salt Lake County had 1.5 more deaths than what was expected on the basis of the standard rates for the state of Utah; therefore, the expected mortality rate is 5.04% [(98.5 × 100)/1,953]. To make the comparison to the standard rates, we calculate a

Exhibit 1–5 Basic Steps for Indirect Standardization

1. Determine the standard mortality rates for pneumonia in the state of Utah for the age groups of interest.
2. Multiply the ASDR for the state of Utah (column c) times the number of county discharges in each age category to obtain the expected number of deaths for each category (columns c × d = column f) in Salt Lake County, Utah.
3. Sum the number of expected deaths.
4. Compute the standard mortality ratio (SMR), which compares the number of actual or observed deaths to the number of expected deaths. In Table 1–6, the number of actual or observed deaths is 100, and the number of expected deaths is 98.5.
5. Multiply the SMR by 100. The SMR is interpreted as a percentage lesser or greater than that of the standard population.

standard mortality ratio (SMR). The SMR compares the actual number of deaths in the group under study (Salt Lake County) to the expected number of deaths based on the standard population rates that were applied to the study group. For the data in Table 1–6, the SMR is calculated as:

$$\text{SMR} = \frac{\text{Observed death rate}}{\text{Expected death rate}} = \frac{0.0512}{0.0504}$$

$$= 1.016 \times 100 = 101.6\%$$

Table 1–6 Mortality Rates Due to Pneumonia (ICD-9-CM Codes 480–486) 2002, Ages 35+, State of Utah versus Salt Lake County, Utah

	State of Utah			*Salt Lake County, Utah*		
Age	*(a) Utah Discharges*	*(b) No. Deaths*	*(c) ASDR (b × 100)/a*	*(d) County Discharges*	*(e) Observed Deaths*	*(f) Expected Deaths (c × d)*
35–45	344	7	2.03%	151	3	3.1
45–54	533	9	1.69%	227	3	3.8
55–64	684	17	2.49%	237	5	5.9
65–74	1,071	53	4.95%	371	17	18.4
75+	2,542	177	6.96%	967	72	67.3
Total	5,174	263	5.08%	1,953	100	98.5
			Observed Death Rate	5.12%		
			Expected Death Rate	5.04%		
			SMR =	0.0512 / 0.0504	1.1016	

Source: Utah Inpatient Hospital Discharge Dataset, Utah Office of Health Care Statistics, www.health.state.ut.us.

If the calculated SMR is equal to 100, the number of observed deaths is the same as the number of expected deaths. If the SMR is greater than 100, the number of observed deaths is greater than the number of expected deaths. The interpretation of the SMR is that Salt Lake County's pneumonia death rate is 1% greater than that for the state of Utah. Stated another way, the death rate is 1% greater than what would be expected on the basis of the **mortality rates** due to pneumonia for the entire state of Utah.

In summary, rates are adjusted to remove the effect of the confounding factor for which the adjustment has been made—in this case, age. However, it is always necessary to calculate the crude rate because this represents the actual event. An adjusted rate is used for comparative purposes; adjusted rates do not reveal the underlying raw data that are shown by the crude rates.

Race- and Sex-Specific Death Rates

Mortality rates may be calculated for any variable of interest, such as race or sex, using the same basic formula specified for calculating the crude death rate. Historically in the United States, men have had higher mortality rates than women, but the gap may be narrowing. In 1995, the U.S. sex-specific rate was 9.2 per 1,000 for men and 8.6 per 1,000 for women. However, in 2001, the sex-specific death rate for men was 8.45 per 1,000 for men and 8.49 per 1,000 for women (Table 1–7).

Table 1–7 Sex-Specific Death Rates, United States, 2001

	Women			*Men*		
Age	*Population*	*Deaths*	*Rate/ 1,000*	*Population*	*Deaths*	*Rate/ 1,000*
Under 1 Year	1,968,011	12,091	6.14	2,057,922	15,477	7.52
1–4 years	7,491,412	2,208	0.29	7,841,553	2,899	0.37
5–9 years	9,861,089	1,366	0.14	10,347,035	1,727	0.17
10–14 years	10,199,195	1,561	0.15	10,711,245	2,441	0.23
15–19 years	9,847,662	3,789	0.38	10,423,650	9,766	0.94
20–24 years	9,630,499	4,500	0.47	10,080,924	14,197	1.41
25–34 years	19,698,788	12,926	0.66	20,116,087	28,757	1.43
35–44 years	22,675,474	33,510	1.48	22,464,812	58,164	2.59
45–54 years	19,971,971	63,217	3.17	19,256,395	104,848	5.44
55–64 years	13,160,005	99,181	7.54	12,155,918	144,958	11.92
65–74 years	10,020,545	189,379	18.90	8,301,935	241,581	29.10
75–84 years	7,585,929	361,187	47.61	4,996,556	340,742	68.20
85 years and over	3,127,729	447,998	143.23	1,320,580	217,533	164.73
Total	145,238,309	1,232,913	8.49	140,074,612	1,183,090	8.45

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-line Database, wonder.cdc.gov.

It would be misleading to review the **sex-specific death rates** without review of the individual age-specific rates. Table 1–7 indicates that the death rate for men is higher for every age group. If we want to determine why the death rate of men is higher than that for women, we can compare causes of death by sex and age group. For example, in the combined age groups from 15 to 44 years, the death rate for men is higher than that for women because accidental death is the leading cause of death for men in these age groups. Sex-specific diseases may account for the differences in the death rates for other age groups, such as prostate cancer in men and breast cancer in women. Calculating the age-specific rates and the sex-specific rates can help us better understand what is taking place in the health care environment.

Cause-Specific Death Rates

The **cause-specific death rate** is the death rate due to a specified cause. It may be stated for an entire population or for any age, sex, or race. The numerator is the number of deaths due to a specified cause and the denominator is the size of the population at midyear. It is usually expressed in terms of a rate per 100,000 ($10^n = 10^5 = 100{,}000$). The formula is:

$$\frac{\text{Deaths assigned to a specified cause during a given time interval} \times 100{,}000}{\text{Estimated midinterval population}}$$

Table 1–8 presents the cause-specific death rates for males and females. The cause-specific death rate for pneumonia in the population aged 45 or older is 62.76 per 100,000 for women and 60.08 per 100,000 for men. While the overall cause-specific death rate for women is higher for women than for men, the cause-specific rates for each age group are consistently higher for men than for women. In reviewing the rates in Table 1–6, we can also see that the death rate increases with age for both men and women.

Table 1–8 Cause-Specific Mortality Rates, By Sex, Due to Influenza and Pneumonia (ICD-10 Codes J10–J18.9), Age 45+, United States, 2001

	Women			*Men*			
Age	*Population*	*Deaths*	*Rate/ 100,000*	*Population*	*Deaths*	*Rate/ 100,000*	*ASDR*
45–54 years	19,971,971	702	3.51	19,256,395	1,099	5.71	4.59
55–64 years	13,160,005	1,117	8.49	12,155,918	1,587	13.06	10.68
65–74 years	10,020,545	2,918	29.12	8,301,935	3,732	44.95	36.29
75–84 years	7,585,929	9,383	123.69	4,996,556	9,294	186.01	148.44
85 years and over	3,127,729	19,689	629.50	1,320,580	10,502	795.26	678.71
Total	53,866,179	33,809	62.76	46,031,384	26,214	56.95	60.08

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-line Database, wonder.cdc.gov.

Case Fatality Rate

The **case fatality rate** or killing power of a disease measures the probability of death among the diagnosed cases of a disease. The higher the ratio, the more virulent the infection. It is most often used as a measure in acute infectious disease. The case fatality rate is not useful in chronic disease because such diseases have a longer and more variable course.

The formula for the case-fatality rate is:

$$\frac{\text{No. of deaths due to a disease during a given time interval} \times 100}{\text{No. of cases of the disease in the same time interval}}$$

Proportionate Mortality Ratio

The **proportionate mortality ratio** (PMR) describes the proportion of all deaths for a given time interval that are due to a specific cause. Each cause is expressed as a percentage of all deaths, and the sum of all the causes is 1.00 (100%). The PMR is not a mortality rate, since the denominator is all deaths, not the population in which the deaths occurred. Its formula is:

$$\frac{\text{No. of deaths due to a disease during a given time interval} \times 100}{\text{No. of deaths from all causes in the same time interval}}$$

The PMR is often used to make comparisons between and within age groups and occupational groups, as well as for the general population. The PMR for pneumonia appears in Table 1–9.

Maternal Mortality Rate

The **maternal mortality rate** measures deaths associated with pregnancy. Pregnancy often places a woman at risk for medical problems that would not usually be encountered in the nonpregnant state, such as hemorrhage or toxemia of pregnancy. Pregnancy also complicates chronic conditions such as diabetes mellitus and heart disease. In some women, pregnancy precipitates gestational diabetes. The maternal mortality rate is calculated only for deaths that are related to pregnancy; thus, if a pregnant woman is killed in an automobile accident, the death is not considered a pregnancy-related death.

The numerator is the number of deaths assigned to causes related to pregnancy during a given time period; the denominator is the number of live births reported during the same period. Because the maternal mortality rate is usually very small, it is usually expressed as the number of deaths per 100,000 live births.

Table 1–9 Proportionate Mortality Ratios for Influenza and Pneumonia (ICD-10 Codes J10–J18.9), United States, 2001

Age	*Influenza and Pneumonia Deaths*	*Total Deaths*	*PMR/100*
0–4 years	411	32,675	1.26
5–9 years	46	3,093	1.49
10–14 years	46	4,002	1.15
15–19 years	66	13,555	0.49
20–24 years	115	18,697	0.62
25–34 years	339	41,683	0.81
35–44 years	983	91,674	1.07
45–54 years	1,801	168,065	1.07
55–64 years	2,704	244,139	1.11
65–74 years	6,650	430,960	1.54
75–84 years	18,677	701,929	2.66
85 years and over	30,191	665,531	4.54
Total	62,029	2,416,003	2.57

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-line Database, wonder.cdc.gov.

Rates of Infant Mortality

There are three rates of infant mortality, all of which are based on age. Of the three, the infant mortality rate is the most commonly used measure for comparing health status between nations. All three rates are expressed in terms of the number of deaths per 1,000.

Neonatal Mortality Rate

The neonatal period is defined as the period from birth up to but not including 28 days of age. The numerator is the number of deaths of infants under 28 days of age during a given time period; the denominator is the total number of live births reported during the same period. The **neonatal mortality rate** may be used as an indirect measure of the quality of prenatal care and/or the mother's prenatal behavior (e.g., tobacco, alcohol, and drug use).

Postneonatal Mortality Rate

The postneonatal period is the time period from 28 days of age up to but not including one year of age. The numerator is the number of deaths among children from age 28 days up to but not including one year of age during a given time period; the denominator is the total number of live births reported less the number of neonatal deaths during the same period. The **postneonatal mortality rate** is often used as an indicator of the quality of the infant's home environment.

Infant Mortality Rate

In effect, the **infant mortality rate** is a summary of the neonatal and postneonatal mortality rates. The numerator is the number of deaths among children under one year of age; the denominator is the number of live births reported during the same period. Table 1–10 provides a summary of these rates.

Table 1–10 Frequently Used Mortality Measures

Measure	*Numerator (x)*	*Denominator*	10^n
Crude death rate	Total no. of deaths reported during given time interval	Estimated midinterval population	1,000 or 10,000
Cause-specific death rate	Total no. of deaths due to a specific cause during a given time interval	Estimated midinterval population	100,000
Proportionate mortality ratio	Total no. of deaths due to a specific cause during a given time interval	Total no. of deaths from all causes during the same time interval	100 or 1,000
Case fatality rate	Total no. of deaths assigned to a specific disease during a given time interval	Total no. of cases of the disease during the same time interval	100
Neonatal mortality rate	No. of deaths under 28 days of age during a given time interval	No. of live births during the same time interval	1,000
Postneonatal rate	No. of deaths from 28 days up to and not including one year of age, during a given time interval	No. of live births during the same time interval less neonatal deaths	1,000
Infant mortality rate	No. of deaths under one year of age during a given time interval	No. of live births during the same time interval	1,000
Maternal mortality rate	No. of deaths assigned to pregnancy-related causes during a given time interval	No. of live births during the same time interval	100,000

FREQUENTLY USED MEASURES OF MORBIDITY

Some commonly used measures to describe the presence of disease in a community or a specific location, such as a nursing home, are incidence and prevalence rates. Disease can be illness, injury, or disability, and measures can be further elaborated into specific measures of age, sex, race, or other characteristics of a particular population.

Incidence Rate

The **incidence rate** is the commonly used measure for comparing frequency of disease in populations. Populations are compared using rates instead of raw numbers because rates adjust for differences in the size of the populations. The incidence rate expresses the probability or risk of illness in a population over a period of time. The formula for calculating the incidence rate is:

$$\frac{\text{Total no. of new cases of a specific disease during a given time interval} \times 10^n}{\text{Total population at risk during the same time interval}}$$

For the incidence rate, the denominator represents the population from which the case in the numerator arose, such as a nursing home, school, or company. For 10^n, a value is selected so that the smallest rate calculated results in a whole number.

Prevalence Rate

The **prevalence rate** is the proportion of persons in a population that have a particular disease at a specific point in time, or over a specified period of time. The formula for calculating the prevalence rate is:

$$\frac{\text{All new and preexisting cases of a specific disease during a given time interval} \times 10^n}{\text{Total population during the same time period}}$$

Incidence and prevalence rates are often confused. The rates differ based on which cases are included in the numerator. The numerator of the incidence rate is *new cases* occurring during a given time period; the numerator of the prevalence rate is *all cases* present during a given time period. In comparing the two, you can see that the incidence rate includes only individuals whose illness began during a specified period of time, whereas the numerator for the prevalence rate includes all individuals ill from a specified cause, regardless of when the illness began. A case is counted in prevalence until the individual recovers. Exhibit 1–6 presents an example of incidence and prevalence rates in a nursing home.

At times we may be interested in tracking prevalence rates more closely—for example, tracking *Klebsiella pneumoniae* on a daily basis. We can do this by calculating the **point**

Exhibit 1–6 Calculation of Incidence and Prevalence Rates of *Klebsiella pneumoniae* at the Manor Nursing Home, Month of January

At Manor Nursing Home, 10 new cases of *Klebsiella pneumoniae* occurred in January. For the month of January there were a total of 17 cases of *Klebsiella pneumoniae*. The facility had 250 residents during January.

What are the incidence and prevalence rates for *Klebsiella pneumoniae* during January?

Incidence Rate

1. Identify the variable of interest (numerator) and population at risk (denominator).
   Numerator: Total no. of new cases of *Klebsiella pneumoniae* in January
   Denominator: Total no. of nursing home residents in January
2. Identify the numerator and denominator.
   Numerator: 10
   Denominator: 250
3. Set up the rate.
   10/250
4. Divide the numerator by the denominator and multiply by 100 ($10^n = 10^2$).
   $(10/250) \times 100 = 0.04 = 4.0\%$

The incidence rate for *Klebsiella pneumoniae* for the month of January is 4.0%.

Prevalence Rate

1. Identify the variable of interest (numerator) and population at risk (denominator).
   Numerator: Total no. of cases of *Klebsiella pneumoniae* in January
   Denominator: Total no. of nursing home residents in January
2. Identify the numerator and denominator.
   Numerator: 17
   Denominator: 250
3. Set up the rate.
   17/250
4. Divide the numerator by the denominator and multiply by 100 ($10^n = 10^2$). $(17/250) \times 100 = 0.068\%$

The prevalence rate for *Klebsiella pneumoniae* for the month of January is 6.8%.

prevalence rate. The point prevalence rate is the number of cases of a specific disease at a specific point in time. The point prevalence rate is more narrow in its time frame than the general prevalence rate. Table 1–11 displays the point prevalence rates for each day during one week in January.

For a summary of morbidity measures, see Table 1–12.

Table 1–11 Point Prevalence Rates of *Klebsiella pneumoniae* for the Manor Nursing Home, Week of January 3

	Sun.	*Mon.*	*Tues.*	*Weds.*	*Thurs.*	*Fri.*	*Sat.*
No. of cases	10	12	14	13	15	16	16
No. of residents	250	250	250	250	250	250	250
Point Prevalence rate	4.0%	4.8%	5.6%	5.2%	6.0%	6.4%	6.4%

Table 1–12 Frequently Used Measures of Morbidity

Measure	*Numerator*	*Denominator*
Basic formula for computing rates	No. of events occurring during a given time interval	No. of cases or population at risk during the same time interval
Incidence rate	Total no. of new cases of a specific disease during a given time interval	Total population at risk during the same time interval
Prevalence rate	All new and preexisting cases of a specific disease during a given time interval	Total population during the same time interval
Relative risk	Risk for exposed group	Risk for unexposed group
Relative risk using incidence rates	Incidence rate for group of primary interest	Incidence rate for comparison group
Attributable risk	Risk for exposed group minus risk for unexposed group	Risk for exposed group

RELATIVE MEASURES OF DISEASE FREQUENCY

Risk Ratio/Relative Risk

Relative risk (RR) is a ratio that compares the risk of disease or other health event between two groups. What we are comparing is the actual risk of illness between the two groups. In calculating relative risk, we are using the actual rates of illness for each group to make the comparison. The two groups may be differentiated by demographic variables, such as sex or race, or by exposure to a suspected risk factor.

The group of primary interest is labeled as the exposed group, and the comparison group is labeled the unexposed group. The exposed group is placed in the numerator, and the unexposed group is placed in the denominator:

$$\frac{\text{Risk for exposed group}}{\text{Risk for unexposed group}}$$

A risk ratio of 1.0 indicates that the risk is identical in both groups; a risk ratio greater than 1.0 indicates that the risk is greater for the numerator group; and a risk ratio of less than 1.0 indicates that the risk is less for the numerator group.

As an example, we can compare the risk of death due to malignancies in men versus women in Michigan in 2001. First, the collected data are summarized in a two-by-two table. Two-by-two refers to two variables, each with two categories, as shown in Table 1–13.

Table 1–13 Relative Risk of Death Due to Malignancies, Women versus Men Aged 65+, State of Michigan, 2001

	Death Due to Pneumonia		
Sex	*Yes*	*No*	*Total*
Men	7,153	21,507	28,660
	(a)	(b)	(a + b)
Women	6,565	28,890	35,455
	(c)	(d)	(c + d)

Risk of illness among men:
$a/(a + b) = 7{,}153/(7{,}15321{,}507) = 0.2496$
Risk of illness among women
$c/(c + d) = 6{,}565/(6{,}565 + 28{,}890)\ 0.1852$
Risk ratio, men to women: 0.2496/1852 = 1.34
Thus, the risk of death due to malignancy among men aged 65+ is 1.3 times greater than the risk of death due to malignancy in women in the same age group.

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-line Database, wonder.cdc.gov.

To determine the risk of death among men, we compare the total number of men who died from malignancies ($a = 7{,}153$) to the total number of men in the group of interest ($a + b = 7{,}153 + 21{,}507$). The same procedure is followed to determine the risk of death due to pneumonia among women. The two ratios are then compared to determine the RR of death due to malignancies among men as compared to women. A summary of these calculations appears in Table 1–13. Note that the RR in each group is somewhat high, 25.0% and 18.5% respectively. Deaths due to malignancies were the second leading cause of death in the state of Michigan in 2001.

Instead of using the risk ratios to compare risks between groups, we can use actual rates to make the same comparisons. In Table 1–14, hypothetical mortality rates are used to com-

Table 1–14 Lung Cancer Data

Cigarettes/Day	*Death Rate/1,000/Year*
0	0.07
1–14	0.57
15–24	1.39
25+	2.27

Rate ratios:
1–14 cigarettes/day to nonsmokers:
0.57/0.07 = 8.1
15–24 cigarettes/day to nonsmokers:
1.39/0.07 = 19.9
25+ cigarettes/day to nonsmokers:
2.27/0.07 = 32.4

Thus, the risk is 8.1 times greater for those who smoke 1 to 14 cigarettes per day than for nonsmokers; 19.9 times greater for those who smoke 15 to 24 cigarettes per day than for nonsmokers; and 32.4 times greater for those who smoke 25 cigarettes per day than for nonsmokers.

Source: Adapted from *Principles of Epidemiology: An Introduction to Applied Epidemiology and Biostatistics,* p. 95, 1992, United States Department of Health and Human Services, Public Health Service.

pare the risk of death due to lung cancer by number of cigarettes smoked per day. Using the same procedure, we can compare the risk of stroke between men who smoke and men who do not smoke. In this example, we are trying to determine if there is a greater risk of stroke among men who smoke than among men who do not smoke. The statistic is called "relative risk using incidence rates" and is calculated as:

$$\frac{\text{Incidence rate for group of primary interest}}{\text{Incidence rate for comparison group}}$$

The data for this example are presented in Table 1–15. Note that these ratios represent only RR, or the possibility of acquiring an illness, in comparison to another group.

Odds Ratio

The **odds ratio** (OR) is another relative measure of occurrence of illness. The odds in favor of a particular event are defined as the frequency with which the event occurs divided by the frequency with which it does not occur. Estimates of RR and the OR are both used to measure the strength of the association between exposure and disease. The OR is an estimate of

Table 1–15 Twelve-Year Risk of Stroke Among Male Smokers and Nonsmokers

	Stroke		
Smokers	*Yes*	*No*	*Total*
Yes	171	3,264	3,435
No	117	4,320	4,437
Total	288	7,584	7,872

Risk of stroke among smokers:
171/3,435 = 0.049
Risk of stroke among nonsmokers:
117/4,437 = 0.026
Risk of male smokers to male nonsmokers:
0.049/0.026 = 1.88
Thus, the risk of stroke is 1.88, or almost two times greater in men who smoke than men who do not smoke.

RR. It is calculated from data obtained from retrospective studies where actual incidence rates are not calculated.

To calculate the OR, a two-by-two table is first constructed as shown in Table 1–16. Exhibit 1–7 displays the calculation of the odds ratio using the data from Table 1–15. The results indicate that the odds of having a stroke is 1.93 times greater in men who smoke than in men who do not smoke.

Table 1–16 Two-by-Two Table for Odds Ratio

	Disease	
Risk Factor	*Cases*	*Non-cases*
Present	*a*	*b*
Absent	*c*	*d*

Odds Ratio = $(a \times d)/(b \times c)$, where a = number of persons with disease and with exposure of interest, b = number of persons without disease and with exposure of interest, c = number of persons with disease but without exposure of interest, and d = number of persons without disease and without exposure of interest.
$a + c$ = total persons with disease (cases)
$b + d$ = total persons without disease (controls)

Exhibit 1–7 Procedure for Calculating Odds Ratio (OR)

$$OR = (a/b) \div (c/d)$$

$$= \frac{(a \times d)}{(b \times c)}$$

$$OR = \frac{171 \times 4{,}320}{3{,}264 \times 117}$$

$$= 1.93$$

The probability of having a stroke is 1.93 times greater in men who smoke than in men who do not smoke.

The interpretation of the OR is similar to that for RR. If the exposure is not related to the diagnosis, the OR will equal 1; if the exposure is positively related to the disease, the OR will be greater than 1; and if the exposure is negative, the OR will be less than 1. We could also apply this same ratio, or any others, to the acute care setting. An outcomes evaluator learns that patients on the surgical unit were exposed to the *E. coli* bacterium. Data were collected for two weeks to determine if the odds for obtaining *E. coli* infection were greater for patients on the surgical units than for patients hospitalized on the medical unit. The data are displayed in Table 1–17. As you can see from the calculations for the OR, the odds or probability of obtaining an *E. coli* infection is 2.68 times greater for a patient hospitalized on the surgical unit than for a patient hospitalized on a medical unit.

Table 1–17 *E.Coli* Infections of Medical and Surgical Patients

	Nosocomial Infection		
Hospital Unit	*Yes*	*No*	*Total*
Surgical Unit	20	628	648
Medical Unit	10	842	852
Total	30	1,470	1,500

The odds ratio is calculated as follows:
OR = $(a \times d)/(b \times c) = (20 \times 842)/(10 \times 628) = 2.68$

When the health outcome is uncommon, the OR approximates the RR. Using the same data from Table 1–17, we can determine the RR as follows:

Risk of infection on surgical unit: $20/648 = 0.031$

Risk of infection on medical unit: $10/852 = 0.012$

Risk of infection on surgical unit compared to medical unit: $0.031/0.012 = 2.58$

As you can see, the results for both the OR and the RR are similar: 2.68 and 2.58, respectively.

Attributable Risk

The **attributable risk** (AR) is a measure of the impact of a disease or other causative factor on a population. With this calculation, we assume that the occurrence of the disease in a group not exposed to the risk factor represents the baseline or expected risk for that disease; any risk above that level in the exposed group is attributed to exposure to the risk factor. Basically, the assumption is that the disease will occur in some individuals even without exposure to a given risk factor. The AR measures the additional risk of illness as a result of an individual's exposure to the risk factor. With AR, we attempt to answer the question, "How much of the disease that occurs can be attributed to a certain exposure?" and subsequently, "How much of the risk of disease can we prevent if we eliminate the exposure to the risk factor in question?"

$$\frac{\text{(Risk for exposed group)} - \text{(risk for unexposed group)} \times 100}{\text{Risk for exposed group}}$$

Using the lung cancer data from Table 1–14, we calculate the attributable proportion as outlined in Exhibit 1–8.

Exhibit 1–8 Calculation of Attributable Proportion

1. Identify the exposed group rate. Lung cancer death rate for smokers of 1–14 cigarettes per day = 0.57 per 1,000 per year

2. Identify the unexposed group rate.

   0.07 per 1,000 per year

3. Calculate the attributable proportion.

$$\frac{0.57 - 0.07 \times 100 = 87.7}{0.57}$$

The conclusion from the calculation of the attributable proportion is that 87.7% of the lung cancer cases are due to or attributed to smoking 1 to 14 cigarettes per day. Approximately 12% (1.00 − 0.877) of the cases in this group would have occurred without exposure to the risk factor—in this case, cigarettes. By carrying out the calculations for the remaining two groups, we can see that the AR increases with the number of cigarettes smoked per day.

$$\text{AR 15–24 cigarettes/day} = [(1.39 - 0.07)/1.39] \times 100 = 95.0\%$$
$$\text{AR 25+ cigarettes/day} = [(2.27 - 0.07)/2.27] \times 100 = 96.9\%$$

Approximately 5% and 3%, respectively, of the individuals in these two groups would have acquired the disease regardless of whether or not they smoked cigarettes.

KAPLAN-MEIER SURVIVAL ANALYSIS

Many individuals within the health information management profession are employed in tumor registries or in the capacity of assisting researchers in analyzing data from clinical trials. In clinical trials, the researcher is interested in determining whether a specific medical or surgical intervention improves survival for a particular condition. A major criterion in measuring the success of a clinical trial is the survival time of individuals undergoing the experimental treatment. In survival analysis we are examining the survival rates as a result of a clinical trial involving a medical or surgical intervention. A major problem in conducting survival analysis is that patients may be lost to follow up or some may be **censored**. A censored patient is one who for some reason is unable to complete the study.

There are several methods for analyzing survival rates, but we will limit the discussion to the **Kaplan Meier method**, a type of life table analysis, since it is most often used in analysis of data collected from clinical trials. **Kaplan-Meier survival analysis** requires a dichotomous outcome such as survival/death or improvement/no improvement.

The major reason for using the Kaplan Meier method is that it takes into account some of the problems commonly encountered when conducting prospective studies. The Kaplan Meier method compensates for subjects who are lost to follow-up or who are unable to complete the study. To conduct an accurate survival analysis, we need to know:

- the reason for patients' withdrawal from the study (i.e., death, loss to follow-up, or censorship)
- the date of withdrawal from study (i.e., date of death, date patient last seen alive or lost to follow-up, or date withdrawn from study)

When survival time is censored, the subject is alive at the time of analysis, or was alive at the time last seen. Survival times tagged with a "+" indicate that they are censored. Table 1–18 presents some hypothetical data for 10 patients in a clinical trial for treatment of bladder cancer. The survival times, in months (column 1), for each patient are rank ordered from lowest to highest.

Each row in Table 1–18 represents an interval. The first row is the first study interval. An interval is a death-free time period. So row 1, column 6, represents a death-free time period of less than 23 months. This is interpreted as meaning that the probability (p_x) of surviving up to but less than 23 months is 1.000 (10/10). The p_x of the first interval is always 1.000 because the first death ends the first interval. The occurrence of a death ends one death-free interval and begins another.

Table 1–18 Hypothetical Data on Survival Times for Bladder Cancer Patients

(1) Survival Time Mo.	(2) No. Living Prior to Subject's Death	(3) No. Living After Subject's Death	(4) # Lost to Follow-Up	(5) p_x	(6) Interval for p_x (Mo.)	(7) p_x at End of Interval
—	—	—	—	1.000	0 to <23	1.000
23	10	9	—	0.900	23 to <34	0.900
34	9	8	—	0.889	34 to <37	0.800
37	8	7	—	0.875	37 to <41	0.700
40+			1	—	—	—
41	6	5	—	0.833	41 to <42	0.583
42	5	4	—	0.800	42 to <43	0.466
43	4	3	—	0.750	43 to <45	0.350
45	3	2	—	0.667	45 to <47	0.233
47	2	1	—	0.500	47 to <48	0.117
48+	1	1	1	1.000	>48	0.117

Column 1 in Table 1–18 indicates the survival time, in months, for each subject. Two patients were lost to follow-up, as indicated by "+" — one at 40 months and one at 48 months. Patients lost to follow-up are not included in the calculations of survival rates. Columns 2, 3, and 4 indicate the number surviving before and after each death and the number lost to follow-up during that interval. Column 5 is the proportion of patients surviving the interval and is obtained by dividing the proportion surviving from the beginning of the interval—from the time of the previous death to just before the next death. For example, for the interval "23 to <34," 10 patients were alive at the start of the interval, and 9 were alive at the end. To obtain p_x, divide 9 by 10 to obtain 0.900.

Column 6 is the death-free period—that is, the time of the last death to the time of the next death. Column 7, p_x, is the proportion of subjects surviving from the beginning of the study to the end of the interval. The p_x is obtained by multiplying the p_x values of all the intervals up to and including the row of interest. For the survival time of 34 months, p_x is obtained by multiplying $1.000 \times 0.900 \times 0.889 = 0.800$. Based on the calculations in Table 1–18, the probability of surviving 48 months is 0.117.

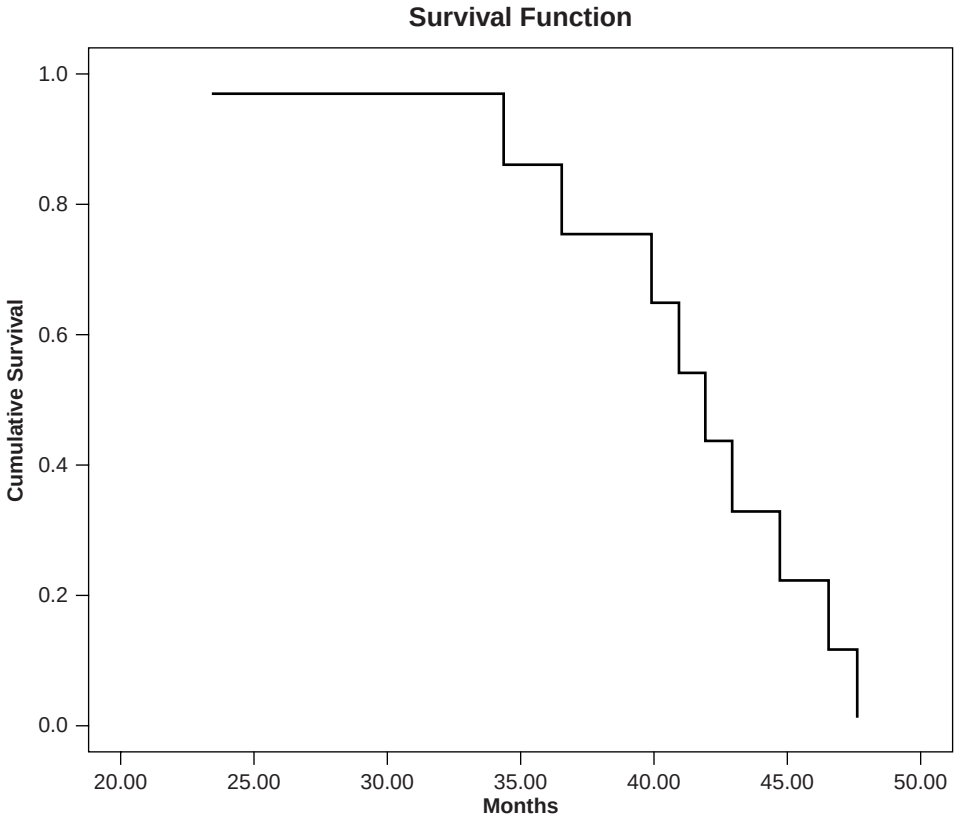
We can use SPSS (Statistical Package for the Social Sciences) to conduct the Kaplan-Meier survival analysis. SPSS is a microcomputer statistical package that we will use throughout this text to solve statistical problems. For the Kaplan-Meier survival analysis, two columns on the data sheet need to be completed. The first column indicates the survival time, in months, for each case; the second column indicates whether the survival time is censored. This can be accomplished by assigning "1" for uncensored survival times and "2" for censored survival times under the "Define Variable" selection. An example of the SPSS data sheet appears in Exhibit 1–9.

Exhibit 1–9 SPSS Data Sheet for Survival Data

Survival Time (Mo.)	*Status*
23.00	Uncensored
34.00	Uncensored
37.00	Uncensored
40.00	Censored
41.00	Uncensored
42.00	Uncensored
43.00	Uncensored
45.00	Uncensored
47.00	Uncensored
48.00	Censored

After completing the data sheet, select "Survival" and then "Kaplan Meier" under the "Statistics" menu. The output, including the survival graph, appears in Figure 1–2. Note that the SPSS printout provides only the p_x (cumulative survival)—the probability of surviving to the end of the interval.

Figure 1–2 SPSS Output for Kaplan-Meier Survival Analysis



Survival Analysis for MONTHS (survival time in months)

Time	*Status*	*Cumulative Survival*	*Standard Error*	*Cumulative Events*	*Number Remaining*
23.00	Uncensored	0.9000	0.0949	1	9
34.00	Uncensored	0.8000	0.1265	2	8
37.00	Uncensored	0.7000	0.1449	3	7
40.00	Censored	0.6000	0.1549	4	6
41.00	Uncensored	0.5000	0.1581	5	5
42.00	Uncensored	0.4000	0.1549	6	4
43.00	Uncensored	0.3000	0.1449	7	3
45.00	Uncensored	0.2000	0.1265	8	2
47.00	Uncensored	0.1000	0.0949	9	1
48.00	Censored	0.0000	0.0000	10	0

Number of Cases: 10 Censored: 0 (.00%) Events: 10

Survival Time	*Standard Error*	*95% Confidence Interval*
Mean: 40.00	2.32	(35.45, 44.55)
Median: 41.00	1.58	(37.90, 44.10)

SPSS provides a summary of the number of cases included in the analysis, including the number of censored cases. The confidence intervals for the mean and median survival times also are provided. (We will discuss confidence intervals in Chapter 5.) The graph depicts the cumulative survival rate for the group under study. Time, in months, is displayed on the *x*-axis, and proportion surviving is displayed on the *y*-axis.

CONCLUSION

In this chapter, we have discussed rates, ratios, and proportions in the form of mortality and morbidity rates and RR. Facility-based morbidity and mortality rates can be compared with community, state, and national rates after adjustment. We can adjust rates using either the direct or the indirect method. Crude rates are important for internal analysis or other non-comparative purposes.

We also reviewed various ratios that are used to measure frequency of disease. Using the various risk ratios and the OR, we can compare risk of certain diseases and causes of morbidity between groups.

Last, we discussed one method commonly used for survival analysis—the Kaplan Meier method. Survival analysis is a tool often used in tumor registries and when analyzing results of clinical trials.

ADDITIONAL RESOURCES

Gail, M.H. and Benichou, J., eds. 2000. *Encyclopedia of epidemiologic methods.* Wiley Reference Series in Biostatistics. Hoboken, NJ: Wiley Publishers.

Gordis, L. 1996. *Epidemiology.* Philadelphia: W.B. Saunders.

Jekel, J.F., et al. 1996. *Epidemiology, biostatistics and preventive medicine.* Philadelphia: W.B. Saunders.

Morton, R., et al. 1989. *A study guide to epidemiology and biostatistics.* Gaithersburg, MD: Aspen Publishers, Inc.

Anderson, K.N., Anderson, L.E., and Glanze, W.D. *Mosby's medical, nursing, and allied health dictionary.* 4th ed. 1994. St. Louis, MO: C.V. Mosby.

Norman, G.R., and Streiner, D.L. 2000. *Biostatistics: The bare essentials.* Ontario, Canada: Hamilton.

Ohio Department of Health. 1996. *Prevention Monthly* 19, no. 3:6.

Pagano, M., and Gauvreau, K. 2000. *Principles of biostatistics.* Pacific Grove, CA: Brooks Cole Publishing.

Ratelle, S. 2001. *Preventive medicine and public health: PreTest self-assessment and review.* New York: McGraw-Hill/ Appleton & Lange McGraw-Hill.

U.S. Department of Health and Human Services (US DHHS), Public Health Service. 1992. *Principles of epidemiology: An introduction to applied epidemiology and biostatistics.*

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS), Office of Analysis, Epidemiology, and Health Promotion (OAEHP), Compressed Mortality File (CMF) compiled from CMF 1968–1988 Series 20, No. 2A 2000, CMF 1989–1998, Series 20, No. 2E 2003 and CMF 1999–2001, Series 20, No. 2G 2004 on CDC WONDER on-line database. wonder.cdc.gov.

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for HIV, STD, and TB Prevention, Division of HIV/AIDS Prevention, AIDS Public Information Data Sets, CDC WONDER on-line data base. wonder.cdc.gov.

Utah Inpatient Hospital Discharge Data Set. http://hlunix.ex.state.ut.us/had.

Appendix 1–A

Exercises for Solving Problems

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
2. Describe the differences and similarities between rates, ratios, and proportions.
3. Outline the procedure for age-adjusting crude mortality rates by the direct standardization method.
4. Describe the differences between the direct and indirect standardization methods of adjusting mortality and morbidity rates.
5. Describe the differences between neonatal mortality rate, postneonatal mortality rate, and infant mortality rate.
6. Describe the difference between incidence and prevalence rates.

MULTIPLE CHOICE

For questions 1 and 2, refer to the following table:

Age Group	*Population*	*No. of Deaths*
< 30	15,000	20
30–65	17,000	55
> 65	6,000	155

1. What is the crude mortality rate?
   a. 230
   b. 6.1 per 1,000
   c. 8.6 per 1,000
   d. 6.1 per 10,000

2. The age-specific death rate for the over-65 age group is:
   a. 155
   b. 25.8 per 1,000
   c. 1.55 per 10,000
   d. 25.8 per 10,000

PROBLEMS

1. Review the hypothetical data on deaths in the MICU in Table 1–A–1 and answer the questions that follow:
   a. What is the ratio of male deaths to female deaths?
   b. What proportion of the patients who died were admitted from the Emergency Department? What proportion were transfers from other hospitals?
   c. The total number of patients discharged from DRG 475 was 61. What is the case fatality rate for DRG 475?
   d. The total number of patients discharged from DRG 483 was 51. What is the case fatality rate for DRG 483?
   e. What is the relative risk of death for patients discharged from DRG 475 compared to discharges from DRG 483?

Table 1–A–1 Critical Care Hospital, Deaths in the MICU by DRG

DRG	*DRG Title*	*Adm. Source*	*Gender*	*LOS*
001	Craniotomy Age >17 W Cc	SNF	Male	2
014	Intracranial Hemorrhage & Stroke W Infarct	Other	Male	3
014	Intracranial Hemorrhage & Stroke W Infarct	Emerdept	Female	3
020	Nervous System Infection Except Viral Meningitis	Other	Female	15
075	Major Chest Procedures	Hospital	Male	6
105	Cardiac Valve & Oth Major Cardiothoracic Proc W/O Card Cath	Hospital	Female	23
123	Circulatory Disorders W Ami, Expired	Hospital	Male	7
123	Circulatory Disorders W Ami, Expired	Other	Male	1
123	Circulatory Disorders W Ami, Expired	Other	Male	4
123	Circulatory Disorders W Ami, Expired	Emerdept	Male	5
172	Digestive Malignancy W Cc	Emerdept	Male	1
172	Digestive Malignancy W Cc	Physician	Male	1
188	Other Digestive System Diagnoses Age >17 W Cc	SNF	Female	1
191	Pancreas, Liver & Shunt Procedures W Cc	Hospital	Male	9
202	Cirrhosis & Alcoholic Hepatitis	Physician	Female	15
202	Cirrhosis & Alcoholic Hepatitis	Other	Male	1
202	Cirrhosis & Alcoholic Hepatitis	Emerdept	Male	24

continued

DRG	DRG Title	Adm Source	Gender	LOS
205	Disorders Of Liver Except Malig,Cirr,Alc Hepa W Cc	Physician	Male	20
331	Other Kidney & Urinary Tract Diagnoses Age >17 W Cc	Physician	Male	44
357	Uterine & Adnexa Proc For Ovarian Or Adnexal Malignancy	Physician	Female	24
416	Septicemia Age >17	Hospital	Female	4
416	Septicemia Age >17	Other	Male	2
449	Poisoning & Toxic Effects Of Drugs Age >17 W Cc	Other	Male	1
473	Acute Leukemia W/O Major O.R. Procedure Age >17	Physician	Male	5
475	Respiratory System Diagnosis With Ventilator Support	Physician	Male	25
475	Respiratory System Diagnosis With Ventilator Support	Physician	Female	1
475	Respiratory System Diagnosis With Ventilator Support	Hospital	Female	1
475	Respiratory System Diagnosis With Ventilator Support	Other	Female	1
475	Respiratory System Diagnosis With Ventilator Support	Hospital	Female	21
475	Respiratory System Diagnosis With Ventilator Support	Other	Male	5
475	Respiratory System Diagnosis With Ventilator Support	Hospital	Female	8
475	Respiratory System Diagnosis With Ventilator Support	Emerdept	Female	10
475	Respiratory System Diagnosis With Ventilator Support	Clinic	Male	13
475	Respiratory System Diagnosis With Ventilator Support	SNF	Female	1
475	Respiratory System Diagnosis With Ventilator Support	Emerdept	Male	5
475	Respiratory System Diagnosis With Ventilator Support	Clinic	Female	4
475	Respiratory System Diagnosis With Ventilator Support	Emerdept	Male	3
475	Respiratory System Diagnosis With Ventilator Support	Other	Female	12
475	Respiratory System Diagnosis With Ventilator Support	Other	Female	5
483	Trac W Mech Vent 96+Hrs Or Pdx Except Face, Mouth & Neck Dx	Hospital	Female	30
483	Trac W Mech Vent 96+Hrs Or Pdx Except Face, Mouth & Neck Dx	Hospital	Male	19
483	Trac W Mech Vent 96+Hrs Or Pdx Except Face, Mouth & Neck Dx	Hospital	Male	22
483	Trac W Mech Vent 96+Hrs Or Pdx Except Face, Mouth & Neck Dx	Physician	Female	46
483	Trac W Mech Vent 96+Hrs Or Pdx Except Face, Mouth & Neck Dx	Other	Female	28

2. Review the data in Table 1–A–2 and answer the questions that follow.
   a. What is the case fatality rate for AIDS for the years 1981 through 1995?
   b. The midyear population for the state of Ohio in 1994 was 11,140,950. What is the incidence rate for AIDS for 1994?

Table 1–A–2 AIDS Cases in Ohio 1981–1995

Year of Diagnosis	*Total No. of New Cases*	*Cases Dead*
1981	2	2
1982	7	7
1983	27	25
1984	58	56
1985	120	113
1986	211	198
1987	401	374
1988	540	482
1989	631	537
1990	682	577
1991	763	644
1992	775	587
1993	1935	908
1994	947	259
1995	259	63

Source: Department of Health HIV/AIDS Surveillance Program, Columbus, OH, www.odh.state.oh.us.

3. Review the data in Table 1–A–3 and answer the questions that follow.
   a. What is the male-to-female ratio for AIDS in Ohio? In the United States?
   b. Out of the total number of AIDS cases in Ohio, what proportion are women? Of the total cases in the United States, what proportion are women?
   c. What proportion of the total AIDS cases in Ohio are ages 30 to 39? What proportion in the United States are ages 30 to 39?
   d. Calculate the proportion of AIDS cases in Ohio by race. Calculate the proportion of AIDS cases in the United States by race.
   e. How do the preceding ratios and proportions, Ohio versus United States, compare?

Table 1–A–3 Ohio AIDS Cases by Age, Race, and Sex, as of June 30, 2003; U.S. AIDS Cases 1981–1999

Demographics	*Total Ohio*	*Total U.S.*
Age		
<13	96	8,718
13–19	72	3,725
20–24	331	25,904
25–29	776	97,676
30–39	4,686	329,066
40–49	5,362	190,087
50–64	2,254	68,196
65+	217	10,002
Subtotal	13,794	733,374
Race/Ethnicity		
White	6,943	318,354
Black	5,742	272,881
Hispanic	642	133,703
Other	74	7,479
Unknown	393	957
Subtotal	13,794	733,374
Sex		
Male	10,766	609,329
Female	2,634	124,045
Unknown	394	
Subtotal	13,794	733,374

Source: Ohio HIV/AIDS Statistical Summary, HIV Infection and AIDS Cases Diagnosed through June 2003, Ohio Department of Health, www.odh.state.oh.us

US DHHS, Public Health Service, CDC, National Center for HIV, STD, and TB Prevention, AIDS Public Information Data Set, CDC WONDER On-line Database, wonder.cdc.gov

4. Complete the columns in Table 1–A–4.
   a. Compute the age-specific death rates for whites and blacks.
   b. Compute the 2001 overall crude death rate for the state of California and the crude death rates for whites and blacks.
   c. Compute the 2001 age-adjusted death rates for whites and blacks in the state of California using the standardized method.
   d. Is there a difference between the age-adjusted mortality rates for whites and blacks? If so, explain the reason for the discrepancy.

Table 1–A–4 Age-Specific Mortality Rates, State of California, 2001

Age	*(a) White Pop.*	*(b) Deaths*	*(c) White ASDR*	*(d) Black Pop.*	*(e) Deaths*	*(f) Black ASDR*	*(g) Comb. Pop. Total*	*(h) Expected No. of Deaths Whites (g × c)*	*(i) Expected No. of Deaths Blacks (g × f)*
<1	428,238	2,131		33,774	435		462,012		
1–4	1,565,447	413		170,587	80		1,736,034		
5–9	2,120,923	291		240,189	45		2,361,112		
10–14	2,084,668	311		244,031	55		2,328,699		
15–19	1,929,503	1,129		208,006	185		2,137,509		
20–24	1,916,977	1,569		186,458	274		2,103,435		
25–34	4,123,447	3,399		373,455	644		4,496,902		
35–44	4,318,242	7,394		415,178	1,258		4,733,420		
45–54	3,554,132	13,766		304,914	2,339		3,859,046		
55–64	2,201,539	18,939		176,743	2,647		2,378,282		
65–74	1,531,032	33,192		11,657	3,517		1,542,689		
75–84	1,119,160	59,115		62,592	3,998		1,181,752		
85+	395,512	57,200		20,542	2,906		416,054		
Total	27,288,820	198,849		2,448,126	18,383		29,736,946		

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-line Database, wonder.cdc.gov.

5. Calculate the odds ratio for the data in Table 1–13. Interpret the results.

6. At Critical Care Hospital, the complication rate for hip replacement surgery is 8.96%. The relevant statistics appear in Table 1–A–5. The administrative staff at the hospital is concerned that the hospital complication rate does not compare favorably with the overall complication rate of all patients with hip replacement surgery in the county. The complication rate for the county is 5.5%. The county complication rate for patients age 65 or older is 8.0%; for those under age 65, the complication rate is 3.0%. Using the indirect method of standardization, calculate the complication rate for the hospital that has been adjusted for age.

Table 1–A–5 Critical Care Hospital, Hip Replacement Surgery

Age Group	*No. of Patients*	*No. of Patients with Complications*	*Complication Rate*
≥ 65	170	17	10.00%
< 65	42	2	4.76%
Total	212	19	8.96%

7. The overall mortality rate for patients who have had a cerebrovascular accident (CVA) is 15.8% at CGH. You have been asked to compare the hospital's mortality rate to that of the state. Using the data provided in Table 1–A–6, calculate the age-adjusted death rate and the standard mortality ratio (SMR) for the hospital, using the indirect method of standardization. Explain the results.

Table 1–A–6 Mortality Rates for CVAs, State versus City General Hospital

Severity of Illness	*State Mortality Rate*	*Hospital Discharges for CVA*	*Observed Deaths*	*Expected Deaths*
1	4.2	55	2	
2	5.9	116	8	
3	7.8	195	20	
4	20.9	147	29	
5	34.6	62	32	
		575	91	

INTERNET ACTIVITY

An important skill for the health information management professional is the ability to search the Internet for information. This can be particularly useful when one is searching for comparative information. This activity is designed to provide experience working with an on-line interactive database and to provide experience analyzing and summarizing the results of data queries.

Instructions

1. The Utah Department of Health has an on-line interactive database that is available for public use. The database is constructed from the Uniform Hospital Discharge Data Set (UHDDS). Information on DRGs and ICD-9-CM codes (International Classification of Diseases, 9th revision, Clinical Modification) can be obtained through queries. The public data set contains data for the years since 1992. The website address ishlunix.hl.state.ut.us/.
2. Once at the site, click "Descriptive Statistics." This should take you to the Utah Hospital Discharge Query System. The Utah External Injury Data System will also be accessed.
3. Answer the questions that follow. An alternative website is the Centers for Disease Control and Prevention's data sets at http://wonder.cdc.gov.

Questions

1. For the diagnosis of acute myocardial infarction, ICD-9-CM category 410:
   a. Prepare a bar graph that displays the number of deaths due to AMI by year, 1998 through 2002.
   b. Prepare a line graph that displays the number of deaths by gender for the years 1998 through 2002. What are your conclusions?
   c. Prepare a table that displays the number of deaths due to AMI by age group. Use the table to prepare a bar graph of the same information.
   d. Prepare a bar graph that displays the average length of stay by gender for the years 1992 through 2002. What are your conclusions after reviewing the data?
   e. Prepare a line graph that displays the median charges by year, 1992 through 2002. What does the graph indicate?

2. How many patients with coronary atherosclerosis, ICD-9-CM category 414, had a coronary artery bypass graft (CABG) procedure, ICD-9-CM procedure category 36?
   a. Prepare a line graph that displays both the total number of discharges with a principal diagnosis of coronary atherosclerosis and the total number with the CABG procedure.
   b. Construct a bar graph that displays average length of stay, by year and gender, for patients with coronary atherosclerosis and CABG procedure for the years 1998–2002.
   c. Construct a bar graph that displays the number of CABG procedures, by gender, for the years 1998–2002. What are your conclusions?

3. Determine the number of patient discharges with pathological fractures, ICD-9-CM code 733.1, by year, 1998 through 2002, and by gender. You are interested in patients aged 65 years and over. Prepare a line graph displaying the number of discharges by year and by gender. Discuss your findings.

4. In table form, how many patients were discharged, by year, 1998 through 2002, and by gender, with malignant neoplasms of the trachea, bronchus, and lung? Use the selection option that is available on the database. Prepare a bar or line graph that displays the percentage of patients, by gender, who expired from these illnesses.

5. For ICD-9-CM code 185, for the years 1998 through 2002:
   a. Prepare a bar graph or pie chart, by third-party payer, of men, aged 45 and older, discharged with a diagnosis of prostate cancer.
   b. Prepare a bar graph that displays the number of men, by age group, discharged with prostate cancer.
   c. Discuss your findings.

6. For patients discharged with pneumonia during the years 1998 through 2002 (use the selection option that is available on the database):
   a. Prepare a table that reports the average length of stay for patients discharged with pneumonia by year and by gender. Include only patients who are aged 65 years and older.
   b. Prepare a bar or line graph to display your results.
   c. Discuss your findings.

CHAPTER 2

Graphic Display of Data

KEY TERMS

- Tables
  - Table shell
  - Box head
  - Stub
  - Cell
  - Note
  - Source
- Bar charts
  - Grouped bar chart
  - Stacked bar chart
  - 100% component bar chart
- Pie chart
- Histogram
- Frequency polygon
- Line graph
- Scatter diagram

LEARNING OBJECTIVES

At the conclusion of this chapter, you should be able to:

1. Define key terms.
2. Determine which graphic technique is appropriate for the type of information to be conveyed.
3. Determine the appropriate graphic techniques for the various scales of measurement.
4. Outline the essential components of tables.
5. Correctly prepare tables for one, two, three, and/or four variables.
6. Outline the principles for construction of bar charts and pie charts.
7. Differentiate between the following types of bar charts: one-variable bar chart, grouped bar charts, stacked bar charts, and 100% component bar charts.
8. Correctly prepare bar charts and pie charts.

9. Correctly prepare the following types of graphs: histograms, frequency polygons, line graphs, and scatter diagrams.
10. Differentiate between histograms and bar charts.
11. Differentiate between frequency polygons and line graphs.

The purpose of tables, charts, and graphs is to summarize and display data clearly and effectively. They are all means of summarizing quantities of information to the reader. Tables, charts, and graphs offer the opportunity to analyze data sets and to explore, understand, and present distributions, trends, and relationships in the data. The primary purpose of tables, charts, and graphs is to communicate information about the data to the user.

Whether the graphic technique used is a table, chart, or graph, it should

- display the data
- allow the viewer to think about what the data convey
- avoid distortion of the data
- encourage the reader to make comparisons
- reveal data at several levels, from a broad overview to the fine detail
- serve a reasonably clear purpose: description, exploration, tabulation, or decoration
- be closely related to the statistical and verbal descriptions of the data set

CONSTRUCTION OF TABLES

A **table** is an orderly arrangement of values that groups data into rows and columns. Almost any type of quantitative information can be grouped into tables. For example, we can use tables to display frequencies for such vital statistics as morbidity rates or hospital admission and discharge data. Tables are useful for demonstrating patterns and other kinds of relationships. They also serve as a basis for more visual displays of data, such as graphs and charts, where some of the detail may be lost. Because tables generally do not capture the interest of the reader, they should be used sparingly.

Table Shells

Although data cannot be analyzed until they have been collected, it is useful to prepare a **table shell** that shows how the data will be organized and displayed. It also helps one work through the data collection process in advance to ensure that once the data have been collected they can be analyzed in the manner desired. The basic shell for construction of tables appears in Exhibit 2–1. Table shells are tables that are complete except for the data. In summary, a table should be self-explanatory, even if it is taken out of its original context. A table should convey all the information necessary for the reader to understand the data. Check the table to be sure that

- It is a logical unit.
- It is self-explanatory. Ask yourself if the table can stand on its own if photocopied and removed from its context.
- All sources are specified.
- Headings are specific and understandable for every column and row.
- Row and column totals are checked for accuracy.
- Cells are not left blank; enter "0" or "-".
- Categories are mutually exclusive and exhaustive.

Exhibit 2–1 Table Shell

TITLE

Box Head		*Sex*					
		Male		*Female*		*Total*	
	Age	*No.*	*%*	*No.*	*%*	*No.*	*%*
Stub		Row Variable →→→→		→→→→		→→→→	
	<45			Column Variable			
	45–54			↓			
	55–64			↓			
	65–74			↓			
	75+			↓			

Note:

Source: Adapted from *Self-Instructional Manual for Cancer Registries, Book 7: Statistics and Epidemiology for Cancer Registries*, p. 23, United States Department of Health and Human Services, Public Health Service, National Institutes of Health, National Cancer Institute.

Consideration should also be given to alignment of data in tables. Guidelines for aligning text include the following: align text in a table to the left; text that serves as a column label may be centered; numeric values should be aligned to the right. If the numeric values contain decimals, they should be decimal-aligned. Some word processing and microcomputer statistical software programs have features that assist in the formatting of tables. The essential components of a table are outlined in Exhibit 2–2.

Exhibit 2–2 Essential Components of Tables

TITLE	The title should be as complete as possible and should clearly relate the content of the table. It should answer the following questions: • What are the data? (e.g., counts, percentages) • Who? (e.g., white females with breast cancer; black males with lung cancer) • Where are the data from? (e.g., hospital, state, community) • When? (e.g., year, month) For example: Site distribution by Age and Sex of Cancer Patients upon First Admission to General Hospital
BOX HEAD	The box head contains the captions or column headings. The heading of each column should contain as few words as possible but should explain briefly exactly what the data in the column represent.
STUB	The row captions are known as the stub. Items in the stub should be grouped to facilitate interpretation of data. For example, group ages into five-year intervals.
CELL	The box formed by the intersection of a column and a row.
Optional Items:	
NOTE	Anything in the table that cannot be understood by the reader from the title, box head, or stub should be explained by notes. Notes contain numbers, preliminary or revised numbers, or explanations for any unusual numbers. Definitions, abbreviations, and/or qualifications for captions or cell names should be footnoted. A note usually applies to a specific cell(s) within the table, and a symbol, such as ** or #, can be used to key the cell to the note. If several notes are required, it is better to use small letters than to use symbols for numbers. Note that numbers may be confused with the numbers within the table.
SOURCE	If data from a source outside your research are used, the exact reference to the source should be given. Indicating the source lends authenticity to the data and allows the reader to locate the original source if more information is needed.

Source: Adapted from *Self-Instructional Manual for Cancer Registries, Book 7: Statistics and Epidemiology for Cancer Registries*, p. 24, United States Department of Health and Human Services, Public Health Service, National Institutes of Health, National Cancer Institute.

A One-Variable Table

The most basic table is a frequency distribution with just one variable. The first column shows the values or categories of the variable represented by the data, such as age or sex. The second column shows the number of persons or events that fall into each category. A third column may be added to show the percentage of persons or events in each category. Because of rounding, column totals for percentages often add up to 99.9% or 100.1%. Even when this occurs, the total given should be 100.0%, with a footnote explaining that the difference is due to rounding. An example of a one-variable table, which displays some hypothetical admissions data, is presented in Table 2–1. The variable is sex, which is divided into two mutually exclusive categories, male and female.

Table 2–1 XYZ Hospital Admissions by Sex

Sex	*Admissions*	*%*
Male	30	60.0%
Female	20	40.0%
Total	50	100.0%

Two- and Three-Variable Tables

We can use tables to display data that have more than one variable. Data can be tabulated to show counts by two or three variables, such as age and sex. A two-variable table that is cross-tabulated is usually called a two-by-two contingency table. In Table 2–2, "lung cancer patients" is classified on two variables, race and sex; race is the row variable and sex is the column variable. Contingency tables, which will be discussed in greater detail in Chapter 9, are often used in calculating measures of association such as chi square.

Table 2–2 XYZ Hospital, Lung Cancer Patients by Race and Sex, 2004

	Sex		
Race	*Male*	*Female*	*Total*
White	316	204	520
Black	35	15	50
Total	351	219	570

Tables 2–3 and 2-4 are examples of data classified on three and four variables. Because tables classifying data on more than two variables can be quite confusing to the reader, they should be avoided if at all possible.

In a three-way classification table, it sometimes becomes quite challenging to arrange the data in a readable format. A multidimensional relationship must be shown in a two-dimensional space. In Table 2–3, we have expanded the classification of the lung cancer data to include not only race/ethnicity and sex but also geographic region. The row categories are first divided by geographic region, and then by race/ethnicity. In Table 2–4, a three-way table, types of cancer are first divided by primary site and then classified by race, and sex.

Table 2–3 Age-Adjusted Rates by Geographic Location, Race/Ethnicity and Sex, Cancer of the Colon and Rectum, 1997–2001

		Sex	
Geographic Region	*Race/Ethnicity*	*Male*	*Female*
San Francisco–Oakland	White	61.4	44.3
	Black	61.7	49.2
	American Indian/Alaska Native	7.4	0.0
	Asian or Pacific Islander	51.6	39.0
	Hispanic	60.9	38.6
Connecticut	White	70.9	52.3
	Black	64.6	51.7
	American Indian/Alaska Native	18.3	60.1
	Asian or Pacific Islander	60.2	6.0
	Hispanic	71.7	52.0
Detroit	White	69.2	48.6
	Black	80.4	60.5
	American Indian/Alaska Native	22.6	12.9
	Asian or Pacific Islander	47.6	27.1

Source: Ries, L.A.G., Eisner, M.P., Kosary, C.L., Hankey, B.F., Miller, B.A., Clegg, L., Mariotto, A., Feuer, E.J., and Edwards, B.K., (eds). SEER Cancer Statistics Review, 1975–2001. National Cancer Institute, Bethesda, MD, http://seer.cancer.gov.

Table 2–4 Number of New Cancer Cases, 1997–2001, by Selected Primary Site, Race and Sex, SEER Geographic Areas

	All Races			*Whites*			*Blacks*		
Site	*Total*	*Male*	*Female*	*Total*	*Male*	*Female*	*Total*	*Male*	*Female*
Oral & Pharynx	18,688	12,452	6,216	14,890	9,917	4,973	1,888	1,329	559
Liver	10,395	7,027	3,368	6,731	4,468	2,263	1,081	759	322
Pancreas	18,790	9,214	9,576	14,999	7,424	7,575	2,116	996	1,120
Lung & Bronchus	105,298	58,192	47,106	86,163	46,518	39,645	11,314	6,815	4,499
Hodgkin's Lymphoma	5,101	2,780	2,321	4,313	2,345	1,968	505	277	228
Colon & Rectum	91,850	46,176	45,674	74,274	37,348	36,926	9,744	4,118	4,626

Source: Ries, L.A.G., Eisner, M.P., Kosary, C.L., Hankey, B.F., Miller, B.A., Clegg, L., Mariotto, A., Feuer, E.J., and Edwards, B.K., (eds). SEER Cancer Statistics Review, 1975–2001. National Cancer Institute, Bethesda, MD, http://seer.cancer.gov.

CHARTS

Graphs and charts of various types are the best means for presenting data for quick visualization of relationships. Graphs and charts emphasize the main points and analyze and clarify relationships between variables that may otherwise remain elusive.

Regardless of the type of graph or chart being prepared, several principles of construction should be followed. First, it is important to avoid distortion of the data. To avoid distortion, the representation of numbers on the graph should be directly proportional to the numerical quantities that are being represented on the graph. It is also important to consider proportion and scale. Graphs should accommodate the eye in that they should emphasize the horizontal. Graphs should be greater in length than they are in height. The three-quarter high rule is a useful guide: the height (*y*-axis) of the graph should be three-fourths the length (*x*-axis) of the graph. A longer horizontal axis helps to point out the causal variable in more detail. Other helpful hints in preparing graphs or charts include spelling out abbreviations in a note so that misunderstandings are avoided; using colors to help clarify groupings that may appear in the graph; and using both upper- and lowercase letters in titles, as the use of all capital letters can be unfriendly to the eyes.

There are many types of charts. We will first discuss the construction of charts for data that fall into categories.

Bar Charts

One-Variable Bar Chart

We can use **bar charts** to display data for one or more variables. Bar charts are appropriate for displaying data that are categorical. The simplest bar chart is the one-variable bar chart. Each category of the variable is represented by a bar. In Figure 2–1, the bar represents one variable—the crude death rate for cancer of the trachea, bronchus, and lung—which is placed in categories by years (1990 through 1998). There is one bar representing the crude death rate for each of the nine years in the bar chart. Guidelines for construction of a bar chart are summarized in Exhibit 2–3.

Figure 2–1 Crude Death Rate by Year, Cancers of the Trachea, Bronchus, and Lung, ICD-9-CM Codes 162.0–162.9, 1990–1998

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-Line Database, wonder.cdc.gov.

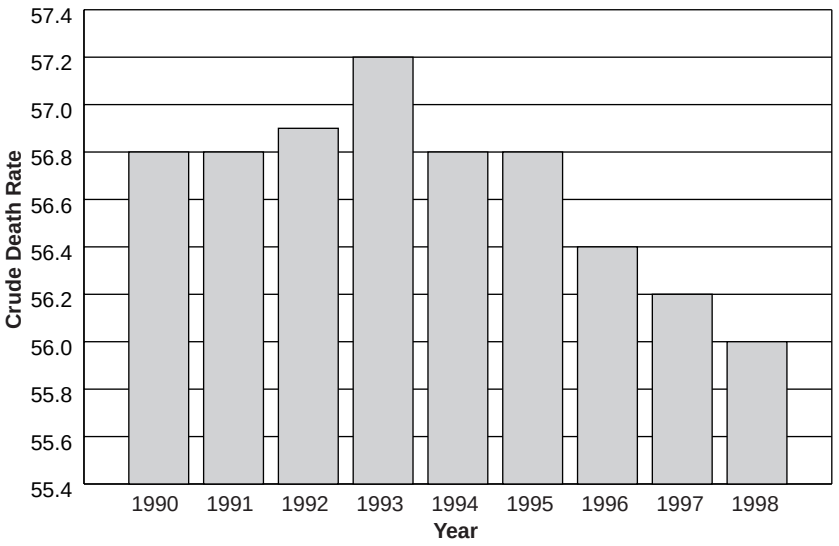


Exhibit 2–3 Guidelines for Constructing a Bar Chart

When constructing a bar chart, keep the following points in mind:

- Arrange the bar categories in a natural order, such as alphabetical order, order by increasing age, or an order that will produce increasing or decreasing bar lengths.
- The bars may be positioned vertically or horizontally.
- The bars should be of the same width.
- The length of the bars should be in proportion to the frequency of the event.
- Avoid using more than three bars (categories) within a group of bars.
- Leave a space between adjacent groups of bars but not between bars within a group.
- Code different variables by differences in bar color, shading, cross-hatching, and so on. Include a legend that interprets your code.

Source: Adapted from *Self-Instructional Manual for Cancer Registries, Book 7: Statistics and Epidemiology for Cancer Registries,* p. 251, United States Department of Health and Human Services, Public Health Service, National Institutes of Health, National Cancer Institute.

The length or height of each bar is proportional to the number of persons or events in the category. The presentation of the information in this bar chart makes it easy to see at a glance that the crude death rate was the greatest in 1993 and the least in 1998. One can also readily see that the crude death rate has been on the decline since 1995.

Bar charts may be drawn either horizontally or vertically. Figure 2–2 presents the same information that appears in Figure 2–1, but in a horizontal format. Personal preference determines the format used.

It is not uncommon to confuse a bar chart with a histogram. A bar chart is used to display data that fall into groups or categories, whereas histograms are used to illustrate frequency distributions of continuous variables. In a bar chart, the bars that represent the categories of the variables are separated, whereas in a histogram the bars are joined. A **histogram** is used to display the frequency distribution of a continuous variable, such as age. A bar chart is used to display the frequency distribution of a variable that is discrete with noncontinuous categories such as race or sex.

Computer software makes it easy to present bar charts in either two- or three-dimensional form. When bars are presented in three-dimensional form, it is sometimes difficult for the reader to estimate the true height of the bar. In a 3-D bar chart, the back edges of the bar are higher than the front edge, as in Figure 2–3. To make sure that the reader correctly interprets the bar, label the data points at a point on the bars, as shown in Figure 2–3.

Figure 2–2 Crude Death Rate by Year, Cancers of the Trachea, Bronchus, and Lung, ICD-9-CM Codes 162.0–162.9, 1990–1998

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-Line Database, wonder.cdc.gov.

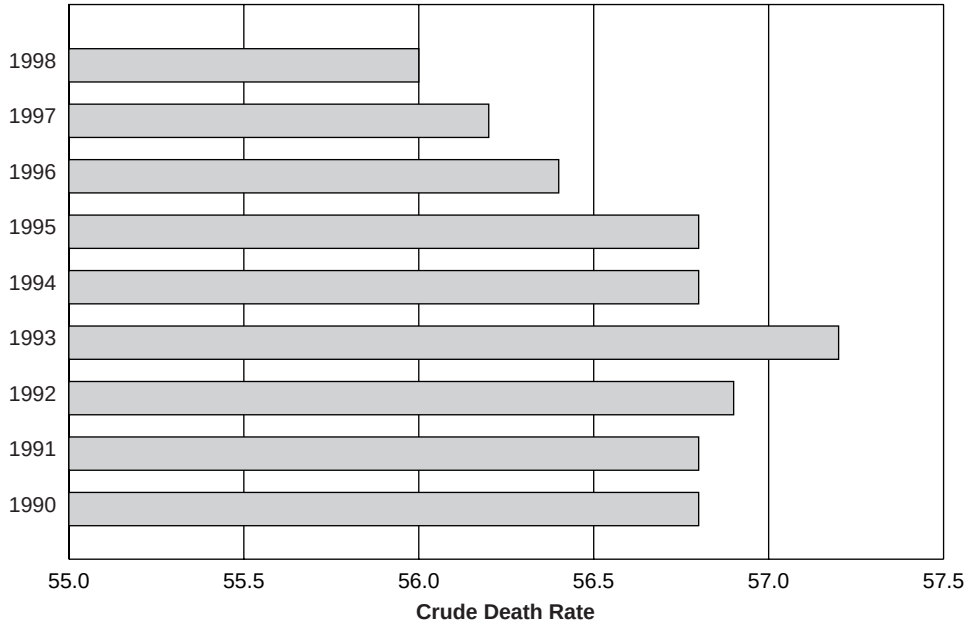
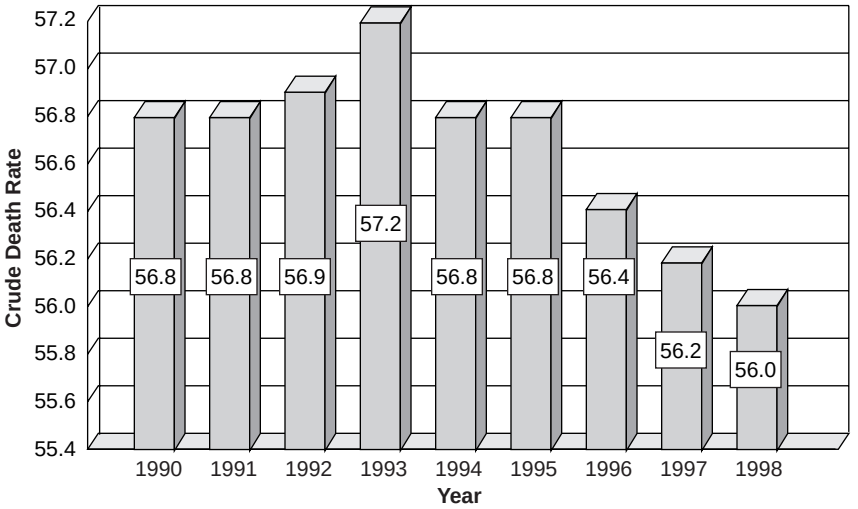


Figure 2–3 Crude Death Rate by Year, Cancers of the Trachea, Bronchus, and Lung, ICD-9-CM Codes 162.0–162.9, 1990–1998

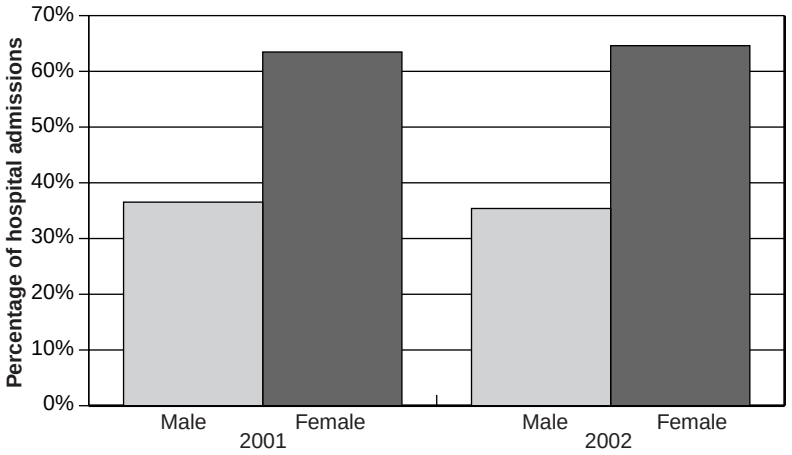
Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-Line Database, wonder.cdc.gov.



Grouped Bar Charts

A **grouped bar chart** is used to display information from tables containing two or three variables. An example of a grouped bar chart can be demonstrated by the variable sex, which has two categories: male and female. Bars within a group are usually joined; in this case, the grouping is by year. The number of bars within a grouping should be limited to three. There must also be a legend to indicate what categories the bars represent. In viewing the grouped bar chart in Figure 2–4, we can easily see that proportionately, more women than men were admitted to the hospital for the years 2001 and 2002.

Figure 2–4 Percentage of Hospital Admissions by Sex, 2001 and 2002



Stacked Bar Charts

In a **stacked bar chart**, bar segments for each data category are stacked like building blocks on top of one another to form a single bar. In a stacked bar chart, the bar represents the total number of cases that occurred in a category; the segments of the bar represent the frequency of cases within the category. As an example, the data that appear in Table 2–5 are presented as a stacked bar chart in Figure 2–5. Each bar in the stacked bar chart represents the total number of cancer cases for a specific primary site; the bar segments represent the number of males and the number of females affected within the total number of cases.

Stacked bar charts should be used with caution, since they are very difficult to interpret. Except for the bottom category, the categories do not rest on a flat baseline. What this means is that where one category of the variable ends, the next begins. Each category rides the bumps of those below it.

From the stacked bar graph in Figure 2–5, it can be readily seen that except for cancer of the colon/rectum and pancreas, men are affected more often than women. But the exact number of cases for women in each category is difficult to determine. Stacked bar charts are deceptive, so they are often used to exaggerate or hide information.

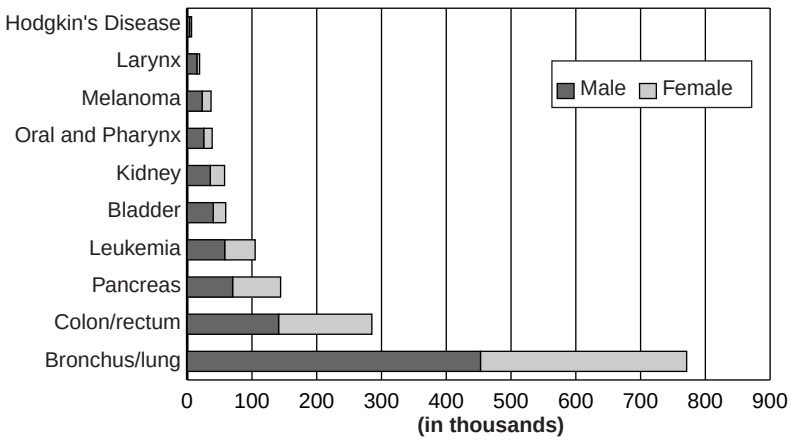
Table 2–5 Number of Deaths by Selected Primary Cancer Sites, 1997–2001

Primary Site	*Male*	*Female*	*Total*
Bronchus/lung	452,846	318,281	771,127
Colon/rectum	141,124	144,007	285,131
Pancreas	70,155	74,069	144,224
Leukemia	57,916	47,037	104,953
Bladder	40,211	19,265	59,476
Kidney	35,649	22,059	57,708
Oral & Pharynx	25,532	13,005	38,537
Melanoma	23,119	13,727	36,846
Larynx	15,069	4,080	19,149
Hodgkin's Disease	3,723	3,044	6,767
Total	865,344	658,574	1,523,918

Source: Ries, L.A.G., Eisner, M.P., Kosary, C.L., Hankey, B.F., Miller, B.A., Clegg, L., Mariotto, A., Feuer, E.J., and Edwards, B.K., (eds). SEER Cancer Statistics Review, 1975–2001. National Cancer Institute, Bethesda, MD, http://seer.cancer.gov.

Figure 2–5 Number of Deaths by Selected Primary Cancer Sites, 1997–2001

Source: SEER Cancer Statistics Review, 1975–2001. National Cancer Institute, Bethesda, MD, http://seer.cancer.gov.

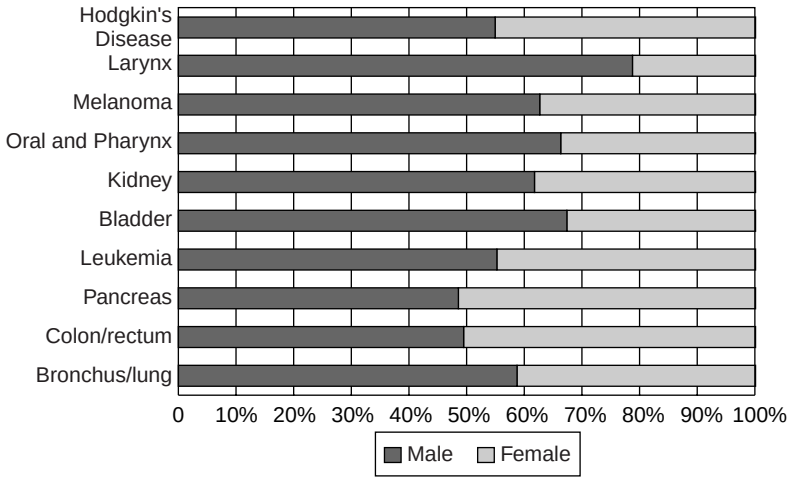


100% Component Bar Charts

The **100% component bar chart** is a variant of the stacked bar chart. In a 100% bar chart, all of the bars are of the same height and show the variable categories as percentages of the total rather than the actual values. Each bar is much like its own pie chart. A set of 100% bar charts can be used instead of multiple pie charts. This is more advantageous because it is easier to make a comparison between bars than between pies. Figure 2–6 presents the same information that appears in Figure 2–5. The stacked bars for each year represent 100% of the various types of cancer cases by sex. Each category of the sex variable is represented in terms of a percentage, in one bar.

Figure 2–6 Percentage of Cancer Deaths by Selected Primary Sites by Sex, 1997–2001

Source: SEER Cancer Statistics Review, 1975–2001. National Cancer Institute, Bethesda, MD, http://seer.cancer.gov.



Pie Charts

A **pie chart** is an easily understood chart in which the sizes of the slices show the proportional contribution of each part of the pie. We can use pie charts to show the component parts of a single group or variable. To calculate the size of each slice of the pie, first determine the proportion of the pie to be represented by each slice. Multiply the proportion by 360—the total number of degrees in a circle. The result will be the size of each slice in degrees.

In the pie chart in Figure 2–7, one slice of the pie, acute myeloid leukemia, represents 34% of the cases, or 34% of the pie. Within the pie chart, this slice of the pie equals 122.4° (360° × 0.34 = 122.4°). All other leukemia types represents 30% of the cases, and the size of its respective slices is 108° (360° × 0.30 = 108°). One category of the pie, chronic lymphoid leukemia, represents 20% of the cases, which is equivalent to 72° (360° × 0.20 = 72°). The remaining slices of the pie represent chronic myeloid leukemia, 9 percent and 32.4° of the pie, and acute lymphoid leukemia, 7 percent and 25.2° of the pie. The sum of the degrees for each slice of the pie is 360° (122.4° + 108° + 72° + 32.4° + 25.2° = 360°). The pie chart in Figure 2–7 demonstrates how the whole pie is divided into segments. By convention, the largest slices of the pie begin at 12 o'clock, as in Figure 2–7. The slices of the pie should be arranged in some logical order. In Figure 2–8, acute lymphoid leukemia appears in the 12 o'clock position. This is an example where pie slices are arranged in alphabetical order rather than according to magnitude.

It is not recommended to use pie charts to compare multiple distributions because they are not optimal for comparing components for more than one group. When components of more than one group are to be compared, a 100% component bar chart should be used.

Figure 2–7 Leukemia Cancer Deaths by Type, 1997–2001, Ordering by Magnitude of Groupings

Source: SEER Cancer Statistics Review, 1975–2001. National Cancer Institute, Bethesda, MD, http://seer.cancer.gov.

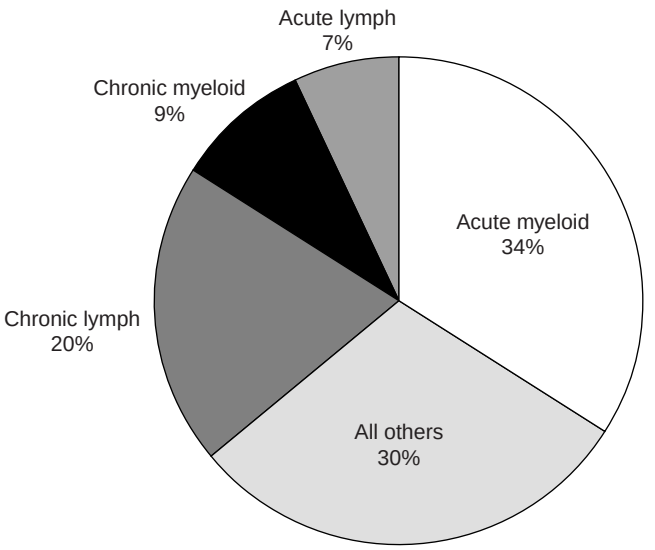
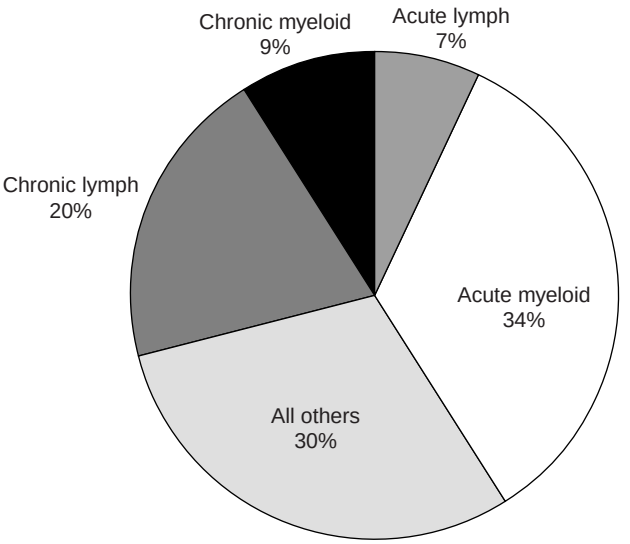


Figure 2–8 Leukemia Cancer Deaths by Type, 1997–2001, Alphabetical Order

Source: SEER Cancer Statistics Review, 1975–2001. National Cancer Institute, Bethesda, MD, http://seer.cancer.gov.



Histograms

Thus far we have discussed graphing data that are in categorical or discrete form. The techniques that will be discussed next are appropriate for data that are continuous in nature.

A histogram is appropriate for displaying a frequency distribution for one continuous variable. The frequency distribution can be presented in either number or percentage form. A histogram consists of a series of bars, each having as its base one class interval and as its height the number (frequency) or percentage of cases in that class. A class interval is a type of category; a class interval can represent one value in a frequency distribution (Figure 2–9) or a group of values in a frequency distribution (Figure 2–10).

In this type of graph, there are no spaces between the bars, since the data points represented are continuous. That is, a data point may fall anywhere in the area covered by the graph. The sum of the heights of the bars represents the total number, or 100% of the cases. When the distribution of the data needs to be emphasized more than the actual values, use a histogram. An example where the class interval represents a single value in a frequency distribution is displayed in Figure 2–9. Each bar of the histogram represents a single age, in contrast to Figure 2–10, where each bar represents an age group.

Figure 2–9 DRG 416, Septicemia, Histogram of Patients Aged 80–89

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-Line Database, wonder.cdc.gov.

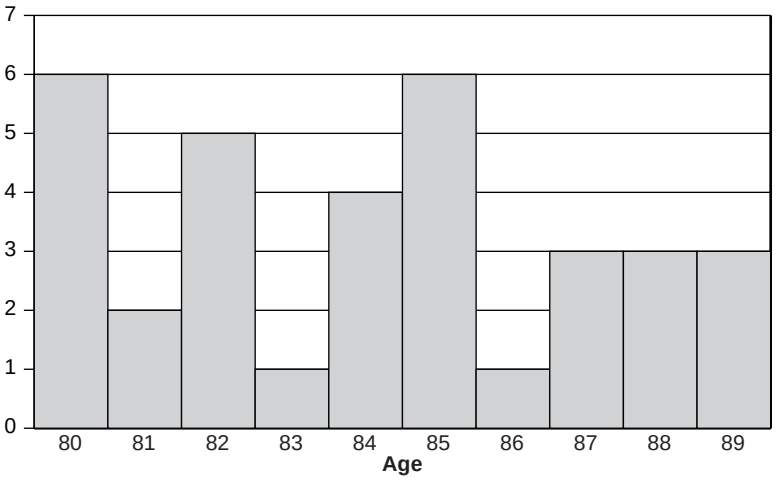
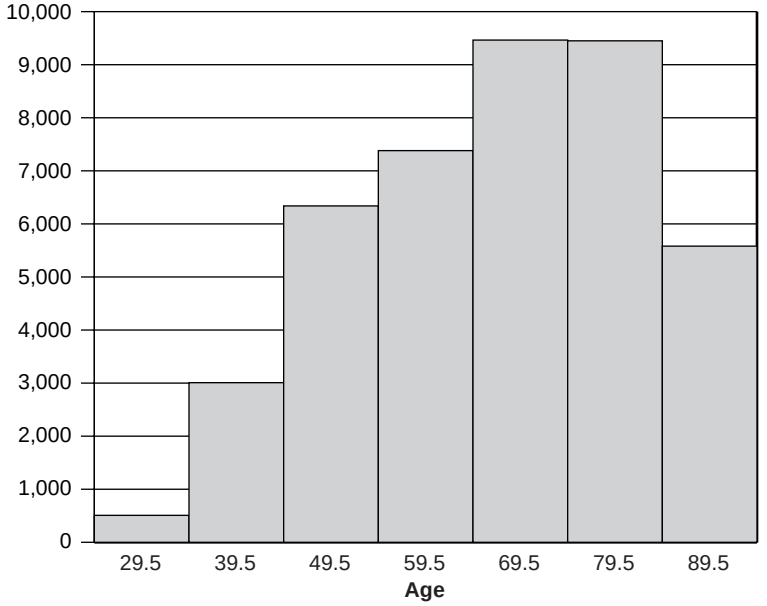


Figure 2–10 Deaths Due to Breast Cancer, ICD-9-CM Codes 174.0–174.9, by Age, 1998

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-Line Database, wonder.cdc.gov.



Age at death for breast cancer is the variable represented in the histogram in Figure 2–10. The values at the bottom of the *x*-axis are the midpoints of the class intervals for the following age groups:

Age Group	*Midpoint of Class Interval*
25–34	29.5
35–44	39.5
45–54	49.5
55–64	59.5
65–74	69.5
75–84	79.5
85+	89.5

In the histogram, it is clear that there are two age groups that account for most of the deaths due to breast cancer: 65–74 and 74–84.

Frequency Polygons

A **frequency polygon** can be used as an alternative to the histogram. Like a histogram, a frequency polygon is a graph of a frequency distribution. To construct a frequency polygon, simply join the midpoints at the top of each bar in the histogram (4.5, 14.5, 24.5, and so on). The advantage of the frequency polygon over the histogram is that several frequency polygons can be plotted on the same graph for comparison purposes. Frequency polygons also are easy to interpret.

When constructing a frequency polygon, make the *x*-axis longer than the *y*-axis to avoid distorting the data. The frequency of the observations is always placed on the *y*-axis, and the scale of the variables under study is placed on the *x*-axis. Frequency values are plotted at the midpoint of each class interval.

The frequency polygon in Figure 2–11 plots the same data that appear in the histogram in Figure 2–10. Since the *x*-axis represents the total distribution, *the line always starts and ends with zero.*

The frequency polygon tells us that the number of deaths due to breast cancer reaches its peak in the age groups 65 to 74 and 75 to 84. A frequency polygon presents this pattern with more clarity than the histogram.

Line Graphs

A **line graph** is often used to display time trends and survival curves. The *x*-axis shows the unit of time from left to right, and the *y*-axis measures the values of the variable being plotted.

A line graph does not represent a frequency distribution. A line graph consists of a line connecting a series of points on an arithmetic scale. Like all graphs, it should be designed so that it is easy to read. The selection of proper scales, complete and accurate titles, and informative legends is important. If a graph is too long and narrow, either vertically or horizontally, it has an awkward appearance and may exaggerate one aspect of the data. The upward trend for median charges for septicemia patients in the state of Utah is displayed Figure 2–12. The line graph is especially useful when there are a large number of values to

Figure 2–11 Deaths Due to Breast Cancer, ICD-9-CM Codes 174.0–174.9, by Age, 1998

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-Line Database, wonder.cdc.gov.

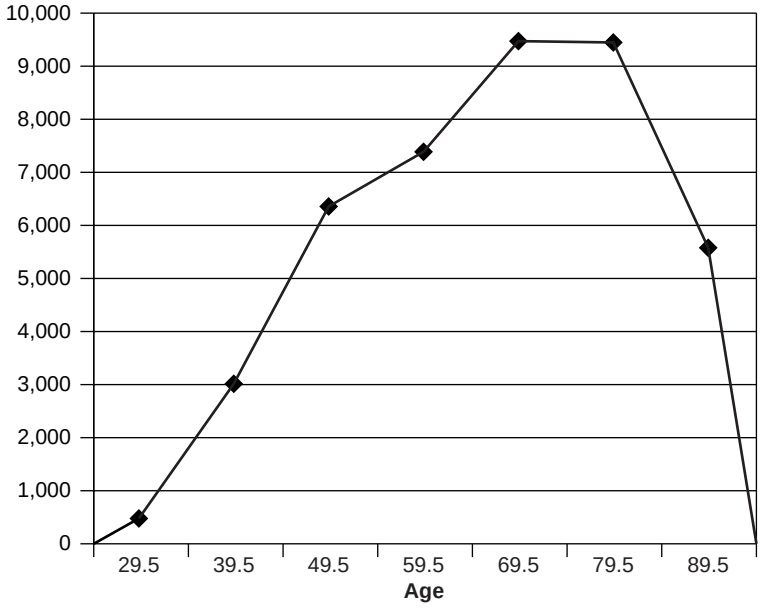
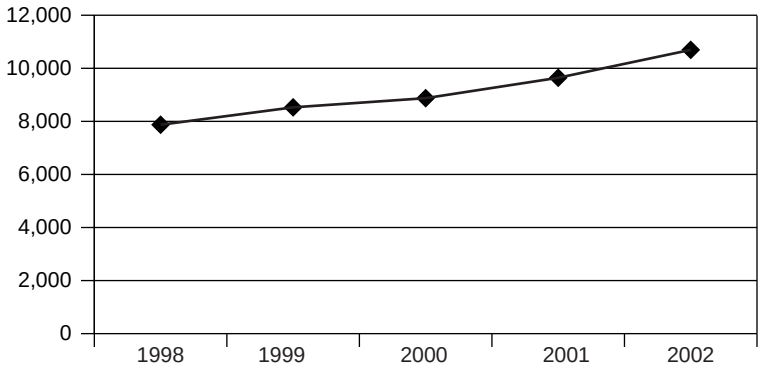


Figure 2–12 DRG 416, Septicemia Age ≥ 17, Median Charges, State of Utah, 1998–2002

Source: Utah Inpatient Hospital Discharge Dataset, Utah Office of Health Care Statistics, www.health.state.ut.us.



be plotted; that is, when you have a continuous variable with an unlimited number of possible points. It also allows the presentation of several sets of data on one graph.

Either actual numbers or percentages may be used on the *y*-axis of the line graph. Use percentages on the *y*-axis when more than one distribution is to be shown on one graph.

A percentage distribution allows comparisons between groups where the actual totals are different.

If more than one set of data is plotted on the same graph, different types of lines (solid or broken) should be used to distinguish between the lines. The number of lines should be kept to a minimum—a line graph can soon become too cluttered. Each line should be identified in a legend or on the graph itself.

There are two kinds of time-trend data: (1) point data, which reflect an instant in time, and (2) period data, which cover an average or total over a specified period of time, such as a one-year or five-year time frame. In point data, the scale marker on the *x*-axis indicates a particular point in time, such as one, two, or three years of survival. On the other hand, in plotting of period data, the horizontal scale lines are used to indicate the interval limits, and the values are plotted at the midpoint at each interval. For example:

Year of Diagnosis	*Midpoint of Interval*
1986–1988	1987
1989–1991	1990
1992–1994	1993
1995–2000	1997.5

Table 2–6 presents an example of point data that are graphed in Figure 2–13. Other examples are presented in Table 2–7 and Figure 2–14.

Figure 2–13 Relative Survival Rates by Year of Diagnosis for Kidney and Renal Pelvis Cancer, 1992–1996

Source: Miller, B.A., Clegg, L., Mariotto, A., Feuer, E.J., and Edwards, B.K., (eds). SEER Cancer Statistics Review, 1975–2001, National Cancer Institute, Bethesda, MD, http://seer.cancer.gov.

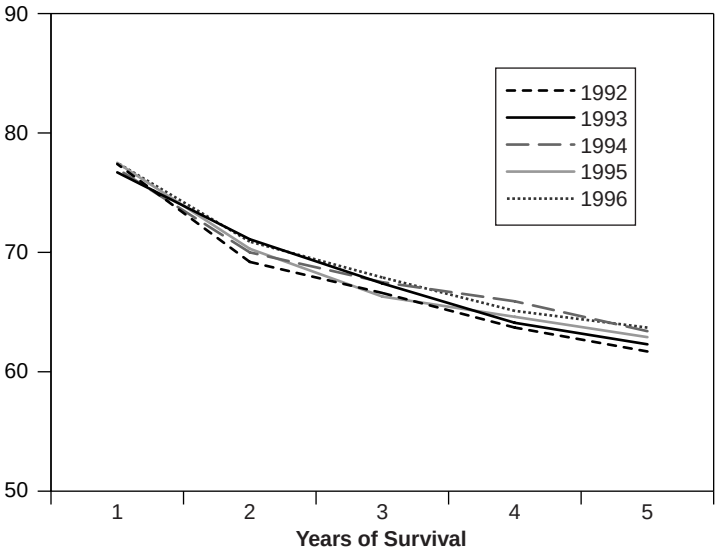


Table 2–6 Survival Rates by Year of Diagnosis, Kidney and Renal Pelvis Cancer, 1992–1996

Years of Survival	*1992*	*1993*	*1994*	*1995*	*1996*
1	77.4	76.7	77.0	77.5	77.5
2	69.2	71.1	70.0	70.3	70.9
3	66.6	67.4	67.5	66.3	67.9
4	63.7	64.1	65.9	64.6	65.1
5	61.7	62.3	63.4	62.9	63.7

Source: Ries, L.A.G., Eisner, M.P., Kosary, C.L., Hankey, B.F., Miller, B.A., Clegg, L., Mariotto, A., Feuer, E.J., and Edwards, B.K., (eds). SEER Cancer Statistics Review, 1975–2001. National Cancer Institute, Bethesda, MD, http://seer.cancer.gov.

Figure 2–14 Five-Year Survival Rates for Kidney and Renal Pelvis Cancer for Patients Diagnosed 1986–1988, 1989–1991, 1992–1994, 1995–2000

Source: Miller, B.A., Clegg, L., Mariotto, A., Feuer, E.J., and Edwards, B.K., (eds). SEER Cancer Statistics Review, 1975–2001, National Cancer Institute, Bethesda, MD, http://seer.cancer.gov.

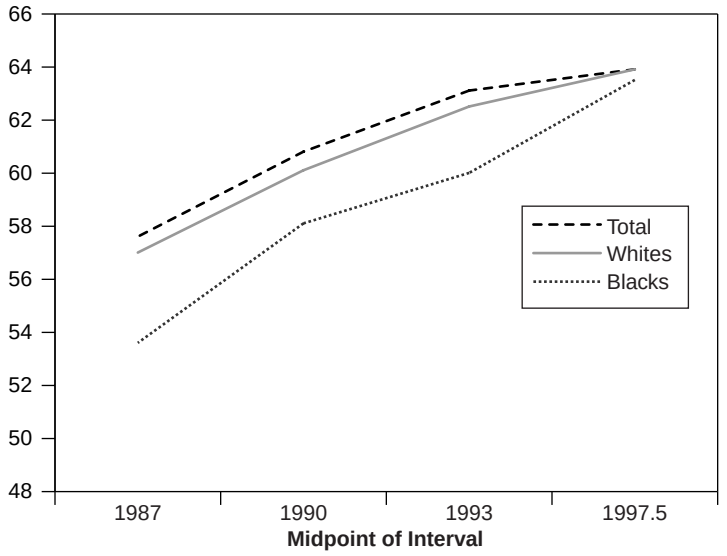


Table 2–7 Five-Year Survival Rates for Kidney and Renal Pelvis Cancer for Patients Diagnosed 1986–1988, 1989–1991, 1992–1994, 1995–2000

		Race		
Year of Diagnosis	*Midpoint of Interval*	*Total*	*Whites*	*Blacks*
1986–1988	1987	57.0	57.6	53.6
1989–1991	1990	60.1	60.8	58.1
1992–1994	1993	62.5	63.1	60.0
1995–2000	1997.5	63.9	63.9	63.5

Source: Ries, L.A.G., Eisner, M.P., Kosary, C.L., Hankey, B.F., Miller, B.A., Clegg, L., Mariotto, A., Feuer, E.J., and Edwards, B.K., (eds). SEER Cancer Statistics Review, 1975–2001. National Cancer Institute, Bethesda, MD, http://seer.cancer.gov.

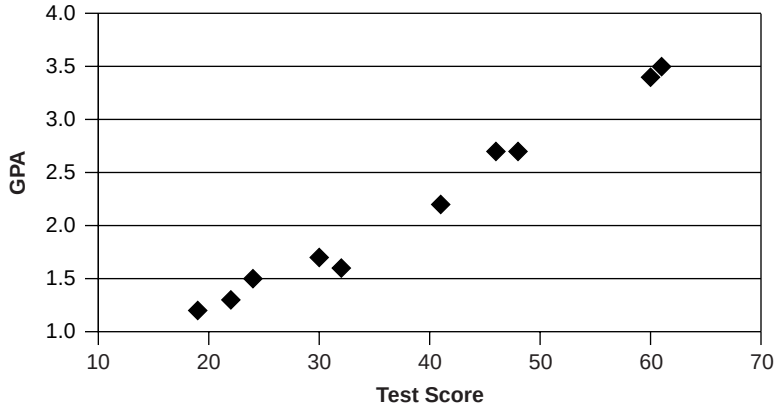
Scatter Diagrams

A **scatter diagram**, or scatter plot, is a graphic technique used to display the relationship between two continuous variables. One variable is plotted on the *x*-axis and the other is plotted on the *y*-axis. To create a scatter diagram, there must be a pair of values for every person, group, or other entity in the data set, one value for each variable. Each pair of values is plotted by placing a point on the graph where the two values intersect. To interpret a scatter diagram, analyze the overall pattern of the plotted points. Plotted points that appear to fall in a straight line indicate a linear relationship between *x* and *y*, whereas widely scattered points indicate no relationship between *x* and *y*. Table 2–8 presents hypothetical data of test scores and the grade point averages of 10 students. Figure 2–15 is a scatter plot that depicts the relationship between the two variables, test scores (*x*) and grade point average (*y*).

Table 2–8 Test Scores and Grade Point Averages of 10 Students

Student	*Test Score (x)*	*GPA (y)*
1	24	1.5
2	61	3.5
3	30	1.7
4	48	2.7
5	60	3.4
6	32	1.6
7	19	1.2
8	22	1.3
9	41	2.2
10	46	2.7

Figure 2–15 Scatter Diagram of Hypothetical Test Scores and Grade Point Average



The scatter diagram in Figure 2–15 indicates a strong linear relationship between the variables test scores and grade point average. Scatter diagrams are used to assist in interpretation of inferential statistics such as correlation and linear regression. We will discuss these topics in Chapter 8.

CONCLUSION

Tables, charts, and graphs are effective methods of summarizing data and displaying data in a clear, concise format. Tables are often used to display data, and they can be used to display data about one or more variables. An advantage of tables is that large amounts of data can be displayed and summarized, as in a four-way table. However, if too much information is included in a table, it can be confusing to the reader.

Bar charts or graphs are often used for displaying categorical data, but they are appropriate for data that are continuous in nature. Bar charts allow for quick visualization of the variable of interest. Relationships between two variables are also easily seen in a bar chart. Bar charts may take the form of a simple bar chart, a grouped bar chart, a stacked bar chart, or a 100% component bar chart. The form selected should be appropriate to the data and easily interpreted by the reader.

Pie charts are useful for displaying the parts of a whole. For example, we could display the proportion of patients admitted by third-party payer, or the proportion of burn patients admitted by severity of burn. Pie charts should be used to display proportions of one nominal-level variable; pie charts are not appropriate for comparing distributions of two or more variables.

Histograms and frequency polygons are used to display the frequency distribution of one continuous variable. Histograms and frequency polygons represent 100% of the cases in a frequency distribution; the shape of the distribution can be easily seen in these two types of graphs.

Line graphs are used to display trends in data; they are also used in survival analysis. A line graph consists of a line connecting a series of points on an arithmetic scale. To avoid distortion in the data, the graph should not be too long or too narrow. When constructing bar graphs and line graphs, the three-quarters-high rule should be used as a guide to avoid data distortion. Either actual numbers or percentages may be displayed in a line graph.

ADDITIONAL RESOURCES

Al-Assaf, A.F., and Schmele, J.A., eds. 1993. *The textbook of total quality management.* Delray Beach, FL: St. Lucie Press. 124–125.

Jones, G.E. 1995. *How to lie with charts.* San Francisco: Cybex, Inc.

Longo, D., and Bohr, D., eds. 1991. *Quantitative methods in quality management.* Chicago: American Hospital Association. 10–11.

Ries, L.A.G., Eisner, M.P., Kosary, C.L., Hankey, B.F., Clegg, L., Mariotto, A., Feuer, E.J., and Edwards, B.K., eds. *SEER Cancer Statistics Review, 1975–2001.* National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr.

SEER's Web-based Training Modules. U.S. National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) Program. http://training.seer.cancer.gov.

Self instructional manual for cancer registries, Book 7: Statistics and epidemiology for cancer registries. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, NIH Publication No. 94-3766. 1994.

Surveillance, Epidemiology, and End Results (SEER) program (www.seer.cancer.gov). *SEER Stat Databases: Incidence.* National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2004, based on the November 2003 submission.

U.S. Department of Health and Human Services, Public Health Service. 1992. *Principles of epidemiology: An introduction to applied epidemiology and biostatistics.* Atlanta, GA: USDHHS.

Utah Hospital Discharge Data, Public Dataset, www.health.state.ut.us.

Appendix 2–A

Exercises for Solving Problems

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
2. The purpose of a table, chart, or graph is to communicate information to the data user. What questions should be considered to accomplish this objective?
3. What questions should be answered in the title of a table, chart, or graph?
4. What points should be considered when constructing a bar chart?
5. Describe the differences between a stacked bar chart and a 100% component bar chart.
6. Differentiate between a bar chart and a histogram.

MULTIPLE CHOICE

1. You want to graph the average length of stay by sex and service for the month of April. The best choice is to use a:
   a. bar graph
   b. histogram
   c. line graph
   d. pie chart
2. You want to graph the number of deaths due to prostate cancer for the years 1998 to 2002. The best choice is to use a:
   a. frequency polygon
   b. histogram
   c. line graph
   d. pie chart

3. A pie chart may be used to display the:
   a. average length of stay by year
   b. percentage of discharges by third-party payer
   c. number of discharges per year and third-party payer
   d. number of patients discharged by sex and service

4. A histogram may be used to display:
   a. discharges by age
   b. discharges by third-party payer
   c. discharges by service
   d. discharges by sex

5. You want to display the number of discharges by sex and service for 1999. The best choice is to use a:
   a. bar chart
   b. cluster line graph
   c. histogram
   d. line graph

PROBLEMS

Prepare the appropriate charts and graphs for the following problems. Include a title for each and identify the data source when indicated.

1. The admissions data in Table 2–A–1 compare actual admissions by hospital service with the budgeted number of hospital admissions for the month of January for Critical Care Hospital. Using computer graphic software, construct a bar chart that compares budgeted admissions with actual admissions. Write a short summary of the results.

Table 2–A–1 Admissions Report for January

Hospital Service	*Budgeted Admissions*	*Actual Admissions*
Medicine	769	728
Surgery	583	578
OB/GYN	440	402
Psychiatry	99	113
Physical Medicine and Rehab	57	48
Other Adult	178	191
Newborn	312	294

2. Using the data in Table 2–A–2, prepare a pie chart for January patient days by service for Critical Care Hospital.

Table 2–A–2 Patient Days by Service

Hospital Service	*Patient Days*
Medicine	4,436
Surgery	4,036
OB/GYN	1,170
Psychiatry	1,223
Physical Medicine and Rehab	1,318
Other Adult	688
Newborn	1,633

3. Table 2–A–3 contains length-of-stay data by service for the month of January for Critical Care Hospital. Construct a stacked bar chart that compares actual average length of stay with the budgeted average length of stay.

Table 2–A–3 Average Length of Stay (ALOS) by Service

Hospital Service	*Budgeted ALOS*	*Actual ALOS*
Medicine	6.39	6.09
Surgery	7.23	6.98
OB/GYN	3.22	2.91
Psychiatry	11.56	10.82
Physical Medicine and Rehab	22.98	27.46
Other Adult	3.93	3.60
Newborn	4.97	5.55

4. Organize the following statistics for the month of January into a table.

Critical Care Cancer Research Institute Statistics for January

Discharges		Discharge Service Days	
Medicine	198	Medicine	1,313
Surgery	152	Surgery	947
Gynecology	74	Gynecology	328
Otolaryngology	48	Otolaryngology	290
Average Length of Stay			
Medicine	6.6		
Surgery	6.2		
Gynecology	4.4		
Otolaryngology	6.0		

5. Exhibit 2–A–1 displays the lengths of stay for 80 patients at the Critical Care Cancer Research Institute. Construct a histogram of these data.

Exhibit 2–A–1 Lengths of Stay for 80 Patients

1	2	3	5	6	10	13	16
1	2	3	5	6	10	13	17
1	2	3	5	6	10	13	17
1	2	4	5	6	10	14	17
1	2	4	5	8	11	14	18
1	2	4	5	8	11	14	19
1	2	4	5	8	11	14	19
1	2	4	6	8	11	15	20
2	3	4	6	8	12	15	20
2	3	5	6	9	12	16	20

6. The average charges for malignant neoplasms of the large intestine and colon and average charges for malignant neoplasms of the trachea, bronchus, and lung appear in Table 2–A–4. Prepare a line graph that compares average charges for both malignancies.

Table 2–A–4 Average Total Charges, Cancers of the Colon and Lung, State of Utah, 1995–2002

Year	*Cancer of the Trachea, Bronchus, and Lung*	*Cancer of the Colon and Rectum*
1995	$13,364	$15,275
1996	$15,403	$14,622
1997	$14,958	$16,511
1998	$15,232	$16,659
1999	$16,331	$19,633
2000	$19,618	$20,424
2001	$21,216	$21,646
2002	$22,256	$22,362

Source: Utah Inpatient Hospital Discharge Dataset, Utah Office of Health Care Statistics, www.health.state.ut.us.

7. Review the data in Table 2–4. Determine the percentage of total male and female cancer cases for each site. Prepare a bar chart to display your results.

CHAPTER 3

Introduction to Measurement

KEY TERMS

- Measurement
- Validity
  - Content validity
  - Construct validity
  - Criterion-related validity
- Sensitivity
- Specificity
- Predictive value
- Reliability
  - Stability
  - Internal consistency
  - Interrater agreement
- Kappa coefficient
- Timeliness
- Scales of measurement
  - Nominal
  - Ordinal
  - Ratio
  - Interval
- Continuous variable

LEARNING OBJECTIVES

At the conclusion of this chapter, you should be able to:

1. Define key terms.
2. Relate the importance of validity and reliability to the measurement process.
3. Differentiate between the following aspects of validity: content, construct, and criterion related.
4. Differentiate between the following aspects of reliability: stability, internal consistency, and interrater agreement.
5. Determine the sensitivity, specificity, and predictive value of a test.
6. Compare and contrast the following scales of measurement: nominal, ordinal, ratio, and interval.
7. Classify variables to the scales of measurement.
8. Identify variables as either discrete or continuous.

WHAT IS MEASUREMENT?

In the daily operations of any organization, whether in business, manufacturing, or health care, data are collected for decision making. To be effective, decision makers must have confidence in the data collected. Confidence requires that the data collected be accurate, timely, and reliable. Assurance of these aspects of data quality is what the **measurement** process is about. Our discussion of analysis of clinical data in health care begins with the topic of measurement. We cannot accurately collect, analyze, and make conclusions regarding clinical data without an understanding of the measurement process. In measurement, we are concerned with measuring an attribute or property of a person, object, or event according to a particular set of rules. In some cases, measurement may be direct, as when we are using a yardstick to measure the property "width of a desk" (object). However, in health care we do not always have the ability to measure persons, objects, or events directly. For example, "quality of care" cannot be measured directly with a yardstick or any other measuring device. Consequently, an indirect measure must be developed.

Whether we are dealing with direct or indirect measures, the result of the measurement process is numbers, and there must be a set of rules for assigning numbers to the objects being measured. The measurement process requires rigorous definition of what will be collected, and the method by which it will be collected, so that the resulting numbers will be meaningful, accurate, and informative. The advantage of standardizing the process is that the results are the same regardless of who is doing the data collection. The resulting uniformity also allows for comparisons between and within institutions. Through measurement, one creates data that can be analyzed using statistical techniques, and that can be presented as meaningful information. Through the measurement process, we transform data into information.

To illustrate, consider the process by which the data for the Caesarean section (C-section) rate is collected. The C-section rate is often used as a performance indicator for hospital obstetric services, especially by managed care plans. A low C-section rate is desirable, since C-sections are more expensive than vaginal deliveries, and there is a longer recovery time for the patient. The first step in the process is to define C-section rate, the event to be measured. Now we must consider the properties that characterize this event—in other words, what data will be collected? The procedure for measuring this event is presented in Exhibit 3–1. As you can see, calculation of the C-section rate is a straightforward process.

However, quantifying some of the performance indicators developed by the Joint Commission on Accreditation of Healthcare Organizations (Joint Commission) and the Agency for Healthcare Research and Quality (AHRQ) is not always so easily accomplished. Since some performance indicators cannot be directly measured, we must use some attribute that is considered to represent the presence or absence of the attribute of interest—an indirect measure. As an analogy, consider the personal attribute of intelli-

Exhibit 3–1 Measurement of Caesarean Section Rate

Caesarean Section Rate (event): The ratio during any given time period of surgical deliveries (Caesarean sections) to the total number of deliveries.

Data To Be Collected (Properties):

1. Total number of Caesarean sections performed for a given period
2. Total number of deliveries for the period

Data Sources:

1. Medical records of discharged obstetrical patients for the period
2. Daily discharge reports for the period
3. Disease and operations indexes of the International Classification of Disease, 9th Revision, Clinical Modification (ICD-9-CM)
4. Labor and delivery room logs

Calculation: $$\frac{\text{Total no. of Caesarean sections for period}}{\text{Total no. of deliveries for same period}} \times 100$$

gence. Intelligence is a personal attribute that cannot be directly measured with a yardstick or with any type of device; instead it is indirectly measured through a standardized IQ test, which consists of questions that are assumed to be representative of intelligence.

Exhibit 3–2 outlines the procedure for collecting data on the performance indicator "esophageal resection mortality rate." This is a performance indicator that measures outcome. The occurrence of this event is considered an indicator of an undesirable outcome and should be avoided. But in calculating the mortality rate, we must first decide who will be counted. Do we count everyone who had an esophageal resection? Do we count only those admitted for esophageal cancer? Do we count those who were diagnosed with esophageal cancer after admission? Standardization of the process helps ensure that everyone is counting the same type of cases. As you can see in Exhibit 3–2, we are interested in counting patients who have a principal or secondary diagnosis of esophageal cancer with resection of the esophagus. This occurrence, to be measured, must be rigorously defined, and procedures for data collection must be strictly followed.

Exhibit 3–2 Example of Measurement of An Inpatient Quality Indicator

Measure: Esophageal Resection Mortality Rate

Type of Measure: Provider level, Mortality indicator for Inpatient Procedures (outcome)

Rationale: Esophageal cancer surgery is a rare procedure that requires technical proficiency, and errors in surgical technique or management may lead to clinically significant complications such as sepsis, pneumonia, anastomotic breakdown, and death.

Relationship to Quality: Better processes of care may reduce mortality for esophageal resection, which represents better quality care.

Definition: Number of deaths per 100 patients with discharge procedure code of esophageal resection.

Numerator: Number of deaths with a code of esophageal resection in any procedure field.

Denominator: Discharges with ICD-9-CM codes of 42.40 through 42.42 in any procedure field and a diagnosis code of esophageal cancer in any field.

Exclude patients transferring to another short-term hospital, MDC 14 (pregnancy, childbirth, and puerperium), and MDC 15 (newborns and neonates).

Face Validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

The primary evidence for esophageal resection mortality as an indicator arises from the volume-outcome literature. The causal relationship between hospital volume and mortality is unclear, and the differing processes that may lead to better outcomes have not been identified.

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

Esophageal resection is a relatively uncommon procedure. Patti et al.* noted that most hospitals perform 10 or fewer procedures during a 5-year period. The precision of this indicator may be improved by using several years data.

Construct Validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

There is no evidence for the construct validity of esophageal resection beyond the volume-outcome relationship.

*Patti, M.G., Corvera, C.U., Glasgow, R.E., et al. A hospital's annual rate of esophagectomy influences the operative mortality rate. *J Gastrointestinal Surg* 1998, 2(2): 186–92.

Source: AHRQ Guide to Inpatient Quality Indicators: Quality of Care in Hospitals—Volume, Mortality, and Utilization. United States Department of Health and Human Services, Agency for Healthcare Research and Quality, June 2002, Revision 2, September 4, 2003, www.ahrq.gov

VALIDITY

Accuracy in measurement cannot happen without **validity**. The measuring instrument, whether a ruler, an IQ test, or a survey instrument, is considered valid if it measures what it is intended to measure and for the intended purpose. A ruler or scale is a direct measure. In

health care, because we cannot measure quality with yardsticks and scales, quality is often assessed through indirect measures.

As an example, the quality measure "Diabetes mellitus: hospital admission rate for long-term complications" developed by AHRQ is displayed in Exhibit 3–3. This is a community measure that focuses on the prevention of "hospital admissions for ambulatory care-sensitive conditions." Effective management of these patients in an ambulatory setting may prevent hospitalization for diabetes mellitus, a chronic condition. Thus, this performance measure is serving as an indirect measure for access to a certain type of ambulatory care in a given community. Inpatient admissions are serving as the proxy measure.

Exhibit 3–3 AHRQ Quality Indicator—Diabetes Mellitus: Hospital Admission Rate for Long-Term Complications

Measure: Diabetes Mellitus: Hospital Admission for Long-Term Complications

Source: AHRQ quality indicators. Guide prevention quality indicators: hospital admission for ambulatory care sensitive conditions.

Description: This indicator assesses the number of admissions for long-term diabetes per 100,000 population.

Rationale: Long-term complications of diabetes mellitus include renal, eye, neurological, and circulatory disorders. Long-term complications occur at some time in the majority of patients with diabetes to some degree. Proper outpatient treatment and adherence to care may reduce this incidence of long-term complications, and lower rates represent better quality of care.

Numerator: Discharges, age 18 years and older, with ICD-9-CM principal diagnosis code for long-term complications of diabetes (renal, eye, neurological, circulatory, or complications not otherwise specified). Patients transferring from another institution, MDC14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and neonates) are excluded.

Denominator: Population in Metropolitan Statistical Area (MSA) or county, age 18 years and older.

Source: AHRQ Guide to Inpatient Quality Indicators: Quality of Care in Hospitals—Volume, Mortality, and Utilization. United States Department of Health and Human Services, Agency for Healthcare Research and Quality, June 2002, Revision 2, September 4, 2003, www.ahrq.gov

To implement this performance measure, we must now ask what qualifies as a long-term diabetic complication. Is each health care facility free to define what qualifies as a diabetic complication? Are both Type I and Type II diabetics to be included? What about complications of gestational diabetes? It should become obvious that if we are to have valid data, qualifying complications must be rigorously defined. Without rigorous definition of the measure, consistent, comparable data will not result from the data collection process, and confidence in the information will be lost. There are many types of validity, but we will limit our discussion to three: content, construct, and criterion-related validity.

Content validity is the adequacy of the sample, or the number of items used to represent the content area being measured. It is the extent to which the instrument makes sense in terms of the property or attribute being measured. This is often referred to as content validity. For example, if we are interested in assessing an individual's competency in ICD-9-CM coding, a sample of ten items from the vast domain of coding guidelines and principles is inadequate. It is not logical that ten items could adequately measure an individual's ability to apply ICD-9-CM guidelines and principles to all body systems. By its very nature, content validity is a matter of judgment and is evaluated by a panel of experts.

Construct validity is the ability of an instrument to measure the selected property or attribute of interest. How do we know that an intelligence test actually measures intelligence, or that a survey instrument on patient satisfaction actually measures patient satisfaction? Construct validity is considered to be the link between theory and the property/attribute under study. A statistical technique of factor analysis is often used to evaluate construct validity. In Exhibit 3–2, Esophageal Resection Mortality Rate, there is a statement regarding the construct validity of this measure—"there is no evidence for construct validity of esophageal resection beyond the volume-outcome relationship." That is to say, the mortality rate may not in itself be a measure of quality care.

In **criterion-related validity**, we are applying a known criterion or gold standard to the measurement instrument. It is assessed by correlating the measure of interest with an external criterion that is known to measure the property of interest. Criterion-related validity can be either concurrent or predictive. An example of criterion-related validity that is predictive is the Scholastic Achievement Test. This test is used to predict the success of an individual in the first year of college. Linear regression may be used to evaluate predictive validity.

SENSITIVITY, SPECIFICITY, AND PREDICTIVE VALUE OF A MEASURE

Sensitivity, specificity, and **predictive value** are aspects of data accuracy. They assist in evaluating the validity of a measure, especially the indirect performance indicators that are often used in health care. A measure is sensitive to the extent that it identifies every case in which the property of interest is truly present (cell *a*, Table 3–1). If the measure is not sensitive, it will not detect the property of interest when it is present (cell *c*). Specificity is the aspect of measurement that results in exclusion of cases when the property of interest is truly absent (cell *d*). If the measurement is not specific, it will falsely detect the property of interest when it is not present (cell *b*). The accuracy of a test or measure is dependent upon the number of false positives and false negatives that occur as a result of using the measure. If the test is accurate, the number of false positives and false negatives will be low. The predictive value is the proportion of positive tests that are truly positive ($a/a + b$). The predictive value of a positive test increases as sensitivity and specificity increase.

To determine the specificity and sensitivity of a proposed measure, a pilot test of the measure may be conducted, with the results displayed in a table such as the one shown in Table 3–1. In this table, the rows represent the true situation—the presence or absence of the performance indicator. The columns represent the possible results of the measure for the

Table 3–1 Assessing Sensitivity and Specificity of a Measure

	True Situation/Event		
Test/Measure	*Performance Indicator Present*	*Performance Indicator Absent*	*Total*
Positive	*a*	*b*	$a + b$
Negative	*c*	*d*	$c + d$
Total	$a + c$	$b + d$	$a + b + c + d$

Sensitivity: $a/(a + c)$
Specificity: $d/(b + d)$
Predictive Value: $a/(a + b)$: the predictive value of a positive
$d/(c + d)$: The predictive value of a negative

There are four possible outcomes, represented by the four cells:
Cell *a*: true positives—the variable of interest is present and the measure reveals its presence
Cell *b*: false positives—the measure indicates the variable of interest to be present, but it is incorrect
Cell *c*: false negatives—the variable of interest is present, but the measure does not reveal it
Cell *d*: true negatives—the variable of interest is absent, and the measure indicates that it is absent

Source: Adapted from Introduction to Clinical Reasoning—Evidence Based Medicine. Medical University of South Carolina. http://www.musc.edu/dc/icrebm/sensitivity.html

performance indicator of interest. The test is positive when it tells us that the performance indicator is present and negative when it tells us that the performance indicator of interest is not present.

A test is accurate to the extent that it does not result in false positives and false negatives. A large number of false positives will lead to the unnecessary examination of cases where no problem exists. In time, this will lead to the loss of the data's credibility. A large number of false negatives can result in overlooking cases that actually contain the performance indicator of interest.

An example may help clarify assessment of sensitivity and specificity: A quality improvement team wants to evaluate the performance measure on diabetic long-term complications because it has not been tested for reliability and validity by the AHRQ (Exhibit 3–3). The team knows in advance that diabetes with the specified complications, such as diabetic peripheral vascular disease and diabetic ulcers, was the reason for admission in 20 cases. How effective is our test in identifying these 20 cases? The results of the pilot test are displayed in Table 3–2.

Table 3–2 Assessing Sensitivity and Specificity of Admission for Diabetes with Specified Complication

	Specified Complication Present		
Admission for Diabetes with Specified Complication	*Yes*	*No*	*Total*
Yes	15	10	25
No	5	70	75
Total	20	80	100

In this example, diabetes mellitus with the specified complications occurred in 20 cases ($a + c = 15 + 5$). The measure correctly identified 15 of the 20 cases; thus, the sensitivity for the measure is 0.75 ($a/a + c$). The test incorrectly identified 10 patients as having the problem. These are false positives. The test correctly identified 70 of the 80 cases that had no problem; thus the specificity is 0.88 [$d/b + d = 70/(10 + 70)$]. The test also failed to identify 5 cases where there were problems; these are false negatives.

Validity is the extent to which a measure actually measures a property or attribute one wants to measure. In this example, the measure correctly identified 15 of the 25 cases in which the diabetic complication was present; thus, the predictive value is 0.60 [$a/a + b = 15/(15 + 10)$]. The test correctly measures the property of interest 60% of the time.

Sensitivity, specificity, and predictive value quantify the accuracy of the quality measure. With a sensitivity of 0.75, the measure misses 25% of the cases in which we are interested. The predictive value is 0.60; thus, 40% of the time the measure tells us that there may be a problem with performance when there is none. In this example, the measure is not highly accurate.

RELIABILITY

Error is integral to the measurement process, whether it is the measurement of weight, height, or blood pressure. Even when measurement is made as accurately as the instrument allows and all procedures are followed, repeated measures do not always give exactly the same results. This is because error is a component of all measurement. However, an instrument that is reliable will tend to have results that are consistent with each other over repeated trials. A measurement process is said to be reliable if repeated measurements over time on the same property or attribute give the same or approximately the same results. A yardstick is reliable because it will provide approximately the same result every time an object is measured, regardless of who is doing the measuring. An example of an unreliable measuring device is a scale that gives widely different weights each time the same object is measured.

Measures should also be unbiased. A measurement process is unbiased if it does not systematically overstate or understate the true value of the attribute/property being measured. An example of a biased but reliable measure is a scale that consistently measures weight 10 pounds less than the actual weight. An unbiased measure is correct on average.

There are three aspects of **reliability**: stability, internal consistency, and interrater agreement. **Stability**, or test-retest reliability, is the extent to which the same results are obtained on repeated applications. Stability is evaluated by administering the instrument to the same group on two separate occasions—hence the name test-retest reliability. A reliability coefficient is calculated from the two sets of scores. The reliability coefficient ranges from 0 (no reliability) to 1.0 (perfect reliability).

Internal consistency measures the extent to which the items on the measurement instrument are homogeneous, or consistent with one another. Internal consistency is measured by Cronbach's alpha, which is often referred to as coefficient alpha. Coefficient alpha also has a range of 0.0 to 1.0; an instrument should have a minimum coefficient alpha of 0.70 for acceptability.

Interrater agreement is the extent to which results are in agreement when different individuals administer the same instrument to the same individuals or groups. This is an important concept in coding. If we have four coders assigning codes to peripheral vascular disease due to insulin-dependent diabetes mellitus, we want all four coders to come up with the same codes. Interrater agreement should be 100%. The kappa statistic is used to assess interrater agreement. This statistic is limited, however, because it requires that the responses be dichotomous. In the coding example, the codes would be evaluated as either right or wrong.

In using kappa, we are interested in determining the extent of agreement, not disagreement, between raters. Kappa is obtained by the following formula:

$$K = A_o - A_E/N - A_E$$

where A_O is the number of observations that are in agreement and A_E is the number of observations in agreement expected by chance alone.

A_O is obtained by summing the table diagonals ($A_O = a + d$), and the percentage of agreement is obtained by A_O/N. A_E is obtained by multiplying the row total times the column total and dividing by the grand total, as follows:

$$A_E = \sum_{i=1}^{k} \text{(row marginal)(column marginal)}/N$$

The range for the **kappa coefficient** is -1.0 to $+1.0$. A negative value indicates that the proportion of agreement that resulted from chance is greater than the proportion of observed agreement. High positive values of kappa indicate strong interrater agreement. Let us now consider an example where two raters have been asked to evaluate the accuracy of a simple random sample of 200 medical records that have been coded. The coded charts were evaluated as right or wrong. The results appear in Table 3–3.

Table 3–3 Results of Coding Assessments by Two Raters

	Rater 2		
Rater 1	*Right*	*Wrong*	*Total*
Right	80	20	100
Wrong	25	75	100
Total	105	95	200

To obtain the number of observations that are in agreement between the two raters, A_O, the diagonals are summed: 80 + 75 = 155. The percent agreement between the two raters is 77.5% (155/200). The number of observations in agreement that is expected by chance, A_E, is obtained by

$$\begin{aligned} A_E &= \sum_{i=1}^{k} (\text{row marginal})(\text{column marginal})/N \\ &= (100)(105)/200 + (100)(95)/200 \\ &= 52.5 + 47.5 \\ &= 100 \\ K &= A_O - A_E)/(N - A) \\ &= 155 - 100)/(200 - 100) \\ &= 0.55 \end{aligned}$$

The interpretation of kappa is that the obtained agreement of 77.5% is approximately 55% greater than what would be achieved by chance alone. Interpretation of the kappa statistic is as follows:

$$\begin{gathered} k < 0.20 \text{ negligible} \\ 0.20 \leq k < 0.40 \text{ minimal} \\ 0.40 \leq k < 0.60 \text{ fair} \\ 0.60 \leq k < 0.80 \text{ good} \\ k \geq 0.80 \text{ excellent} \end{gathered}$$

In our example, the obtained kappa coefficient, 0.55, indicates that the raters obtained fair agreement in their assessment of the sample of coded records.

In summary, a measure is reliable if (1) the same individual measures the same attribute at another time and achieves the same result, (2) the individual applies the measure consistently on all attributes, and (3) several individuals reach the same result when measuring the same attribute.

TIMELINESS

A critical aspect of data quality is **timeliness**. One must consider whether the data are collected and available within a useful time frame. Does the collected information represent the current condition of the patient or state of the organization? Data collected today are more critical and useful to the decision maker than data collected yesterday or two weeks ago. The value of collected data decreases with time.

For data to be accurate, the measure creating the data must be rigorously defined. A common failure in the measurement process is to undertake the measurement process without a clear definition of what is to be measured, who will use the resulting information, and how the information will be used for making decisions. The measurement process must begin with a strong commitment to identifying the information needs of those who will be using the information.

SCALES OF MEASUREMENT

As stated previously, all measurement results in a number. To properly assign numbers to the measurement results, one must understand the **scales of measurement**. Understanding the differences in the scales of measurement is also critical to make appropriate use of graphic displays of data, to correctly select statistical techniques for data analysis, and to facilitate data entry.

Nominal Scale

The **nominal** scale is the lowest level of measurement. In the nominal scale, measures are organized into categories; there is no recognition of order within these categories. Examples of categories on the nominal scale of measurement are sex, religion, and third-party payer. To facilitate computer analysis, numbers are assigned to nominal variable—for example, male = 1, female = 2. The numbers assigned to the various categories carry no numerical weight; the numbers merely serve as labels for the categories.

Statistical summaries of variables on the nominal scale may be prepared through the use of a frequency distribution. On the nominal scale, a frequency distribution would indicate the number of cases falling into each category. Using statistical notation, the frequency distribution would be represented as follows:

$$N = \sum_{k} f_k$$

where N is the total number of cases, f_k is the frequency within category, and $\sum_{k}$ is a summation over all categories k.

As a hypothetical example, consider the nominal variable "method of payment." For initial patient data collection, options may be commercial insurance ($k = 1$), managed care ($k = 2$), Medicare ($k = 3$), Medicaid ($k = 4$), self-pay ($k = 5$), and other ($k = 6$). Thus, there are six categories. A frequency distribution for patients admitted by method of payment is

presented in Table 3–4. In our example, the total number of patients admitted during the week of July 23 is 256 ($N = 256$), and the number of Medicare patients (frequency) who were admitted is 62 ($f_3 = 62$).

Table 3–4 Number of Patients Admitted by Method of Payment, Week of July 23

Commercial Insurance	*Managed Care*	*Medicare*	*Medicaid*	*Self-Pay*	*Other*	*Total*
75	43	62	55	6	15	256

Sometimes it is useful to show the proportion of cases falling into a category k. This proportion is designated a p_k. A proportion is the number of cases that fall within a particular category divided by the total number of cases and is represented as

$$p_k = f_k/N.$$

The proportion of cases falling into each category will sum to 1. Using the same data from Table 3–4, Table 3–5 represents the proportions and cumulative proportions for each category.

Table 3–5 Proportion of Patients Admitted by Method of Payment, Week of July 23

Category	*Frequency (f)*	*Proportion (p)*	*Cum. Proportion*
Commercial Insurance	75	0.293	0.293
Managed Care	43	0.168	0.461
Medicare	62	0.242	0.703
Medicaid	55	0.215	0.918
Self-pay	6	0.023	0.941
Other	15	0.059	1.000
Total (*N*)	256	1.000	

The distribution of a variable into the various categories can also be represented by percentages, which are proportions multiplied by 100. For data that fall on the nominal scale of measurement, frequencies may be represented in tables, charts, and graphs. The most appropriate measure of central tendency for nominal-level data is the mode. For nominal data, the mode tells us the category that has the most observations.

When designing a system of data collection for nominal variables, remember that the categories must be mutually exclusive and exhaustive. That is, each data element can fall into only one category, and all possible categories must be accounted for. It may be appropriate to assign an "other" category, as in the preceding example, when the number of data elements falling into some possible categories may be too small for data analysis or when the category is not important to the purpose of the data collection.

Ordinal Scale

Some nonnumeric scales have an order to their categories; these are called **ordinal** variables. On the ordinal scale, the order of the numbers is meaningful, not the number itself, so the usual arithmetic operations are not meaningful. This is because the intervals or distance between categories are not necessarily equal. It is not appropriate to perform arithmetic operations, such as calculating averages, on ordinal variables.

Examples of ordinal variables are the numbers assigned to indicate class rank, the ordering of adjectives that describe patient condition, and a Likert-type scale that can be used to describe patient satisfaction. These examples are displayed in Table 3–6. A number may be assigned to represent the ordering of the variables. In the case of class rank, we know that an individual classified as a senior has completed more credit hours than a sophomore, but we cannot say that a senior has completed twice as many credit hours as a sophomore. The same is true regarding patient condition. A patient that is in critical condition is not necessarily twice as sick as an individual who is in stable condition. Only the order of the value is meaningful. On this scale, a patient in critical condition is sicker than a patient in guarded condition. The frequency distributions of ordinal variables may be portrayed in the same way as for nominal variables.

Table 3–6 Examples of Ordinal Variables

Class Rank	*Patient Condition*	*Patient Satisfaction*
1—Freshman	1—Resting and Comfortable	1—Strongly agree
2—Sophomore	2—Stable	2—Agree
3—Junior	3—Guarded	3—No opinion
4—Senior	4—Critical	4—Disagree
		5—Strongly disagree

Nominal and ordinal variables are considered discrete variables. Discrete variables have gaps between successive values. Diagnosis-related groups (DRGs) are examples of discrete/nominal variables.

Scales for Metric Variables

Metric variables are numeric variables that answer questions of how much or how many. Metric variables fall on one of two scales of measurement: ratio or interval. Arithmetic operations may be performed on ratio- and interval-scale measures.

- *Ratio Scale.* The **ratio** scale is the highest level of measurement. On the ratio scale there is a defined unit of measure, a real zero point, and the intervals between successive values are equal. For example, consider the variable of length. Length has defined units of measurement, such as inches, and a true zero point—0 inches. With a real zero point, statements such as "Mary is twice as tall as Jill" can be made. Multiplication on the ratio scale by a constant does not change its ratio character, but addition of a constant to

a ratio measure does. For example, if an older sibling is twice as tall as a younger sibling, and both grow 2 inches, the ratio of their heights is no longer 2:1. But if we multiply their respective heights by 2 (e.g., 60" × 2 and 30" × 2), the ratio between the two heights remains 2:1.

- *Interval Scale.* Measures that fall on the **interval** scale have a defined unit of measurement but do not have a true zero point. The most important characteristic of the interval scale is that the intervals between successive values are equal. On the Fahrenheit scale, the interval between 20°F and 21°F is the same as the interval between 21°F and 22°F. But since there is not a true zero on this scale, we cannot say the 40°F is twice as warm as 20°F.

An advantage of metric data is that they can be grouped. Interval and ratio variables are continuous. If a variable is continuous, it may take on fractional values, such as 85.3235°F. With **continuous variables**, there are no gaps between values, since the values progress fractionally. Graphic techniques that can be used to display interval and ratio data include histograms, frequency polygons, and stem and leaf plots. Measures of central tendency that may be reported for interval and ratio data are the mean, median, and mode.

The scale of measurement depends on the method of measurement, not on the attribute being measured. For example, if we score a test by summing the total number of correct answers, the resulting measures fall on the ratio scale. However, if each question is tallied according to the total number who got the question right and the total number who got the question wrong, the measures fall on the nominal scale because the measure falls into one of two categories—"right" or "wrong." When developing measures for any purpose, one must consider on what scale of measurement the collected data will fall so that the appropriate statistical procedures may be selected.

CONCLUSION

Measurement is a process that requires rigorous definition of what is being measured, the data sources, and how the variable measured will be calculated. Several aspects of a measure should be evaluated: validity, reliability, sensitivity, specificity, and predictive value.

The validity of a measure is the extent to which an instrument measures what it is intended to measure. Several aspects of validity were discussed: content validity, construct validity, and criterion-related validity. With content validity, we are interested in determining whether the number of items on an instrument adequately measure the content of interest. With construct validity, we are trying to determine if the items on an instrument actually assess the attribute of interest, such as patient satisfaction. With criterion-related validity, we are assessing whether the measure is correlated with an external criterion: for example, whether a rise in serum glutamicoxaloacetic transaminase enzyme levels is correlated with acute myocardial infarction.

Reliability is the extent to which an instrument gives the same results over time or over repeated measures. Several aspects of reliability were discussed: test-retest reliability, internal consistency, and interrater agreement. Test-retest reliability is the extent to which re-

peated administrations of an instrument provide the same results. Internal consistency is the extent to which the items on an instrument are related to one another. Interrater agreement is the extent to which different individuals who administer the same instrument achieve similar results.

Other critical aspects of a measure are sensitivity, specificity, and predictive value. Sensitivity is the ability of an instrument to detect the property of interest in every case where it exists. Specificity is the ability of an instrument to exclude cases in which the property of interest is truly absent. A test is accurate to the extent that it does not result in false positives and false negatives. Predictive value is the ability of an instrument or test to correctly measure the proportion of positive tests that are truly positive or the proportion of negative tests that are truly negative.

Finally, to understand the measurement process, we must understand the differences in the scales of measurement: nominal, ordinal, interval, and ratio. In the nominal scale of measurement, measures are organized into categories; the nominal scale is the lowest level of measurement. In the ordinal scale of measurement, there is "order" to the categories; i.e., the numbers themselves are not meaningful, but the order is. On the interval and ratio scales of measurement, the intervals between successive values are equal. Arithmetic calculations may be performed on interval and ratio measures. The interval scale differs from the ratio scale in that the interval scale does not have a true zero.

ADDITIONAL RESOURCES

Agency for Healthcare Research and Quality (AHRQ). *Guide to Inpatient Quality Indicators: Quality of Care in Hospitals—Volume, Mortality, and Utilization*. Department of Health and Human Services. June 2002, AHRQ Pub. No. 02-RO204, Revision 2, September 4, 2003. www.ahrq.gov.

Al-Assaf, A.F., and Schmele J.A., eds. 1993. *The textbook of total quality management*. Delray Beach, FL: St. Lucie Press. 124–125.

Federal Register. Volume 68, No. 229, Friday, November 28, 2003. 66975–66978.

Jekel, J., et al. 1996. *Epidemiology, biostatistics and preventive medicine*. Philadelphia: W.B. Saunders Co.

Kane, R.L., ed. 1997. *Understanding health care outcomes research*. Gaithersburg, MD: Aspen Publishers, Inc. 228–229.

Nelson, E. 1995. Measuring for improvement: Why, what, when, how, for whom? *Quality Connection 4*, no. 2:3.

Osborn, C.E. 1998. Developing instruments for assessment of patient outcomes. *Journal of Rehabilitation Outcomes Measurement 2*, no. 6:18–25.

The Statistics Homepage. StatSoft, Inc., 1984–2004. www.statsoft.com.

U.S. Department of Health and Human Services, Public Health Service. 1992. *Principles of epidemiology: An introduction to applied epidemiology and biostatistics.* Atlanta, GA: USDHHS.

Appendix 3–A

Exercises for Solving Problems

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
2. Describe the measurement process.
3. Why are validity and reliability requirements of accuracy in the measurement process?
4. Compare and contrast the following aspects of validity: content validity, construct validity, and criterion-related validity.
5. You weighed your dog this morning on your bathroom scale. His weight was 15 lb. You decided to weigh him again in the evening, and his weight had increased to 20 lb. This is most likely what type of measurement error? Explain your answer.
6. Compare and contrast the following measures of reliability: stability, internal consistency, and interrater agreement.
7. Relate the importance of determining the sensitivity, specificity, and predictive value of a screening measure.
8. Define nominal, ordinal, interval, and ratio scales of measurement.
9. What is the difference between a discrete variable and a continuous variable?

MULTIPLE CHOICE

1. You are conducting a study in which "sex" is the property of interest. The variable "sex" falls upon which of the following scales of measurement?
   a. nominal
   b. ordinal
   c. ratio
   d. interval

2. You are interested in the lengths of stay for patients of Dr. Wells for a quality improvement study. The variable "length of stay" falls upon which of the following scales of measurement?
   a. nominal
   b. ordinal
   c. ratio
   d. interval

3. You are assisting Dr. Scorch in his study of burn patients. He wants to classify the number of burn patients by severity of burn: first degree, second degree, or third degree. The variable "severity of burn" falls upon which of the following scales of measurement?
   a. nominal
   b. ordinal
   c. ratio
   d. interval

4 "Severity of burn" is an example of which of the following types of variables?
   a. continuous
   b. discrete
   c. interval
   d. ratio

Questions 5 through 8 refer to the following case:

You have designed a survey instrument with a 5-point Likert-type scale consisting of statements that measure physician satisfaction with health information management services. On the scale, 1 = strongly agree (SA) and 5 = strongly disagree (SD).

5. You decide to take an average of the scores for each statement. In this case, you would be treating the data as falling upon which of the following scales of measurement?
   a. nominal
   b. ordinal
   c. ratio
   d. interval

6. After reviewing the results, you decide that the mode might be the more appropriate measure of central tendency for these data. In reporting the mode, you are treating the data as falling upon which of the following scales of measurement?
   a. nominal
   b. ordinal
   c. ratio
   d. interval

7. You have distributed this survey to the physicians five times. You are satisfied that the results are consistent. This aspect of reliability is:
   a. internal consistency
   b. test-retest reliability
   c. interrater reliability
   d. all of the above

8. However, upon careful review of the instrument, you decide that there are not enough items to adequately cover services provided by your department. You are concerned about which of the following types of validity?
   a. content
   b. construct
   c. criterion-related
   d. all of the above
9. Which of the following variables would be considered "continuous"?
   a. age
   b. sex
   c. religion
   d. principal diagnosis code
10. Which of the following variables would be considered "discrete"?
   a. age upon admission
   b. patient length of stay
   c. DRG assignment upon discharge
   d. time spent in operating room
11. Both the specificity and sensitivity of a recently developed clinical indicator are 0.99. This means that:
   a. if the indicator is present, it has a 99% chance of testing positive by the measure
   b. if the indicator is absent, it has a 99% chance of testing negative by the measure
   c. the case has a 99% chance of being correctly classified by the measure
   d. (a) and (b)
   e. all of the above

PROBLEMS

1. The hospital readmission rate is often considered an indicator of an undesirable patient outcome. The quality improvement team is interested in reducing the number of readmissions among patients discharged with a principal diagnosis of congestive heart failure (CHF). The team believes that the high readmission rate is due to the difficulty that these patients have in controlling the number of drugs that they typically take. The team believes that by improving patient/family education regarding drug administration, the readmission rate could be reduced. Thus, they have developed the screen "CHF patients taking three or more drugs" to identify these patients before discharge. To evaluate the effectiveness of the measure, the team conducts a study on all CHF patients discharged the previous year. The results appear in Table 3–A–1. Do the following:
   a. Calculate the sensitivity, specificity, and predictive value for this measure.
   b. On the basis of your results, is this an effective measure? Why or why not?
2. Exhibit 3–A–1 displays selected data elements from the proposed Case Mix Assessment Tool for psychiatric inpatients. Identify the scale of measurement for each element, determine if the measure is discrete or continuous, and determine what measure(s) of central tendency may be used for each—mean, median, or mode.

Table 3–A–1 Readmissions of CHF Patients

	CHF Patients		
No. of Drugs Administered	*Readmitted*	*Not Readmitted*	*Total*
≥ 3 drugs	200	40	240
< 3 drugs	100	900	1,000
Total	300	940	1,240

Exhibit 3–A–1 Selected Data Elements from Case Mix Assessment Tool for Behavioral Healthcare

Data Element	*Measurement Scale Discrete/Continuous Variable*	*Measure of Central Tendency*
Gender: 1—Male 2—Female		
Number of Psychiatric Admissions		
Number of Medications		
Legal Status 1—Voluntary 2—Involuntary 3—Criminal court hold		
Unable to control substance abuse within past 30 days 0—no 1—yes		
Activities of daily living 0—Independent 1—Setup help only 2—Supervision 3—Limited assistance 4—Extensive assistance 5—Maximal assistance 6—Total dependence		
White Blood Count: criteria—range 3.8—10.8		
Number of falls in last 30 days		

Source: Federal Register, Volume 69, No. 229, Friday, November 28, 2003, p. 66975.

3. At Werethebest Hospital, 34 Caesarean sections were performed during January; at rival Weresosick Hospital, 54 Caesarean sections were performed during the month of January. During January, Werethebest Hospital had 200 deliveries; Weresosick Hospital had 1,100 deliveries. The national benchmark for the C-section rate is 15%.

The head of obstetrics at Werethebest Hospital claims that their OB service provides better care than that provided at the rival hospital. Do you agree with this assessment?

4. Develop measures for the following obstetric performance indicators proposed by the Joint Commission. Your measure should include a definition, identification of data sources, and procedures for calculation (formula).
   a. Maternal readmissions within 14 days of delivery
   b. Intrahospital maternal deaths occurring within 42 days postpartum
   c. Patients with excessive maternal blood loss

5. As part of the quality improvement team, you have prepared a report on acute myocardial infarction (AMI) mortality, which is displayed in Table 3–A–2. You query DRG 121, Circulatory Disorders with AMI and Cardiovascular (CV) complications, Discharged Alive; DRG 122, Circulatory Disorders with AMI without CV Complications, Discharged Alive; and DRG 123, Circulatory Disorders with AMI, Expired.
   a. Assess Table 3–A–2 for validity and reliability. What corrections should be made to calculate the AMI mortality rate?
   b. What is the AMI mortality rate?
   c. Review Table 3–A–3, which displays the average length of stay for each DRG. What factors should also be considered when presenting the AMI mortality rate? What is the net AMI mortality rate?

Table 3–A–2 Acute Myocardial Infarctions, DRGs 121, 122, and 123

Principal Diagnosis	DRG 121	DRG 122	DRG 123
410.01	3	1	1
410.11	9	3	4
410.31	2	1	0
410.41	11	11	1
410.61	1	0	0
410.71	66	31	9
410.91	25	11	5
421.0	1	0	0
428.0	31	0	0
Total	119	58	20

Table 3–A–3 Average Length of Stay (ALOS) and Numbers of Patients with Length of Stay (LOS) Two Days or Less, DRGs 121, 122, and 123

	DRG 121	*DRG 122*	*DRG 123*
ALOS	5.6 days	3.6 days	3.5 days
LOS $\leq$ 2 days	26	18	10

CHAPTER 4

Measures of Central Tendency and Variability

KEY TERMS

- Population
- Sample
- Population parameter
- Sample statistic
- Measures of central tendency
  - Mode
  - Median
  - Mean
    - Weighted mean
    - Trimmed mean
    - Winsorized mean
- Measures of variability
- Polarization
- Uniformity
- Individuality
  - Range
  - Variance
  - Standard deviation
- Class interval
- Apparent limits
- Real limits
- Percentiles
- Percentile rank

LEARNING OBJECTIVES

At the conclusion of this chapter, you should be able to:

1. Define key terms.
2. Calculate measures of central tendency—mean, median, and mode—for grouped and ungrouped frequency distributions.
3. Calculate measures of variability—range, variance, and standard deviation—for grouped and ungrouped frequency distributions.
4. Use statistical software to calculate measures of central tendency and variation for grouped and ungrouped frequency distributions.
5. Explain why measures of central tendency and variability differ in grouped and ungrouped frequency distributions.

Two main types of measures are used to describe frequency distributions: measures of central tendency and measures of variability. **Measures of central tendency** measure location; measures of central tendency focus on the typical value of a data set. Variation emphasizes differences. **Measures of variability** measure distance and/or dispersion around the typical value of a data set. Generalizations about populations from samples are based on the variation of the variables in the data set.

Descriptive measures may be computed from both populations and samples. A **population** is a group of persons or objects about which a researcher or investigator wishes to draw conclusions. A **sample** is a subset of the population. Measures that result from the compilation of data from populations are called parameters. Measures that result from analysis of data from samples are called statistics. For example, if we were interested in determining the average age of all members of the American Health Information Management Association (AHIMA), all members of the association would make up the population; the average age would be a **population parameter**. But, if we were to draw a sample from the membership list and then calculate the average age, the average would be a **sample statistic**.

MEASURES OF CENTRAL TENDENCY

As previously stated, measures of central tendency summarize the typical value of a variable. When we think of measures of central tendency, we usually think of averages; however, averages or the arithmetic mean may not always be the most appropriate way of summarizing the most typical value of a data set. There are three major measures of central tendency; the selection of the relevant measure of central tendency is related to the scale of measurement used in data collection.

1. The **mode**, symbolized by M_O, is defined as the most frequently occurring observation in a frequency distribution; it is the only measure of central tendency that is appropriate for nominal data.
2. The **median**, symbolized by M_D, is the midpoint of a frequency distribution; it is appropriate for both ordinal and metric data.
3. The **mean**, symbolized by $\overline{X}$ (pronounced "x-bar"), is the arithmetic average; it is appropriate for metric data.

Mode

The mode is the simplest measure of central tendency. With the mode, we are indicating the most frequently occurring observation for metric data and/or the most frequently occurring category for nominal data. The mode is an important measure of central tendency for nominal-level data because an average cannot be taken from data that are placed in categories. Averages work only when a variable has a unit of measurement. The mode is an important measure of central tendency in the sense that it shows what the typical category on a variable is. For example, if we group the membership of AHIMA by sex, on average, the typical health information management (HIM) professional is female. We can say this because approximately 92% of the membership of AHIMA is female.

Another interpretation of the mode is that if the goal is to be as accurate as possible, the mode provides the best "guess" as to the category an observation may take on a variable. That is, no other guess for a random case will be correct as often as the mode. Continuing with our example, if one were to guess the sex of a typical HIM professional, the best "guess" would be that the sex of the practitioner was female. In making this guess, one would be correct approximately 92% of the time.

The mode offers several advantages: It is easy to obtain and to interpret, it is not sensitive to extreme values, and it is easy to communicate and explain to others. However, there are inherent disadvantages in using the mode. First, the mode may not be descriptive of the data because the most common category may not occur very often, especially when there are a large number of categories. Second, the mode may not be unique. That is, two categories or metric measures may be equally likely and more common than any other category or metric measure; when this occurs, we have a bimodal or multimodal distribution. If each category occurs with equal frequency, there is no mode. Third, the mode does not provide information about the entire frequency distribution—it only tells us the most frequently occurring value or category in the frequency distribution.

Also, the mode may be overly affected by sampling variation in cases where there is a bimodal distribution. If several samples are taken from the same population, the mode may fluctuate widely from sample to sample. For nominal-level data, the mode is also sensitive to how categories are combined. The classification scheme should be at the same level of generality for all categories rather than having broad categories for some and more specific categories for others. The mode can be manipulated by making the level of generality of different categories unequal. When reading a statistical analysis that reports the mode, we should always examine the categories to make sure that the modal categories were not manipulated by use of categories at different levels of generality. For example, compare the following hypothetical data on hospital admissions:

Hospital Admissions by Race—A		*Hospital Admissions by Race—B*	
White	60%	White	60%
African American	15%	African American	15%
Hispanic	12%	Other	25%
Asian	8%		
Other	5%		

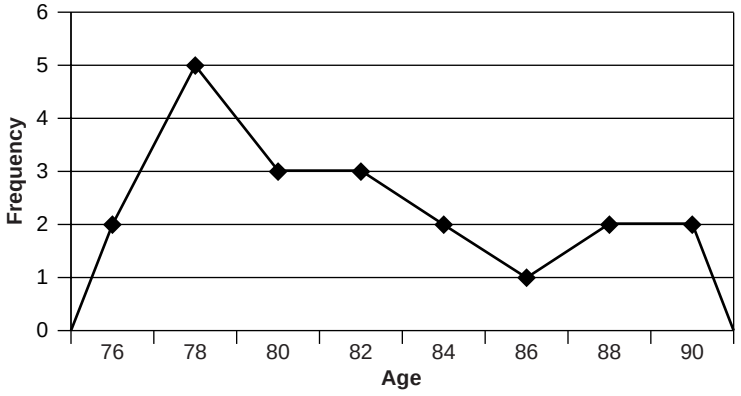
In example A, hospital admissions by race are distributed in a way that might be expected in an urban population. However, in example B, a biased view of admissions is presented by lumping two of the categories, Hispanic and Asian, into the "other" category. In this example, the individual presenting the data may be interested in emphasizing the low percentage in the African American category. Combining two racial categories into the other category creates a biased view of the racial makeup of hospital admissions.

For interval or ratio level data, the mode can be graphically represented in a frequency polygon. Figure 4–1 displays the frequency of the ages of a group of 20 nursing home residents, which are presented below:

76	76	78	78	78	78	78	80	80	80
82	82	82	84	84	86	88	88	90	90

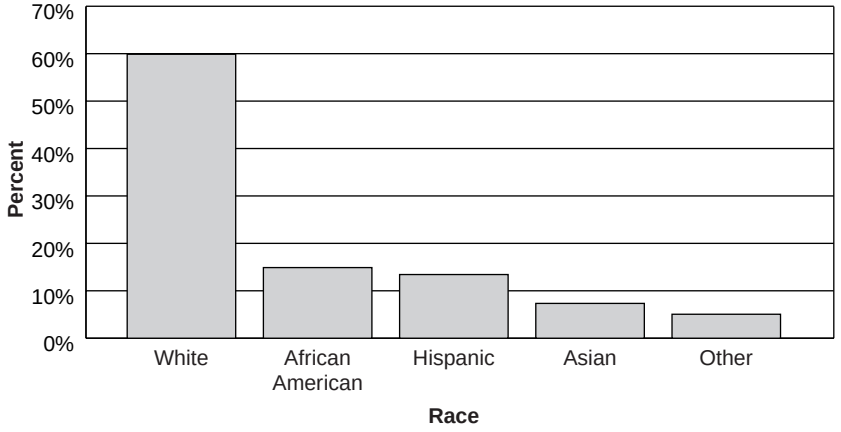
The modal age is 78 ($n = 5$).

Figure 4–1 Frequency Polygon for Ages of 20 Nursing Home Residents



For nominal data, the mode may be depicted in a bar chart (Figure 4–2).

Figure 4–2 Bar Graph of Hospital Admissions by Race



Median

When categories of a variable are ordered, the measure of central tendency should take order into account. The median does so by finding the value of the variable that corresponds to the middle case. It is a positional measure that indicates the point at which 50% of the cases fall above and 50% of the cases fall below.

If there is an odd number of observations in the frequency distribution, the median is the middle number. In the following frequency distribution, the median is 5—there are four cases above and four cases below the value 5.

1 2 3 4 **5** 6 7 8 9

If there is an even number of observations, the median is the midpoint between the two middle observations. It is found by averaging the two middle scores $[(x + y)/2]$. In the following frequency distribution, the median is 5.5 $[(5 + 6)/2]$.

1 2 3 4 **5 6** 7 8 9 10

If the two middle observations take on the same value, the mode is that value. When determining the mode, it does not matter if there are duplicate observations in the frequency distribution. Consider the following frequency distribution:

1 2 2 3 **4 4** 4 5 6 7

In this distribution there are 10 observations, so the median falls between cases between the fifth and sixth observation. Therefore, the median is 4 $[(4 + 4)/2 = 4]$.

The median is an important measure for data that fall on the ordinal scale of measurement. This is due to the limitations of other measures for ordered data. The mode can be used for ordinal data, but it does not take "order" into account, which is the characteristic that makes the measure more than just a nominal classification. The mode can also be unrepresentative for an ordinal measure. It is also not meaningful to take an average of ordered variables because the distance between the intervals is not necessarily equal. This concept is illustrated in Table 4–1, hospital ranking by severity of illness. Hospitals with adjacent ranks differ by as little as 0.001 (1.826 − 1.825) on their severity of illness scores, or by as much as 0.127 (1.753 − 1.626).

Averages are often calculated on ordered variables, but the results can be misleading. Therefore, when choosing a measure of central tendency, you should consider not only the scale of measurement upon which the variable falls but also the purpose of the measure. There are several advantages of using the median. First, it is relatively easy to obtain; second, it is based on the whole distribution rather than just a small portion of the distribution, as is the case with the mode. Third, the median is not influenced by extreme values or unusual outliers, so it is considered a resistant statistic. The median has another advantage in that it can be computed when a distribution is open-ended at the extremes. For example,

Table 4–1 Hospital Ranking by Severity of Illness

Hospital Ranking	*Severity of Illness*	*Distance Between Severity of Illness Scores*
1	2.152	—
2	2.027	0.125 (2.152 − 2.027)
3	1.965	0.062
4	1.876	0.089
5	1.826	0.050
6	1.825	0.001
7	1.753	0.072
8	1.626	0.127
9	1.594	0.032

consider the median length of stay for a group of five patients who were admitted to the hospital on the same day. If two patients are discharged on day 2 and one patient is discharged on day 4, the median length of stay is four days. The median can be determined without waiting to see how long the remaining two patients stay in the hospital. While we can determine the median in this example, we cannot determine either the mode or the mean until the two remaining patients are discharged.

Mean

The most effective way of summarizing the center of metric data is to average the values on the variable. The mode and median can be computed on metric data but they do not take full advantage of the numeric data inherent in the data in the frequency distribution. The formula for calculating the mean is

$$\overline{X} = \sum_{i=1}^{n} X_i / N$$

where Σ is summation, X_i is each successive observation in the frequency distribution (from the first observation, $i = 1$, to the last observation, n), and N is the total number of observations in the distribution.

To calculate the average daily census for the data in Table 4–2, we substitute into the preceding formula as follows:

$$\begin{aligned}\overline{X} &= \sum_{i=1}^{n} X_i / N \\ &= (167 + 185 + 173 + 182 + 179 + 173 + 170)/7 \\ &= 1{,}229/7 \\ &= 175.6\end{aligned}$$

Table 4–2 Daily Census, Critical Care Hospital, Week of July 2, 20xx

Day of Week	*Daily Census*
Sunday	167
Monday	185
Tuesday	173
Wednesday	182
Thursday	179
Friday	173
Saturday	170
Average Daily Census	175.6

The properties of the mean are presented in Exhibit 4–1.

Exhibit 4–1 Properties of Arithmetic Mean

1. The total sum of the deviations around the mean is zero.
2. The total sum of the negative deviations from the mean is always equal to the sum of the positive deviations from the mean. Therefore, the mean is the balance point for the distribution.
3. The sum of the squared deviations around the mean is smaller than the sum of the squared deviations around any other value.
4. The mean is more stable over repeated measures than any other measure of center.
5. Other important statistics, namely, standard deviation and standard error of the mean, are based on deviations from the mean.

There are two disadvantages associated with the mean. First, the mean can take on a fractional value even when the variable itself can take on only integer values. Consider the example in Table 4–2, which gives the daily inpatient census for one week. The average daily census results in a fractional number even though we do not have fractional patients. However, fractional values are considered more as a problem of interpretation than as a nonmeaningful result. For the data in Table 4–2, we could interpret the average daily census as, "On average, the number of inpatients per day was between 175 and 176."

A second disadvantage is that the mean is sensitive to extreme measures. That is, the mean is strongly influenced by outliers; therefore, it is considered a nonresistant measure. For example, if in Table 4–2, the daily census for Sunday was 225 instead of 167, the average daily census would increase to 183.9.

Weighted Mean

Often in the health care setting, we are involved in analyzing several data sets that contain the same information but for different time intervals. That is, we have several samples with separate means for each, and each sample may be of a different size. Review the data presented in Table 4–3.

Table 4–3 Calculation of Arithmetic and Weighted Means for Average Length of Stay (ALOS)

Month	*Discharges (n)*	*Discharge Days*	*ALOS*
Jan	947	4,228	4.46
Feb	763	3,965	5.20
Mar	574	1,842	3.21

Average of Means: (4.46 + 5.20 + 3.21)/3 = 4.29
Weighted Mean: [4.46(947)] + [5.20(763)] + [3.21 (574)]/2,284 = 4.39

What is the overall mean for the three months? One might be tempted to sum the means and divide by 3, which would result in an "average of the means." This would be inappropriate, because it would not take into account the difference in the sample sizes for each month. We need to calculate the **weighted mean**, which takes into account the difference in the size of each sample. We calculate the weighted mean by

$$\text{Weighted } \overline{X} = \sum N_i \overline{X}_i / N$$

where Σ is summation, N_i is the number of observations in each frequency distribution, N is the total number of observations in combined frequency distributions, and $\overline{X}_i$ is the mean of each distribution.

To calculate the weighted mean for the data in Table 4–3, we have

$$\begin{aligned}\text{Weighted } \overline{X} &= \sum N_i \overline{X}_i / N \\ &= [4.46(947)] + [5.20(763)] + [3.21(574)] \\ &= 4.39\end{aligned}$$

Compare the calculations in Table 4–3. The weighted mean takes into account the difference in the number of discharges for each month and is therefore more precise.

The mean can also be calculated for dichotomous data. We have already discussed the mode as the appropriate measure of central tendency for categorical data. Sometimes, however, when we have only two categories, it is more relevant to express the mean as the proportion of cases that fall into a certain category. When we are dealing with dichotomous data, we must first code the data; by convention, the variable of interest is coded as "1." For example, if the number of females in a class in mathematics is the variable of interest, "sex" is coded as follows: 1 = female; 0 = male. The proportion of cases with a score of 1 is denoted as *p*. *p* is a mean with an intuitive interpretation—the proportion of cases that fall in the category scored "1." Review the data presented in Table 4–4. In this context, the mean is related to the category of interest, not necessarily the most typical category. The interpretation is that the proportion of women in the math class is 0.15.

Table 4–4 Proportion of Females in Math Class

Sex	*Code*	*f*	*p*
Female	1	30	0.15(p)
Male	0	170	0.85 ($1 - p$)
Total		200	1.00

Mean = $p = f_1/n = 30/200 = 0.15$

As we have already discussed, one of the disadvantages of using the mean is that it is sensitive to extreme measures. To eliminate the effects of extreme measures, the outliers may be "trimmed" from the frequency distribution before the mean is calculated. An example of **trimmed mean** occurs in some competitive sports, where the top and bottom scores are discarded before the mean score is computed.

Another method for improving the calculated mean's resistance to extreme measures is to "winsorize" the mean. In winsorizing, the most extreme values are changed to equal the next less extreme values rather than being dropped totally from the data set, as in trimming. For example, the 5% trimmed mean drops the highest 5% of the observations and lowest 5% of the observations before the mean is computed. The 5% **winsorized mean** with 20 observations changes the highest value (highest 5%) to the second highest value, and changes the lowest value (lowest 5%) to the second lowest value. Exhibit 4–2 compares these methods.

In Exhibit 4–2, we can see that the arithmetic mean is 1,257.7, which is vastly different from the trimmed mean, 1,217.5, and the winsorized mean, 1,216.6. The latter two adjusted means are actually similar to the median, 1,220. Which measure of central tendency best represents the distribution? This is where judgment is important. The statistical analyst

Exhibit 4–2 Calculation for Arithmetic Mean, Trimmed Mean, and Winsorized Mean

Data Set:

660	1,070	1,220	1,430
740	1,100	1,250	1,475
800	1,100	1,250	1,550
820	1,140	1,250	1,600
880	1,150	1,300	1,700
930	1,150	1,300	1,850
1,000	1,200	1,400	2,900

Before calculations are made, arrange the data in order from lowest to highest.

Arithmetic Mean: Sum the observations and divide by the total number of observations:
$660 + 740 + \ldots + 2{,}900 = 35{,}215/28 = 1{,}257.7$

Trimmed Mean: Eliminate the lowest number, which is 660.
Eliminate the highest number, which is 2,900.

Subtract these values from the previous sum: $35{,}215 - (660 + 2{,}900) = 31{,}655$

Divide by the remaining number of observations: $31{,}655/26 = 1{,}217.5$

Winsorized Mean: Identify the lowest 5%, 660 and 740, which are each replaced by 800.
Identify the highest 5%, 1,850 and 2,900, which are replaced by 1,700.

$32{,}215 - 600 - 740 - 1{,}850 + 2{,}900 = 29{,}065$
$29{,}065 + 800 + 800 + 1{,}700 + 1{,}700 = 34{,}065$

Divide by the total number of observations:
$34{,}065/28 = 1{,}216.6$

Median: 1,220

Summary:

Arithmetic Mean:	1,257.7
Trimmed Mean:	1,217.5
Winsorized Mean:	1,216.6

must “eyeball” the raw data to make this decision. It appears that the highest score, 2,900, is strongly influencing the arithmetic mean. Therefore, the trimmed mean, the winsorized mean, and/or the median better represent this data set. The data analyst should select the measure of central tendency that best describes the typical value in the frequency distribution. The analyst should include an explanation of why an alternative to the more traditional measure of central tendency, the mean, was used to describe the frequency distribution.

We can use SPSS or other statistical software to calculate the mean, median, and mode. The SPSS output is displayed in Exhibit 4–3. From the “Analyze” menu, choose “Descriptive statistics,” then “Frequencies.” In the Frequencies dialog box, select “Statistics,” and click “Mean, Median, and Mode.” SPSS summarizes our data set in a frequency table.

Exhibit 4–3 SPSS Output for Measures of Central Tendency

Statistics

Data Set

N	
Valid	*28*
Missing	*0*
Mean	*1257.6786*
Median	*1210.0000*
Mode	*1250.00*

To obtain the mean, median, and mode using SPSS:

- From the menus, choose:
  Analyze
  →Descriptive statistics
  →Frequencies
- In the Frequencies dialog box, click
  Statistics
  →Select Mean, Median, Mode

MEASURES OF VARIABILITY

Measures of central tendency are not the only statistics used to summarize a frequency distribution. We also want to consider the spread of the distribution, which tells us how widely the observations are spread out around the measure of central tendency. The most commonly used measures of spread are the variance and the standard deviation. The scales of measurement appropriate for the use of the variance and standard deviation are the interval and ratio scales.

Measures of spread increase in value with greater variation on the variable. Measures of spread equal zero when there is no variation. Maximum spread for metric and ordinal variables occurs when cases are evenly split between two extreme groups. This is called **polarization**. Maximum dispersion for nominal variables is defined as when there is an even distribution of cases across the categories regardless of the number of categories; this is called **uniformity**. When each category of a nominal variable occurs just once, it is called **individuality**.

Range

The simplest measure of spread is the **range**. It is simply the difference between the smallest and largest values in a frequency distribution:

$$\text{Range} = X_{\max} - X_{\min}$$

The range is easy to calculate but is affected by extreme measures. Therefore, it is a nonresistant measure of spread. The range varies widely from sample to sample. Only the two most extreme scores affect its value, so it is not sensitive to other values in the distribution. Also, the range is dependent upon sample size: in general, the larger the sample size, the greater the range.

Two frequency distributions may have the same range, but the observations may differ greatly in variability. For example, consider the following two frequency distributions:

Distribution 1

1	2	3	4	5	6	7	8	9	10

Distribution 2

1	1.5	3	3.5	3.7	7	8	8.26	10	10

The range for both distributions is 9 ($10 - 1 = 9$). But if we compare the two distributions, we see that there is more variation in distribution 2 than in distribution 1. This is confirmed when the variance for each distribution is calculated—the variances for distributions 1 and 2 are 3.03 and 3.44, respectively.

Variance and Standard Deviation

The **variance** (s^2) is the average of the squared deviations from the mean. The variance of a frequency distribution will be larger when the observations within the distribution are widely spread. The variance (and, as we shall see, the **standard deviation**) is maximized when the data are polarized. The formula for calculating the variance is:

$$s^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2 / N - 1$$

The squared deviations of the mean are calculated by subtracting the mean of a frequency distribution from each value in the distribution, $X - \overline{X}$. The difference between the two values is then squared, $(X - \overline{X})^2$. The squared differences are summed and divided by $N - 1$.

The term $N - 1$ is a concept referred to as the number of degrees of freedom. If the mean of a frequency distribution is known, then only $N - 1$ observations are free to vary. Stated another way, if we know the mean, and the $N - 1$ scores, we can determine the nth score. The effect of dividing by $N - 1$ increases the value of s^2 slightly, and is considered to be a less biased estimate of the population variance. However, when N is large, the effect of using $N - 1$ instead of N is negligible. We will encounter this concept again in later chapters.

As an example, we will calculate the variance for the census data that appear in Table 4–2.The average daily census is 175.6. To calculate the variance, we set up the problem as follows:

Day	*Census*	*Census − Mean* $(X - \bar{X})$	*(Census − Mean)* $(X - \bar{X})$
Sunday	167	−8.6	73.96
Monday	185	9.4	88.36
Tuesday	173	−2.6	6.76
Wednesday	182	6.4	40.96
Thursday	179	3.4	11.56
Friday	173	−2.6	6.76
Saturday	170	−5.6	31.36
Total	1,229	≈0*	259.72

*Approximate due to rounding.

Note that in the property of the mean that is described in Exhibit 4–1, the sum of the deviations from the mean is equal to 0 and is displayed in the preceding calculations. (In this example, the sum is approximately equal to 0.0 because the mean is rounded.) Substituting into the formula, we calculate the variance as:

$$\begin{aligned} s^2 &= \sum_{i=1}^{n} (X_i - \bar{X})^2/N - 1 \\ &= 259.72/6 \\ &= 43.28 \end{aligned}$$

The variance is equal to 43.28, but what does this mean? The interpretation of the variance is not easy at the descriptive level because the original units of measure are squared to arrive at the variance. However, if we take the square root of the variance, we return to the original units of measurement. The square root of variance is the standard deviation (s):

$$s = \sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2/N - 1}$$

Continuing with the census example, the standard deviation is calculated as

$$\begin{aligned} s &= \sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2/N - 1} \\ &= \sqrt{43.28} \\ &= 6.58 \end{aligned}$$

The standard deviation is the most widely used measure of variation that is used in descriptive statistics. The standard deviation measures variability in the same units of measurement as the sample (i.e., height, age, length of stay, and so on). Since the standard deviation is easier to interpret, it is the preferred measure of dispersion for a frequency distribution. The standard deviation is interpreted in relation to the normal distribution, which we will discuss in Chapter 5.

As we already noted, the variance is of little use in descriptive statistics, but it is important in procedures related to statistical inference. The standard deviation, which is the square root of the variance, is defined in terms of deviations from the mean. It is an important measure at the descriptive level. Procedures for calculating the mean, variance, and standard deviation of a frequency distribution are outlined in Exhibit 4–4.

Exhibit 4–4 Calculation of Mean, Variance, and Standard Deviation for Ages of Nursing Home Residents

Patient No.	*Age*	$(X - \bar{X})$	$(X - \bar{X})^2$	*Patient No.*	*Age*	$(X - \bar{X})$	$(X - \bar{X})^2$
1	76	−6	36	11	82	0	0
2	76	−6	36	12	82	0	0
3	78	−4	16	13	82	0	0
4	78	−4	16	14	84	+2	4
5	78	−4	16	15	84	+2	4
6	78	−4	16	16	86	+4	16
7	80	−2	4	17	88	+6	36
8	80	−2	4	18	88	+6	36
9	80	−2	4	19	90	+8	64
10	80	−2	4	20	90	+8	64
				Σ	1,640	0	376

Using a random sample of 20 nursing home residents, calculate the average age of nursing home residents.
Mean:

1. Sum the observations (Σ): 1,640
2. Divide by the number of observations (N): 20
   1,640/20 = 82

Interpretation: The average age or typical age of any nursing home resident is 82.

Variance:

1. Subtract the mean from each observation: $(X - \bar{X})$
2. Square each deviation from the mean: $(X - \bar{X})^2$
3. Sum the squared deviations from the mean: $\sum (X - \bar{X})^2$
4. Divide the sum of the squared deviations from the mean by $N - 1$ (19).

$$s^2 = \sum (X - \bar{X})^2/N - 1$$
$$= 376/19$$
$$= 19.8$$

The standard deviation is the square root of the variance:

$$s = \sqrt{\sum (X - \bar{X})^2/N - 1}$$
$$= 4.45$$

Interpretation: Approximately 68% of the nursing home residents are between the ages of 77.55 ($\bar{X} + 1s$).

CALCULATING MEASURES OF CENTRAL TENDENCY AND VARIABILITY USING SPSS

Thus far, we have reviewed the formulas for calculating the mean, range, variance, and standard deviation. Exhibit 4–5 displays the SPSS output for the measures of central tendency and variation for the nursing home data presented in Exhibit 4–4. To calculate these measures using SPSS, select "Descriptive Statistics" from the Analyze menu, then select "Frequencies." In the Frequencies dialog box, select "Options." You can now "click" the measures of central tendency and variation that you are interested in. A frequency table is prepared as part of the output. This allows you to verify the actual observations that were entered on the data sheet. Each age, as well as the number of times it was entered on the data sheet, appears in columns 2 and 3, respectively.

Exhibit 4–5 SPSS Output for Nursing Home Data

Statistics

Age

N	
Valid	20
Missing	0
Mean	82.0000
Median	81.0000
Mode	78.00*
Std. Deviation	4.44854
Variance	19.7895
Range	14.00

*Mulitple modes exist. The smallest value is shown.

Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	76.00	2	10.0	10.0	10.0
	78.00	4	20.0	20.0	30.0
	80.00	4	20.0	20.0	50.0
	82.00	3	15.0	15.0	65.0
	84.00	2	10.0	10.0	75.0
	86.00	1	5.0	5.0	80.0
	88.00	2	10.0	10.0	90.0
	90.00	2	10.0	10.0	100.0
	Total	20	100.0	100.0	

Note that for the statistics table, there is a note indicating that the distribution contains multiple modes with the smallest value displayed in the report. Thus, the mode may not be a fair representation of the distribution. The frequency table tells us that there are two modes, 78 and 80.

In reviewing the data, you can see that the results are similar to the results that were calculated with the assistance of a hand-held calculator. Note that the results indicate the number of observations that were included in the calculation. This should always be reviewed to verify that the correct number of observations was entered on the SPSS data sheet. Computer packages also provide more precision in our results. SPSS carried out the calculations to more than four decimal places.

DICHOTOMOUS DATA

Just as we could compute the mean for dichotomous data, we can also determine the variance and standard deviation for dichotomous data:

$$\text{Variance: } s^2 = p(1 - p)$$
$$\text{Standard deviation: } s = \sqrt{p(1 - p)}$$

where p is equal to the variable of interest. In our coding scheme, for the variable "sex," the category of interest, "female," is coded 1, and the "male" category is coded 0. Calculations of the variance and standard deviation for the variable "sex" are presented in Exhibit 4–6.

Exhibit 4–6 Calculation of Variance and Standard Deviation for Dichotomous Data

Sex	*Code*	*f*	*p*
Female	1	30	0.15 (p)
Male	0	170	0.85 ($1 - p$)
Total		200	1.00

Mean: $p = f_1/n = 30/200 = 0.15$
Variance: $s^2 = p(1 - p) = 0.15(1 - 0.15) = 0.1275$
Standard deviation: $s = \sqrt{p(1 - p)} = 0.357$

GROUPED FREQUENCY DISTRIBUTIONS

Since it is inconvenient to work with large data sets, sometimes data are grouped into class intervals for analysis. Table 4–5 presents an example of an ungrouped frequency distribution. There are 58 patients with varying lengths of stay (LOSs). In this format, the data are

difficult to interpret, so we will group the LOSs into **class intervals**. Intervals are ways of classifying and/or summarizing raw data into categories.

Table 4–5 Patient Length of Stay (LOS)

	Frequency	
LOS (X)	*f*	*fX*
1	0	0
2	2	4
3	6	18
4	6	24
5	6	30
6	11	66
7	6	42
8	8	64
9	5	45
10	3	30
11	1	11
12	2	24
13	1	13
14	1	14
Total	58	385

Grouping Data

The first step in grouping data is to determine the number of class intervals (categories) to use. There is no fixed rule as to how many intervals are appropriate; some recommend 5 to 15 class intervals, while others recommend 10 to 20. All agree, however, that there should not be more than 20. As an example, we will group the LOS frequency distribution that appears in Table 4–5. To construct the grouped frequency distribution, we will arbitrarily select five as the number of class intervals for grouping the LOS data. Next, we need to determine the width of each of these intervals. This is accomplished by dividing the range of the distribution by the number of class intervals: $(14 - 2)/5 = 2.4$. The nearest odd integer value is used to select the width of the class interval; thus, the width will be 3. To construct the class interval, we need to determine what the highest and lowest class intervals should be. These class intervals will contain the highest and lowest values in the frequency distribution. In the LOS data in Table 4–5, the number 14 is the highest observation and the number 2 is the lowest observation; thus, the highest and lowest class intervals must contain these values. The class interval containing the highest value should be placed at the top of the grouped frequency distribution, and the lowest class interval should be placed at the bottom of the grouped frequency distribution. Intervals should be continuous throughout distribution: that is, there should not be any gaps in values in the distribution. Using the ungrouped data, tally the frequencies that occur in each interval. The LOS data in Table 4–5 are presented as a grouped frequency distribution in Table 4–6.

Table 4–6 Grouped Frequency Distribution, LOS Data

Class Interval Apparent Limits	*Class Interval Real Limits*	*Tally*	*f*	*M*
13–15	12.5–15.5	//	2	14
10–12	9.5–12.5	///// /	6	11
7–9	6.5–9.5	///// ///// ///// ////	19	8
4–6	3.5–6.5	///// ///// ///// ///// ///	23	5
1–3	0.5–3.5	///// ///	8	2
			58	

Table 4–6 groups the LOS data into five class intervals, each having a width of 3. Also, note that there are two columns—one indicating the **apparent limits** of the class interval and the other indicating the **real limits** of the class interval. The real limits depict the continuous nature of the distribution. The interval width is the difference between the upper and lower real limits of the class interval—for example, $3.5 - 0.5 = 3$. The highest value, 14, is contained in the class interval "13–15," and the lowest value, 2, is contained in the class interval "1–3." Keep in mind that the distribution should not contain any gaps; even if there are no values that fall within a given interval, it should be included in the distribution. The midpoint (M) of the intervals is obtained by adding the apparent limits of the class intervals and dividing by 2—for example, $(1 + 3)/2 = 2$; $(4 + 6)/2 = 5$. The midpoint will be used in calculating the mean, median, and mode of the grouped distribution. As an alternative to the preceeding procedure for constructing class intervals, we can use the "Sturges" rule as a guide:

$$k = 1 + 3.3\log_n(n)$$

where $\log_{10}$ is the usual base 10 logarithm, k is the number of class intervals, and n is the number of data items to be grouped into class intervals.

Using the LOS data from Table 4–5, the number of class intervals (k) required for the grouped frequency distribution would be calculated as follows:

$$\begin{aligned} k &= 1 + 3.3\log_n(n) \\ \log_{10}(58) &= 1.7653 \\ k &= 1 + 3.3(1.7653) \\ &= 6.82 \end{aligned}$$

Since our result is a fraction, we will round to the nearest whole number. Thus, seven class intervals are needed for grouping the frequency distribution in Table 4–5.

After the number of class intervals has been determined, the next step is to determine the width of the intervals:

$$\text{width} = \text{max} - \text{min (range)}/k$$

where max is the largest value in the data set, min is the smallest value in the data set, and k is the number of class intervals.

$$\text{width} = (14 - 2)/7 = 12/7 = 1.7 = 2$$

The results appear in Table 4–7. The data in Table 4–7 are usually summarized as presented in Table 4–8. The proportion of the total in each class interval is called the relative frequency (rel f) and is obtained by dividing the frequency (f) or number of observations in a given interval by the total number of observations (N). The cumulative frequency (cum f) is obtained by summing the observations in each interval with the number of observations (frequencies) in the next interval, beginning with the lowest class interval and proceeding upward. The first class interval (1–2) contains two observations; thus, the cumulative frequency is 2. The next class interval contains 12 observations; thus, the cumulative frequency is 14 (2 + 12). Proceed in this manner until the last class interval is reached. Exhibit 4–7 outlines the steps for constructing a grouped frequency distribution.

Table 4–7 Class Intervals for LOS Data

Apparent Limits	*Real Limits*	*Tally*	*f*	*M*
13–14	12.5–14.5	//	2	13.5
11–12	10.5–12.5	///	3	11.5
9–10	8.5–10.5	~~////~~ ///	8	9.5
7–8	6.5–8.5	~~////~~ ~~////~~ ~~////~~	14	7.5
5–6	4.5–6.5	~~////~~ ~~////~~ ~~////~~ //	17	5.5
3–4	2.5–4.5	~~////~~ ~~////~~ //	12	3.5
1–2	0.5–2.5	//	2	1.5
Total			58	

Table 4–8 Grouped Frequency Distribution of LOS Data

Class Interval	*f*	*Cum f*	*Rel f*	
13–14	2	58 (56 + 2)	0.034 (2/58)	100.0 (96.5 + 3.4)
11–12	3	56 (53 + 3)	0.052	96.5 (91.3 + 5.2)
9–10	8	53 (45 + 8)	0.138	91.3 (77.5 + 13.8)
7–8	14	45 (31 + 14)	0.241	77.5 (53.4 + 24.1)
5–6	17	31 (14 + 17)	0.293	53.4 (24.1 + 29.3)
3–4	12	14 (2 + 12)	0.207	24.1 (3.4 + 20.7)
1–2	2	2	0.034	3.4
Total	58			

Exhibit 4–7 Steps in Constructing a Grouped Frequency Distribution

1. Determine the number of class intervals needed by the Sturges rule. It is suggested that the number of intervals be limited to 20.
2. Determine the width of the class intervals by dividing the range of the frequency distribution by the number of class intervals obtained in step 1. Intervals should be set up so that one score cannot belong to more than one class interval.
3. Determine the point at which the lowest class interval should begin.
4. Record the limits of all class intervals, placing the interval containing the highest score value at the top. Intervals should be continuous and of the same width. Do not leave out class intervals in which no observations occur; to do so would create a misleading impression.
5. Using the tally system, place a tally for each observation in the corresponding class intervals.
6. Summarize the tallies for each class interval in a frequency (f) column.
7. Record the total (f) at the bottom of the frequency column.

There are several disadvantages associated with grouped frequency distributions. First, precision in the resulting statistical calculations is lost. When values are grouped into an interval, the values may be spread evenly throughout the interval, or the values may be concentrated at either end of the interval. When data are grouped, the assumption is that the observations are spread evenly across the class intervals. Second, as we have seen, different groupings can result from the same frequency distribution. Grouping a set of values does not result in a unique set of grouped scores. That is, if the same frequency distribution were grouped into slightly different class intervals, the sample statistics would not be exactly the same.

Calculating the Mean from a Grouped Frequency Distribution

The first step in calculating the mean from a grouped frequency distribution is to determine the midpoint of each interval. Table 4–9 presents the grouped LOS with the corresponding midpoint (X) for each interval. The midpoint represents all observations that fall

Table 4–9 Grouped LOS Data for Calculating the Mean

Class Interval	*Midpoint (X)*	*f*	*f(X)*
13–14	13.5	2	27
11–12	11.5	3	34.5
9–10	9.5	8	76
7–8	7.5	14	105
5–6	5.5	17	93.5
3–4	3.5	12	42
1–2	41.5	2	3
		$\Sigma = 58$	$\Sigma = 381$

$$\bar{X} = \sum fX/N$$
$$= 381/58$$
$$= 6.57$$

into that interval. For example, the observations 13 and 14 (from Table 4–5) fall into the class interval "13–14"; the midpoint, 13.5, represents both of these observations. The mean is obtained by multiplying the midpoint (X) of each class interval by its corresponding frequency (f); these observations are summed and then divided by the total number of observations (N):

$$\bar{X} = \sum fX / N$$

where X is the midpoint of each interval.

Therefore, for the grouped LOS data, the mean is

$$\bar{X} = 381/58$$
$$= 6.57$$

The mean for the grouped data, 6.57 (Table 4–9), does not equal the mean for the ungrouped data, 6.64 (385/58). This discrepancy illustrates the loss of precision that occurs when grouping frequency distributions.

Calculating the Median from Grouped Data

Recall that the median is the point in a distribution that 50% of the observations fall above and 50% of the observations fall below. To determine the median in a grouped distribution, we first need to determine which observation meets this definition. For the LOS data in Table 4–9, the median would be the 29th value, as calculated below:

$$50\% \text{ of } N = 0.5 \times 58 = 29$$

Thus, the 29th value will be the median for the grouped distribution. Inspection of the data in Table 4–8 shows that the 29th observation falls in the interval "5–6." How is this determined? Begin counting the frequencies (f) in the lowest class interval, 1 to 2, until the interval containing the 29th observation is located ($2 + 12 + 17 = 31$). The interval 5 to 6 contains the 15th, 16th, 17th, 18th, 19th, and so on, up to the 31st observation. Thus the 29th observation occurs in this interval. The cumulative frequency column provides this information for us. The median is obtained by

$$\text{Mdn} = L[w(1/2n - \text{c})]f_{\text{mdn}}$$

where L is the real lower limit of the class interval containing the median, W is the width of the interval, c is the total number of values falling below the interval containing the median (cumulative frequency), and f_{mdn} is the frequency of the values in the interval containing the median value.

Substituting into our formula, since $L = 4.5$, $w = 2$, $c = 14$, and $f_{\text{mdn}} = 17$, we have

$$\begin{aligned} & 4.5 + \{2[1/2(58) - 14]\}/17 \\ &= 4.5 + [2(15)]1/17 \\ &= 4.5 + (30/17) \\ &= 4.5 + 1.76 \\ &= 6.26 \end{aligned}$$

The median for the grouped frequency distribution is 6.26 days; the median for the ungrouped distribution is 6 days.

Calculating the Mode from Grouped Data

There are several ways to determine the mode from grouped data. The simplest is to take the midpoint of the most frequently occurring class interval. For the LOS data in Table 4–9, the mode is 5.5 because 17 observations fall in this interval. This is referred to as the crude mode.

The second way to determine the mode from a grouped distribution adjusts the modal value in relation to the relative frequencies in the class intervals adjacent to the class interval containing the modal value. It pulls the modal value toward the adjacent class interval that has greater frequency. This mode is referred to as the refined mode, and is calculated as follows:

$$\text{Refined mode} = w(f_{\text{mo}} - f_{\text{b}})/(f_{\text{mo}} - f_{\text{b}}) + (f_{\text{mo}} - f_{\text{a}})$$

where L is the real lower limit of the class interval containing the modal value, w is the width of the class interval, f_{mo} is the number of values (f) in the class interval containing the mode, f_{b} is the number of values (f) in the adjacent class interval below the class in-

terval containing the mode, and f_a is the number of values (f) in the adjacent class interval above the class interval containing the mode.

For the grouped LOS data in Table 4–10, $L = 4.5$, $w = 2$, $f_{mo} = 17$, $f_b = 12$, and $f_a = 14$, so the refined mode is calculated as follows:

$$\begin{aligned}\text{Refined mode} &= w(f_{mo} - f_b/(f_{mo} - f_b) + (f_{mo} - f_a) \\ &= 4.5 + [2(17 - 12)]/(17 - 12) + (17 - 14) \\ &= 4.5 + 2[5/(5 + 3)] \\ &= 4.5 + (10/8) \\ &= 4.5 + 1.25 \\ &= 5.75\end{aligned}$$

Table 4–10 Grouped LOS Data for Calculating the Mean

Class Interval	*Midpoint (X)*	*f*	*fX*	*F(X²)*
13–14	13.5	2	27	364.5
11–12	11.5	3	34.5	396.75
9–10	9.5	8	76	722.0
7–8	7.5	14	105	787.5
5–6	5.5	17	93.5	514.25
3–4	3.5	12	42	147.0
1–2	1.5	2	3	4.5
		$\Sigma = 58$	$\Sigma = 381$	$\Sigma = 2{,}936.5$

Thus, the refined mode is 5.75, and the crude mode is 5.5. You can see from the calculations for the refined mode that the number of observations in the class interval that is immediately above ($f_a = 14$), the interval containing the mode is pulling the modal value in that direction. Compare the refined mode to the mode in the ungrouped frequency distribution for the LOS data in Table 4–5, where the mode is 6. The length of stay—six days—occurred 11 times.

Calculating the Variance and Standard Deviation from Grouped Data

To compute the variance and standard deviation from grouped data, we will use what is called the raw score formula. The data from Table 4–9 are reproduced in Table 4–10, but we need to add another column—$f(X^2)$—which is the midpoint (X) squared multiplied by the frequency (f).

The raw score formula for calculating the variance is

$$s^2 = \sum x^2/(N - 1)$$

and the standard deviation is

$$s = \sqrt{\sum x^2/(N - 1)}$$

where the Σx^2 is

$$\sum x^2 = \sum f(X^2) - [\sum (X)^2/N]$$

To calculate the variance and standard deviation, we have

$$\begin{aligned}\sum x^2 &= \sum f(X^2) - [\sum (X)^2/N] \\ &= 2{,}936.5 - (381^2/58) \\ &= 2{,}936.5 - 2{,}502.78 \\ &= 433.7\end{aligned}$$

$$\begin{aligned}s^2 &= \sum x^2/(N - 1) \\ &= 433.7/57 \\ &= 7.61\end{aligned}$$

and

$$\begin{aligned}s &= \sqrt{\sum x^2/(N - 1)} \\ &= \sqrt{7.61} \\ &= 2.76\end{aligned}$$

The variance and standard deviation for our LOS data are 7.61 and 2.76, respectively. Measures of central tendency and variation for the ungrouped and grouped frequency distributions are compared in Table 4–11. Note that the grouped distribution results in greater variation than the ungrouped distribution because only the midpoint of each interval is considered in the calculations, rather than the entire data set.

Table 4–11 Measures of Central Tendency and Variation, Ungrouped Data versus Grouped Data

	Ungrouped Distribution	*Grouped Distribution*
Mean	6.64	6.57
Median	6	6.26
Mode	6	5.75
Variance	2.76	7.61
SD	1.66	2.76

Percentiles

Health information management professionals are often asked to "benchmark" characteristics of the organization within which we work to the performance of peer organizations. An example is the semiannual PEPP (Exhibit 4–8) report from the state Peer Review Organization (PRO), now called Quality Improvement Organizations (QIOs). This report compares the performance of one hospital against all others in the state on certain DRG pairs. National DRG data indicated that certain DRGs are subject to "upcoding"—that is, assigning codes that would move an inpatient discharge from a lower weighted DRG to a higher weighted DRG. Peer data are used to identify "outliers" that might indicate a health care organization is "undercoding" or "overcoding" certain types of Medicare cases. The data are reported in the form of percentiles.

Exhibit 4–8 Example PEPP Paired DRG Report

	Hospital			*State*		
Indicator	*1st DRG*	*2nd DRG*	*% Higher*	*10th %tile*	*90th %tile*	*Median (50th %tile)*
014*/015	67	21	76%	47%	78%	63%
079*/089	27	86	24%	8%	30%	19%
088*/096	61	10	86%	80%	96%	87%
089*/096	86	10	90%	81%	97%	88%
121*/124	37	95	28%	34%	100%	72%

*Denotes higher weighted DRG

Percentiles are measures of spread and location. They represent the proportion of scores in a distribution that a specific score is greater than or equal to, or less than or equal to. The *n*th percentile of a population, $n(p)\%$, is the number that $n(p)$ of the population fall below, and $n(1 - p)\%$ of the population fall above. Consider the following examples:

95th percentile	$95\% \leq n\ (0.95)$	$5\% \geq n(0.95)$
90th percentile	$90\% \leq n\ (0.90)$	$10\% \geq n(0.90)$
75th percentile	$75\% \leq n\ (0.75)$	$25\% \geq n(0.75)$
50th percentile	$50\% \leq n\ (0.50)$	$50\% \geq n(0.50)$
25th percentile	$25\% \leq n\ (0.25)$	$75\% \geq n\ (0.25)$
10th percentile	$10\% \leq n\ (0.10)$	$90\% \geq n(0.10)$

The data in Exhibit 4–8 compares hospital-specific data for selected DRGs with state data. The report indicates that for the DRG pair 014, Intracranial Hemorrhage and Stroke with Infarction, and 015, Non-Specific Cardiovascular and Precerebral Occlusion without Infarction, 76% of the cases fell into the higher weighted DRG ([67 + 21]/88 = 76%). This percentage falls between the 10th and 90th percentiles when compared to all hospitals within the state. A percentage greater than 78% might indicate that the hospital was upcoding cases from DRG 015 to DRG 014. A percentage below 47% might indicate that the hospital has opportunities for improving the quality of their coding.

The general interpretation for the PEPP report for this state is that for 90% of the hospitals, up to 78% of their cases fall into the higher weighted DRG, and 10% of the hospitals have up to 47% of their cases fall into the higher weighted DRG. Fifty percent of the hospitals have up to 63% of their cases fall into the higher weighted DRGs. For this hospital, the ranking for the DRG pair is between the 50th and 90th percentile. This conclusion is that it does not appear that the organization is upcoding cases for this particular DRG pair.

Calculating Percentile Ranks

To determine a **percentile rank**, we prepare a frequency distribution table of the length of stay for patients discharged from DRG 127, Heart Failure and Shock. The length of stay ranges from 1 day to 27, and are ranked from the longest to the shortest length of stay. We want to know the percentile rank for a length of stay of five days.

LOS	*f*	*cf*	*Cum %*
27	1	60	100.0%
17	1	59	98.3%
14	1	58	96.7%
11	3	57	95.0%
10	3	54	90.0%
9	1	51	85.0%
8	4	50	83.3%
7	6	46	76.7%
6	3	40	66.7%
5	8	37	61.7%
4	4	29	48.3%
3	10	25	41.7%
2	12	15	25.0%
1	3	3	5.0%

The formula for obtaining the percentile rank is:

$$\text{percentile rank} = (\text{cum } f/N) \times 100$$

We can then determine the percentile rank from the cumulative frequency column. Thus:

$$\begin{aligned}\text{percentile rank} &= (\text{cum } f/N) \times 100\\ &= (37/60) \times 100\\ &= 0.617 \times 100\\ &= 61.7\end{aligned}$$

The interpretation is that 61.7% of the discharges from DRG 127 have a length of stay of five or fewer days.

We can also divide a frequency distribution into quartiles. When a distribution is divided into quartiles we have:

Q_1 (lower quartile)—25th percentile
Q_2 (median)—50th percentile
Q_3 (upper quartile)—75th percentile

If we "eyeball" the data in the table, the LOS that falls at the 25th percentile is two; the LOS that falls at the 50th percentile is between four and five days; and the LOS that falls at the 75th percentile is between six and seven days.

To calculate the LOS of stay for a given percentile, we must first determine which observation, when placed in rank order, corresponds to that percentile. This is the **cumulative frequency**. If we are interested in the length of stay that falls at the 25th percentile, the cumulative frequency (cum f) is calculated as:

$$\text{Cum } f \text{ at nth percentile} = (\text{percentile rank} \times N)/100$$
$$\text{Cum } f \text{ at 25th percentile} = (25 \times 60)/100$$
$$\text{Cum } f = 15$$

Referring to our table, the cumulative frequency of 15 has a corresponding length of stay of two days. Thus, the 25th percentile is 2. Twenty-five percent of the discharges from DRG 127 have a length of stay of two or fewer days.

Calculating the LOS that corresponds to the 50th and 75th percentiles is not as straightforward because they fall between four and five, and six and seven, respectively. Recalling the rules for grouping data into class intervals, we can use a general method for determining the score is:

$$\text{Score at a given percentile} = X_{ll} + \frac{i(\text{cum}f - \text{cum}f_{ll})}{f_i}$$

where

X_{ll} = score at real lower limit of the interval containing cum f
i = width of interval
cum f = cumulative frequency of the score
cum f_{ll} = cumulative frequency at the real lower limit of the interval containing cum f
f_i = number of scores within the interval containing cum f

To calculate the 50th and 75th percentiles (or quartiles) we must first determine the interval that contains the 50th and 75th percentiles:

$$\begin{aligned}\text{cum } f_{50} &= (\text{percentile rank} \times N)/100\\ &= (50 \times 60)/100\\ &= 30\\ \text{cum } f_{75} &= (75 \times 60)/100\\ &= 45\end{aligned}$$

Thus, we are looking for the LOSs that correspond to the 30th and 45th observations. The 30th observation falls in the interval that corresponds to the LOS of five days, and the 45th observation fall in the interval that corresponds to seven days. Thus,

$$\begin{aligned}&= X_{ll} + \frac{i(\text{cum}f - \text{cum}f_{ll})}{f_i}\\ \text{Q}_{50} &= 4.5 + \frac{1(30 - 29)}{8}\\ &= 4.625\end{aligned}$$

and

$$\begin{aligned}&= X_{ll} + \frac{i(\text{cum}f - \text{cum}f_{ll})}{f_i}\\ \text{Q}_{75} &= 6.5 + \frac{1(45 - 40)}{6}\\ &= 7.3\end{aligned}$$

Thus, the length of stay that falls at the 50th percentile is 4.6, and the length of stay that falls at the 75th percentile is 7.3 days.

CONCLUSION

The mean, median, and mode are measures of central tendency that are commonly used to describe frequency distributions. Measures of central tendency describe the “most

typical" value or observation in a frequency distribution. The measure used to describe the distribution should be based on both the scale of measurement and the determination of which measure best describes the most typical observation in the frequency distribution. Measures of spread are the variance, the standard deviation, and the range. The range tells us the distance between the lowest and highest values in a distribution, so in general, it does not provide much information about the frequency distribution. The standard deviation is the most common descriptive statistic used to describe the spread of a frequency distribution for metric variables.

Percentiles are measures of spread and location often used in benchmarking. Percentiles are used to provide comparative information between health care organizations.

ADDITIONAL RESOURCES

Jekel, J., et al. 1996. *Epidemiology, biostatistics and preventive medicine*. Philadelphia: W.B. Saunders.

Stockburger, D.W. 1998. Introductory statistics: Concepts, models and applications. Available on the Internet at http://www.psychstat.smsu.edu/introbook/sbk13m.html.

U.S. Department of Health and Human Services, Public Health Service. 1992. *Principles of epidemiology: An introduction to applied epidemiology and biostatistics*. Atlanta, GA: USDHHS.

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, Office of Analysis, Epidemiology, and Health Promotion, Compressed Mortality File (CMF) compiled from CMF 1968–1988, Series 20, No. SQ 2000, CMF 1989–1998, Series 20, No. 2E 2003 and CMF 1999–2001, Series 20, No. 2G 2004 on CDC Wonder On-Line Database. http://wonder.cdc.gov.

Weisberg, H.F. 1992. *Central tendency and variability*. Newbury Park, CA: Sage Publications.

Appendix 4–A

Exercises for Solving Problems

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
2. What is the difference between a population parameter and a sample statistic?
3. Compare and contrast the following measures of central tendency: mean, median, and mode.
4. Define polarization, individuality, and uniformity in relation to variability in a frequency distribution.
5. Why do measures of central tendency and variation of ungrouped frequency distributions differ from those of grouped frequency distributions?

MULTIPLE CHOICE

1. There are 40 students in section I of a medical terminology class and 20 students in section II. The mean score on a midterm exam for section I is 60, and the mean score for section II is 70. What is the mean score for the two classes combined?
   a. 63.3
   b. 65
   c. 67
   d. not enough information provided to answer question
2. For the two scores 8 and 12, $\sum (X - \overline{X})^2$ is:
   a. 2
   b. 4
   c. 8
   d. 12

3. In a frequency distribution, the lowest score is 25 and the highest score is 50; the mean is 37.5. The range is:
   a. 11
   b. 12.5
   c. 24
   d. 25

4. In grouping a set of scores from a frequency distribution, the width of the class intervals should:
   a. be equal
   b. be at least one
   c. vary according to the frequency within the interval
   d. be no more than 10

5. The first class interval in the grouped frequency distribution is 5–10. The width of the interval is:
   a. 5
   b. 5.5
   c. 6
   d. 6.5

6. The midpoint of the class interval 5–10 is:
   a. 7
   b. 7.5
   c. 8
   d. 8.5

7. The width of a class interval is 3. The midpoint is 9. The apparent lower limit of the interval is:
   a. 7
   b. 7.5
   c. 8
   d. 8.5

8. The "real limits" of the class interval 1–3 are:
   a. 0.5–3.5
   b. 1–3
   c. 0–4
   d. 1.5–2.5

The following table displays a cumulative frequency distribution of the length of stay of patients discharged from Community Behavior Health Care during the week of April 1. Use the table below to answer questions 9 through 14.

Class Interval for Length of Stay	*f*	*Cum. f*	*Cum. %*
20–24	4	20	100
15–19	8	16	80
10–14	6	8	40
5–9	0	2	10
0–4	2	2	10

9. Six patients had lengths of stay that fall:
   a. below 14.5 days
   b. at 12 days
   c. above 9.5 days
   d. between 9.5 days and 14.5 days

10. Twelve patients had lengths of stay that were greater than:
    a. 14.5 days
    b. 15 days
    c. 17 days
    d. 19.5 days

11. Sixteen patients had lengths of stay below:
    a. 20 days
    b. 19.5 days
    c. 17 days
    d. 14.5 days

12. The cumulative frequency value of "8" means that 8 cases fall below:
    a. 14.5 days
    b. 14 days
    c. 12 days
    d. 10 days

13. Forty percent of the patients discharged had lengths of stay that fell below:
    a. 14.5 days
    b. 14 days
    c. 12 days
    d. 10 days

14. Eighty percent of the patients discharged had lengths of stay that fell below:
    a. 19.5 days
    b. 19 days
    c. 17 days
    d. 14.5 days

15. Which of the following is not a measure of spread?
    a. mode
    b. range
    c. standard deviation
    d. variance

16. Review the following two frequency distributions:

    1: 200 210 190 220 195
    2: 210 170 180 235 240

    The standard deviation for distribution 1 is:
    a. the same as that for set 2
    b. less than that for set 2
    c. greater than that for set 2
    d. not enough information provided

17. The standard deviation of a frequency distribution is 5. The variance is:
    a. 10
    b. 15
    c. 20
    d. 25
    e. not enough information provided

18. In a frequency distribution, the mean is 32. If each score is divided by 2, the mean of the new distribution is:
    a. 64
    b. 32
    c. 16
    d. not enough information provided

19. In the frequency distribution 4 4 5 7, the number 4 is the:
    a. mean
    b. mode
    c. median
    d. range

20. Review the grouped frequency distribution that follows:

Class Interval	*f*
40–44	*2*
35–39	*3*
30–34	*6*
25–29	*5*
20–24	*4*

After the data were grouped into class intervals, it was determined that a value of 35 was really a value of 39. Correcting this error will change the calculated value of:
a. the mean
b. the median
c. the mode
d. all of the above
e. none of the above

21. A percentile rank may take on any of the following values except:
a. 37
b. 50
c. 87.4
d. 103

PROBLEMS

1. Review the data in Tables 4–A–1 and 4–A–2 and answer the questions that follow. Use an electronic spreadsheet to assist you in preparing the answers.

Table 4–A–1 Male Deaths Due to Leukemia (ICD-9-CM Codes 200.0–200.9) in the state of Ohio, 1998

Age Group	*Leukemia Deaths in Men*	*p*	*Cum. p*	*M*	*f(M)*
5–14	16				
15–24	20				
25–34	27				
35–44	63				
45–54	118				
55–64	194				
65–74	388				
75–84	418				
85+	124				
Total	1,368				

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC Wonder On-Line Database, wonder.cdc.gov.

Table 4–A–2 Female Deaths Due to Leukemia (ICD-9-CM Codes 200.0–200.9) in the state of Ohio, 1998

Age Group	*Leukemia Deaths in Women*	*p*	*Cum. p*	*M*	*f(M)*
5–14	6				
15–24	10				
25–34	9				
35–44	26				
45–54	61				
55–64	132				
65–74	296				
75–84	463				
85+	194				
Total	1,197				

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC Wonder On-Line Database, wonder.cdc.gov.

a. What is the mean age of death for men? For women?
b. What are the crude modes and median ages of death for men? For women?
c. What are the refined mode and the refined median for men? For women?
d. Compare and contrast the crude and refined results for each group. Explain any disparities that may exist.
e. In analyzing the results of your data for men and women, what conclusions can you draw?

2. Use the data of the ages of 61 patients discharged from DRG 127, Heart Failure and Shock, that appear in Exhibit 4–A–1 to solve the following:

Exhibit 4–A–1 Ages of 61 Patients Discharged from DRG 127, Heart Failure and Shock

37	51	60	65	75	80
37	51	60	65	76	80
39	52	61	66	76	82
39	53	63	66	77	83
42	54	63	69	77	83
47	54	63	70	77	84
47	56	64	72	78	85
47	57	64	73	79	86
49	57	64	73	80	87
51	59	64	75	80	88
					88

a. Use statistical software to calculate the mean, median, mode, variance, and standard deviation for the ungrouped frequency distribution and to prepare a frequency table.
b. Group ages into class intervals. Prepare a table that displays the frequencies for each class interval, the cumulative frequency, the relative proportion, and the cumulative percent.
c. Compute the mean, median, mode, variance, and standard deviation for the grouped data.
d. Compare the results of the grouped and ungrouped frequency distributions.

3. The lengths of stay for a group of patients discharged from DRG 127 are presented in Exhibit 4–A–2.

Exhibit 4–A–2 Length of Stay for Patients Discharged from DRG 127, Heart Failure and Shock

1	2	3	5	8	14
1	3	3	5	8	15
1	3	3	5	8	16
1	3	3	5	10	17
1	3	3	5	10	27
1	3	3	6	10	36
2	3	4	6	10	
2	3	4	6	11	
2	3	4	6	11	
2	3	4	7	11	
2	3	4	8	13	

   a. Use statistical software to calculate the mean, median, mode, variance, and standard deviation for the ungrouped frequency distribution and to prepare a frequency table.
   b. Group ages into class intervals. Prepare a table that displays the frequencies for each class interval, the cumulative frequency, the relative proportion, and the cumulative percent.
   c. Compute the mean, median, mode, variance, and standard deviation for the grouped data.
   d. Compare the results of the grouped and ungrouped frequency distributions.

CLASS ACTIVITY

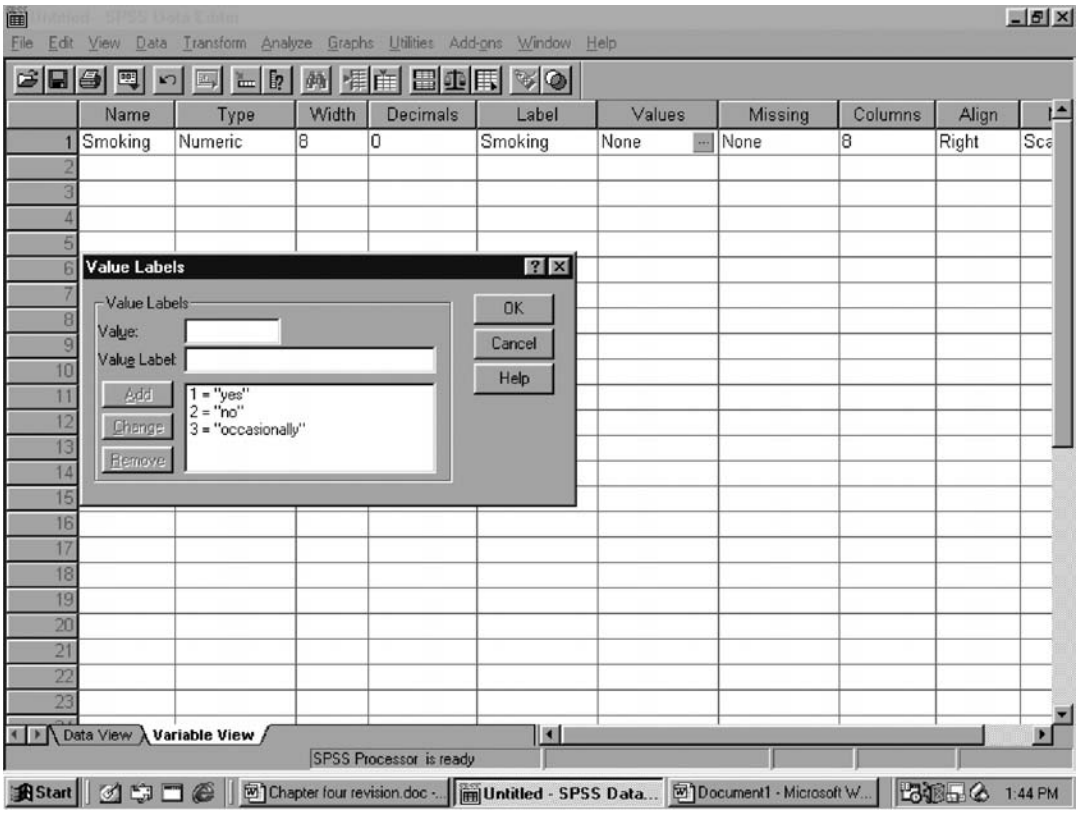
An important aspect of the health information manager's job is to be able to collect and analyze data. But data are not always useful in their raw form; they must be turned into information. This activity is designed to provide you with the experience in preparing a written report based on the analysis of the collected data. For this activity, you will:

   a. set up an SPSS (or other microcomputer statistical package) data sheet
   b. define and label variables for the data sheet
   c. summarize data
   d. prepare charts and graphs

Instructions

1. Complete the survey that follows these instructions.
2. Turn the survey in to the instructor. Copies of each student's responses will be distributed to the class.
3. Set up the SPSS (or other) data sheet. Define the variables for the data input sheet. Categorical variables have been coded on the data sheet. An example of how to enter the variable information on the data sheet appears in Figure 4–A–1.
4. Input the data from each completed survey onto the data sheet.

Figure 4–A–1 Example of How to Enter Variable Information for Class Activity on SPSS Data Sheet.
Source: SPSS 12.0 for Windows, Copyright SPSS, Inc., 2003.



5. Answer the questions that follow.
6. Prepare a two- to three-page written report describing the characteristics of the HIM class based on your analysis of the data. Include statistical tables, charts, and/or graphs to support your findings.

Survey Questions

1. What percentage of students live on campus? Off campus?
2. What percentage of the class is female? Male?
3. What is the average height of the class? What is the average height of the men in the class? What is the average height of the women in the class?
4. What are the average heights of the mothers and the fathers?
5. What proportion of the class is right-handed? What percentage of the mothers is right-handed? How many fathers are right-handed?
6. Prepare a bar graph that displays the actual frequency distribution of the hair color of your classmates. What is the modal hair color? Prepare a pie chart that displays hair color by percentage.
7. Prepare a table displaying the frequency distribution of the eye color of your classmates.
8. What proportion of the class went to a public high school?
9. What is the average amount of time that the class spends studying for HIM classes? Prepare a histogram that displays the frequency distribution of time spent studying.
10. What is the average amount of money spent on haircuts for the class? What is the average amount for men? For women?
11. What is the average amount of CDs owned by the class? How many CDs do the men own? The women?
12. What is the average amount of time spent exercising by the class per week? What is the average amount of time for men? For women?
13. What is the average number of hours per week spent watching television? What is the average amount of time for men? For women?

HIM CLASS SURVEY

1. Where do you live?
   ☐ campus housing (1)
   ☐ off-campus housing (2)
2. Sex
   ☐ male (1)
   ☐ female (2)
3. What is your height in inches? ____
4. What is your father's height in inches? ____

5. What is your mother's height in inches? ____
6. What is your shoe size (length, not width)? ____
7. Do you smoke?
   - ☐ yes (1)
   - ☐ no (2)
   - ☐ occasionally (3)
8. Are you:
   - ☐ left-handed (1)
   - ☐ right-handed (2)
9. Is your father:
   - ☐ left-handed (1)
   - ☐ right-handed (2)
10. Is your mother:
    - ☐ left-handed (1)
    - ☐ right-handed (2)
11. What is your hair color?
    - ☐ black (1)
    - ☐ brown (2)
    - ☐ blond (3)
    - ☐ red (4)
    - ☐ other (5)
12. What is your eye color?
    - ☐ black (1)
    - ☐ brown (2)
    - ☐ blue (3)
    - ☐ green (4)
    - ☐ gray (5)
    - ☐ hazel (6)
    - ☐ other (7)
13. What type of high school did you attend?
    - ☐ public high school (1)
    - ☐ private high school (2)
14. On average, how many hours per week do you spend studying for HIM classes? ____
15. How much did you spend, to the nearest dollar, on your last haircut, including tip? ____
16. How many CDs do you own? ____
17. On average, how many hours per week do you spend exercising? ____
18. On average, how many hours per week do you spend watching television? ____

CHAPTER 5

The Normal Distribution and Statistical Inference

KEY TERMS

- Normal distribution
- Symmetrical
- Asymptotic curve
- Skewness
- Kurtosis
- *z* values
- Standard normal distribution
- Standard normal deviate
- Point estimate
- Central limit theorem
- Standard error of the mean
- Confidence interval
- Sampling methods
  - Simple random sampling
  - Stratified random sampling
  - Systematic sampling
  - Cluster sampling
- Null hypothesis
- Alternative hypothesis
- Type I error
- Type II error
- Level of significance
- *p* value

LEARNING OBJECTIVES

At the conclusion of this chapter, you should be able to:

1. Define key terms.
2. Describe the characteristics of the normal distribution and the standard normal distribution.
3. Compare and contrast the normal distribution and the standard normal distribution.
4. Compare the standard deviation and the standard normal deviate.
5. Convert normal distributions to standard normal distributions using computer statistical software.
6. Explain the central limit theorem.
7. Calculate the standard error of the mean and confidence intervals for samples.
8. Explain how sample size and variation affect the standard error of the mean.

9. Explain the following sampling techniques: simple random sampling, stratified random sampling, systematic sampling, and cluster sampling.
10. Explain the differences between the null and alternative hypotheses.
11. Explain the factors that affect type I and type II errors.
12. Differentiate between the alpha level and the *p* value.
13. Identify the factors that influence sample size.
14. Calculate sample size for given situations.

Much of statistical inference is based on the **normal distribution**, also called the Gaussian distribution for Johann Karl Gauss, the person who best described it. The normal distribution is not a single distribution, but an infinite number of possible distributions. This is important in statistical inference because the population mean can take on any positive or negative value, and the population standard deviation can take on any positive or negative value. Thus, the normal distribution is the most widely used theoretical distribution; many naturally occurring phenomena, such as blood pressure, height, and weight, approximate the normal distribution.

CHARACTERISTICS OF THE NORMAL DISTRIBUTION

There are several characteristics of the normal distribution with which you should be familiar. First, the curve is bell-shaped and **symmetrical** about the mean (the population mean is symbolized by μ, pronounced "mu"). Second, because the distribution is symmetrical, approximately 50% of the observations lie above the mean and 50% of the observations lie below the mean. The total area under the curve is equal to 1.00. Third, in a normal distribution, the mean, median, and mode are equal. The values of the normal distribution range from minus infinity ($-\infty$) to plus infinity ($+\infty$).

As we move out from the center of the normal curve bilaterally, the height of the curve descends gradually at first, then faster, and finally more slowly as it approaches the horizontal axis. Each tail of the curve approaches the *x*-axis but never touches it, no matter how far from center we go. This type of curve is called an **asymptotic curve** because it is considered asymptotic to the horizontal axis.

Figure 5–1 displays the normal distribution and how the values in the distribution are arranged around the population mean, μ. Approximately 68% of the values lie within 1 standard deviation from the mean ($\mu \pm 1.68\ \sigma$); 95% of the observations lie within 1.96 standard deviations from the mean ($\mu \pm 1.96\ \sigma$); and 99% of the observations lie within 2.58 standard deviations of the mean ($\mu \pm 2.58\ \sigma$). These characteristics of the normal curve are important when making inferences about population parameters from sample statistics. The symbols used to distinguish between population parameters and sample statistics appear in Table 5–1.

In a normal distribution, we know that the population μ can take on any value and that the population μ is the midpoint of the distribution. Compare Figures 5–1 and 5–2. Even

Figure 5–1 Histogram of Normal Distribution 1

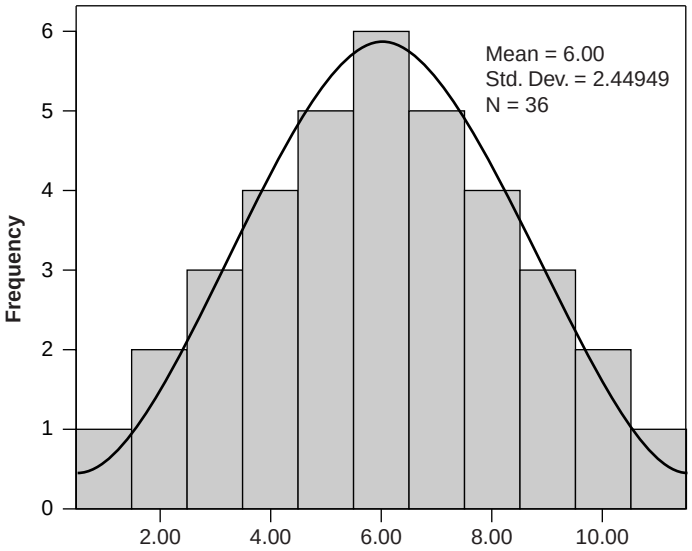


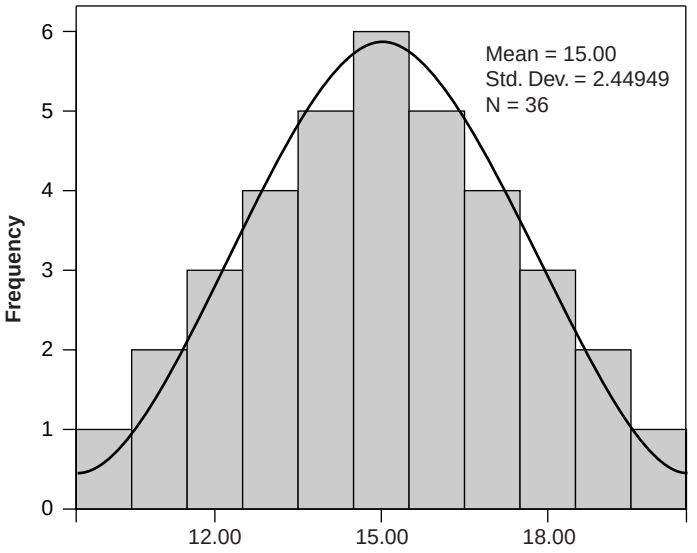
Table 5–1 Statistical Symbols

	Population Parameter	*Sample Statistic*
Mean	μ	$\bar{X}$
Variance	σ	s^2
Standard Deviation	σ^2	s

though the means for each distribution—6 and 15, respectively—are different, the shapes of the distributions are the same. Any change in μ without a corresponding change in σ does not change the shape of the distribution. However, changes in the value σ do change the shape of the distribution but do not affect the midpoint. Basically, changes in σ affect the dispersion of the values in the distribution. Dispersion can affect the **skewness** of the distribution.

A frequency distribution that is asymmetrical is skewed, and in this case the mean, median, and mode will take on different values. Skewness is the horizontal stretching of a frequency distribution to one side or the other, so that one tail is longer than the other. The longer tail has more observations. Because the mean is sensitive to extreme values, the mean moves in the direction of the long tail when a distribution is skewed. When the direction of the long tail is off to the right, the distribution is said to be positively skewed, or skewed to the right. Conversely, when a distribution's long tail is off to the left, the distribution is said to be negatively skewed, or skewed to the left. We can determine if a distribution is skewed through SPSS by checking these items in the dialog box after requesting "Frequencies" or "Descriptives" from the "Analyze" menu. An obtained skewness value greater than 1 is an indication that the distribution differs significantly from normal. Another way to assess the

Figure 5–2 Histogram of Normal Distribution 2



skewness of a distribution is to compare the mean and median. If the mean and median approximate one another, the distribution is probably not significantly skewed.

Kurtosis is the vertical stretching of the frequency distribution. If the distribution appears to be more peaked or more flattened than the normal distribution, it is considered to be kurtotic. For a normal distribution, the value of the kurtosis statistic is zero. A positive kurtosis indicates that the frequency distribution has longer tails than the normal distribution and that the observations are clustered toward the center (peakedness). If the kurtosis statistic is negative, the frequency distribution has tails shorter than the normal distribution, and there is less clustering of the observations (flattened).

THE STANDARD NORMAL DISTRIBUTION (*z* DISTRIBUTION)

Since there are number of normal distributions, which may have any mean and any standard deviation, the observations in the distribution must be standardized when we want to make comparisons between distributions. When we standardize a frequency distribution, such as one for the variable "age," we are transforming the units of measurement (age) to a unit-free form—that is, the age units become ***z* values**. The *z* distribution is referred to as the **standard normal distribution**; it has a mean of 0 and a standard deviation equal to 1. The *z* value, also called the **standard normal deviate**, is the number of standard deviation units that the observed value lies away from the mean, μ. Transforming our raw observations to *z* values makes it possible to make comparisons between distributions.

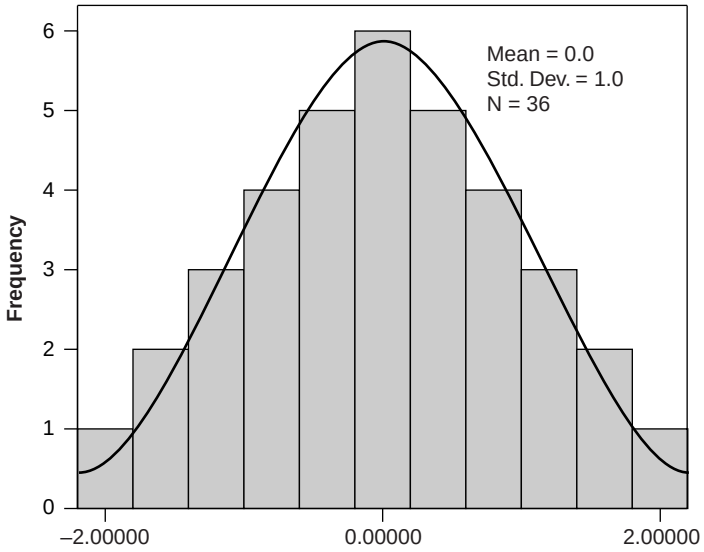
In the standardized normal distribution, the area between the *z* values of ± 1.0 is 68%, the area between the *z* values of ± 1.96 is 95%, and the area between $+3$ and -3 is 99.7%. Any

normal distribution may be transformed into a standardized normal distribution through the formulas in Exhibit 5–1. The transformation of a set of observations to *z* values is a linear transformation that does not change the shape of the distribution. This is illustrated in Figure 5-3, where the normal distribution in Figure 5–1 is transformed into a standard normal distribution with a mean equal to 0 and a standard deviation equal to 1.

Exhibit 5–1 Formulas for Transforming Observed Scores into *z* Scores

z *Values in a Population*	z *Value in a Sample*
$z = (X - \mu)/\sigma$	$z = (X - \overline{X})/s$
Where X is the value of the observation, μ is the mean of the population distribution, and σ is the standard deviation of the population distribution.	Where X is the value of the observation, $\overline{X}$ is the mean of the sample distribution, and s is the standard deviation of the sample distribution.

Figure 5–3 Standard Normal Distribution



For a "real-life" example, we will now compare a normal distribution with a standardized normal distribution to explain some of these concepts. In Table 5–2, a frequency distribution of the age of 100 nursing home residents is displayed. Using these data, we will construct a histogram of the frequency distribution for age using SPSS. We will also request descriptive statistics and statistics on skewness and kurtosis to evaluate the "normality" of the distribution. The histogram appears in Figure 5–4.

Table 5–2 Frequency Distribution of Age of 100 Nursing Home Residents

		Age			
		Frequency	*%*	*Valid %*	*Cumulative %*
Valid	65.00	3	3.0	3.0	3.0
	66.00	3	3.0	3.0	6.0
	67.00	4	4.0	4.0	10.0
	68.00	2	2.0	2.0	12.0
	69.00	3	3.0	3.0	15.0
	70.00	3	3.0	3.0	18.0
	71.00	4	4.0	4.0	22.0
	72.00	5	5.0	5.0	27.0
	73.00	4	4.0	4.0	31.0
	74.00	5	5.0	5.0	36.0
	75.00	12	12.0	12.0	48.0
	76.00	5	5.0	5.0	53.0
	77.00	6	6.0	6.0	59.0
	78.00	7	7.0	7.0	66.0
	79.00	4	4.0	4.0	70.0
	80.00	7	7.0	7.0	77.0
	81.00	4	4.0	4.0	81.0
	82.00	4	4.0	4.0	85.0
	83.00	3	3.0	3.0	88.0
	84.00	2	2.0	2.0	90.0
	86.00	1	3.0	3.0	93.0
	87.00	1	1.0	1.0	94.0
	88.00	1	1.0	1.0	95.0
	89.00	1	1.0	1.0	96.0
	90.00	3	3.0	3.0	99.0
	91.00	1	1.0	1.0	100.0
	Total	100	100.0	100.0	

The descriptive statistics in Exhibit 5–2 indicate that the mean, median, and mode are similar; 76.36, 76.0, and 75, respectively. We will interpret the distribution as "normal" because of the similarity between the three measures of central tendency. The skewness statistic is 0.261, indicating that the distribution does not depart significantly from normal. However, the distribution is somewhat flat, as indicated by the kurtosis statistic, −0.281.

We will now convert our "normal" distribution to a standardized normal distribution. The histogram and descriptive statistics appear in Figure 5–5 and Exhibit 5–4 respectively. We can convert the scores to standard scores by checking the "Save standardized values as variables" option in the "Descriptives" dialog box, as shown in Exhibit 5–3. The SPSS program will create and save a new variable on the data sheet. The mean of the new standardized distribution is 0, and the standard deviation is 1. Though "the bars" on the standardized histogram are not exactly the same as those in Figure 5–4, the skewness statistic and kurtosis statistic remain the same, indicating that the shape of the distribution did not change when the distribution was standardized.

Figure 5–4 SPSS Output for Histogram on Ages of Nursing Home Residents

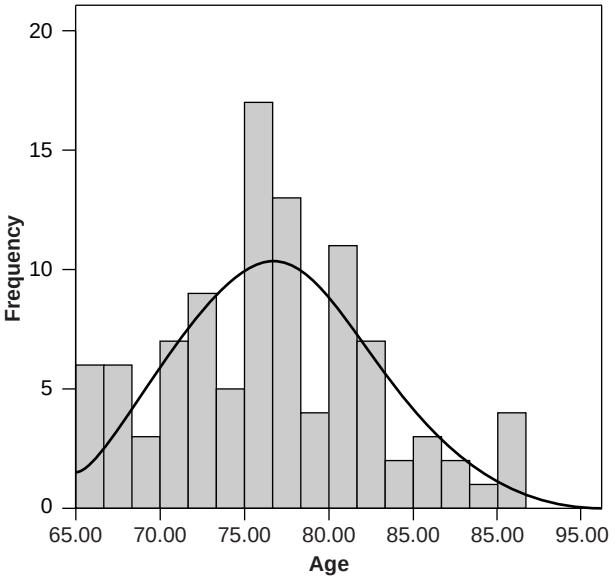
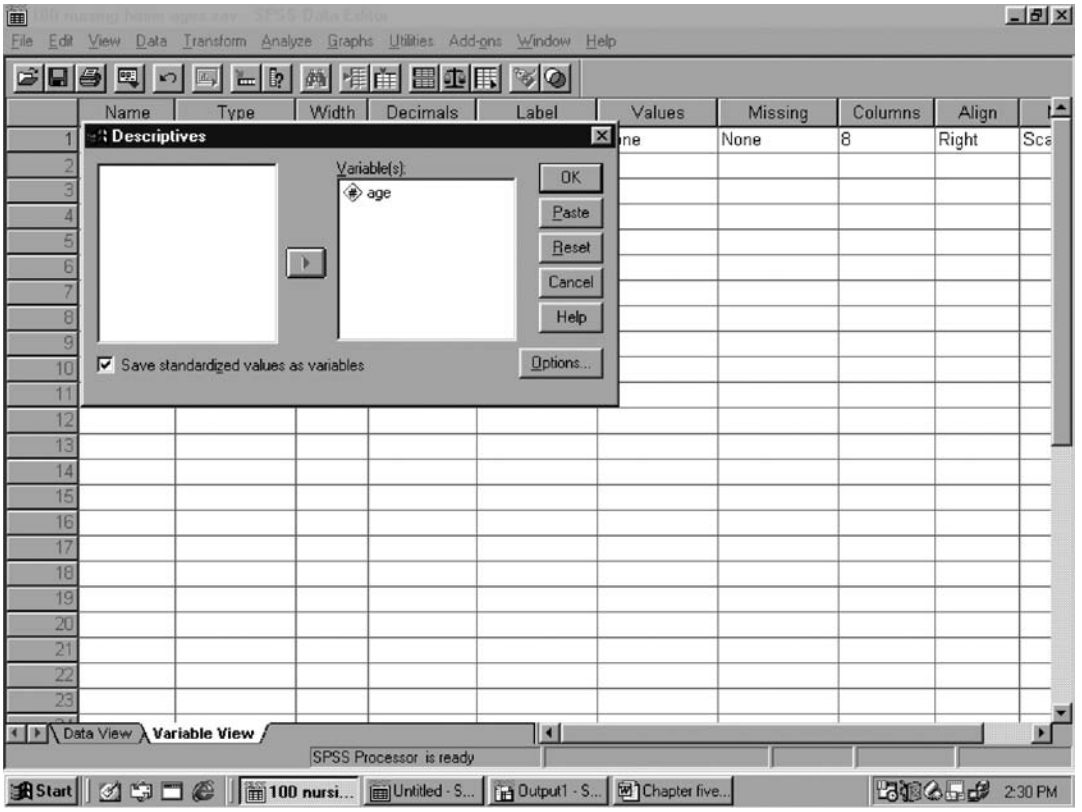


Exhibit 5–2 SPSS Output of Descriptive Statistics, Age of Nursing Home Residents

N	Valid	100
	Missing	0
Mean		76.3600
Median		76.0000
Mode		75.00
Std. Deviation		6.23856
Variance		38.920
Skewness		0.261
Std. Error of Skewness		0.241
Kurtosis		−0.281
Std. Error of Kurtosis		0.478
Range		26.00

Exhibit 5–3 SPSS Screen for Requesting Standardized Variables.

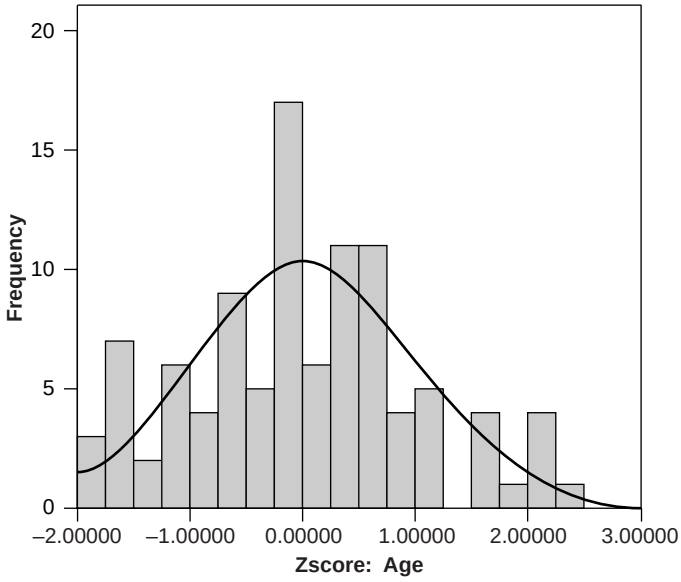


Source: SPSS 12.0 for Windows, SPSS, Inc., 2003.

Exhibit 5–4 SPSS Output for Standardized Descriptive Statistics on Ages of Nursing Home Residents

Z score(AGE)		
N	Valid	100
	Missing	0
Mean		.0000000
Median		−.0577057
Mode		−.21800
Std. Deviation		1.00000000
Variance		1.000
Skewness		.261
Std. Error of Skewness		.241
Kurtosis		−.281
Std. Error of Kurtosis		.478
Range		4.16763

Figure 5–5 SPSS Output for Histogram on Ages of 100 Nursing Home Residents, Standardized Distribution



STATISTICAL INFERENCE

In analyzing health care data, we are usually dealing with data that represent a sample drawn from a larger population. We want to use the sample statistics to describe the larger population. A single sample statistic such as the mean is actually a **point estimate**. A point estimate is a single numerical value computed from the sample that is assumed to best represent the actual population parameter. The properties of the sampling distribution are similar to those of a normal distribution. If the population is normally distributed with a mean μ and a standard deviation σ, the sampling distribution of $\bar{X}$ has the following properties:

1. It has a mean equal to the mean for the population from which the samples were drawn, $\mu_{\bar{x}} = \mu$
2. It has a standard deviation equal to the population standard deviation divided by the square root of the sample size, $\sigma_{\bar{x}} = \sigma\sqrt{n}$
3. It is normally distributed. The sampling distribution of $\bar{X}$ is approximately normal when sampling is from a non-normal population distribution. As the sample size increases, the sample's approximation to normality improves (central limit theorem).

CENTRAL LIMIT THEOREM

When we draw inferences from our sample, we are assuming that the sample represents a frequency distribution that is normally distributed. When a graph of our sample appears normal, we assume that the population from which the sample was drawn is also normally distributed. This is true regardless of the size of the sample drawn from the same population. However, many populations are not normally distributed. Regardless of the type of population distribution, the sampling distribution, if sufficiently large, is approximately normal. The **central limit theorem** summarizes the relationship between the shapes of the population distribution and the sampling distribution of the mean, $\overline{X}$:

> If repeated random samples of size *N* are drawn from a population, and if a mean is calculated for each sample, the distribution of the sample means approaches the normal distribution as *N* becomes large. The mean of the sampling distribution will approach the population mean, μ. This is true even if the population distribution is not normal.

The central limit theorem assures us that regardless of the shape of the population distribution, the sampling distribution of $\overline{X}$ approaches normality as the sample size increases. This is true even when data in individual samples are skewed. In reality, the samples do not have to be very large for the sampling distribution of $\overline{X}$ to be approximately normal. In most instances, the approximation to normality is quite rapid as *N* increases.

These concepts can be illustrated by using the age of nursing home residents' data. Recall that the population mean is 76.4 and the population standard deviation is 6.2. A simple random sample of 38 nursing home residents was drawn from the population. A frequency distribution for the sample is displayed in Table 5–3, and a histogram is displayed in Figure 5–6. The sample mean is 75.2 (Exhibit 5–5), and the sample standard deviation is 6.2. The

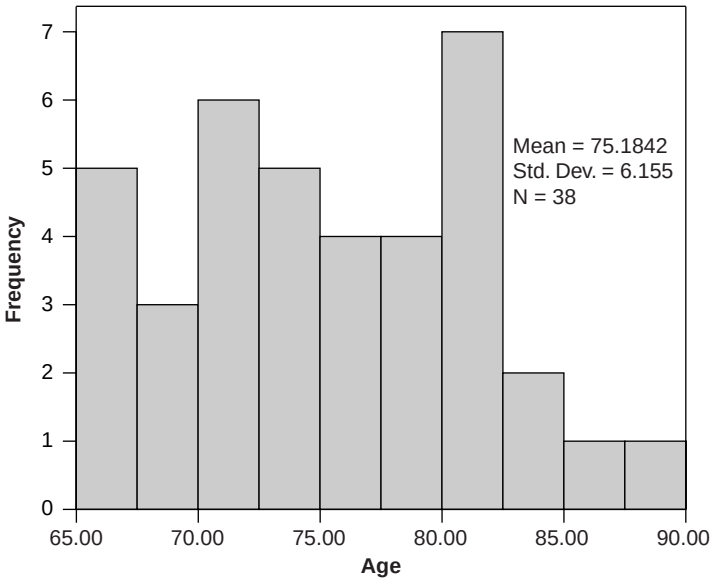
Exhibit 5–5 SPSS Output for Standardized Descriptive Statistics on Ages of a Sample of 38 Nursing Home Residents

	AGE	Valid *N* (listwise)
N	38	38
Minimum	65.00	
Maximum	90.00	
Mean	75.1842	
Std. Deviation	6.15500	
Variance	37.884	
Skewness	.255	
Std. Error for Skewness	.383	
Kurtosis	−.541	
Std. Error for Kurtosis	.750	

Table 5-3 Frequency Distribution of Random Sample of Ages of Nursing Home Residents

Age	*Frequency*	*%*	*Valid %*	*Cumulative %*
65.00	1	2.6	2.6	2.6
66.00	2	5.3	5.3	7.9
67.00	2	5.3	5.3	13.2
68.00	2	5.3	5.3	18.4
69.00	1	2.6	2.6	21.1
70.00	1	2.6	2.6	23.7
71.00	2	5.3	5.3	28.9
72.00	3	7.9	7.9	36.8
73.00	2	5.3	5.3	42.1
74.00	3	7.9	7.9	50.0
75.00	2	5.3	5.3	55.3
76.00	1	2.6	2.6	57.9
77.00	1	2.6	2.6	60.5
78.00	3	7.9	7.9	68.4
79.00	1	2.6	2.6	71.1
80.00	3	7.9	7.9	78.9
81.00	2	5.3	5.3	84.2
82.00	2	5.3	5.3	89.5
83.00	1	2.6	2.6	92.1
84.00	1	2.6	2.6	94.7
86.00	1	2.6	2.6	97.4
90.00	1	2.6	2.6	100.0
Total	38	100.0	100.0	

Figure 5–6 Histogram of Ages of 38 Nursing Home Residents



sample distribution does not significantly depart from normality, as indicated by the skewness statistic of 0.255, which is less than one. However, the distribution is flatter than the population distribution, as indicated by the kurtosis statistic of −0.541. The difference between the population parameters and the sample statistics is sampling error.

STANDARD ERROR OF THE MEAN

When inferences are made about normally distributed data, conclusions are based on the relationships of the standard deviation and mean to the normal curve. The mean of the sample may or may not be the same as the population mean. The difference between the sample mean and the population parameter μ is sampling error. For example, if we draw three different samples from a population, we will get three different means. In addition, if we take many samples from the same population, we will have as many different means, and these means will be normally distributed. The mean of the means will be close to the true population mean.

To determine how close the sample mean is to the population mean, we find the standard deviation of the distribution of means. The standard deviation of the distribution of means is called the **standard error of the mean**, or standard error. The smaller the standard error, the closer the sample mean is likely to be to the population mean. However, we do not need to draw a lot of samples to calculate the standard error; it can be calculated from a single sample:

$$\text{standard error of the mean (SE)} = s/\sqrt{n}$$

The standard error is influenced by the standard deviation and the sample size. The greater the dispersion around the mean, the less certain we are about the actual population mean, and the greater the standard error of the mean. The larger the sample size, the more confidence we have in the mean and the smaller the standard error of the mean. The smaller the standard error, the more reliable the statistic.

The effects of sample size and standard deviation on the standard error are illustrated in the following examples. In a hypothetical frequency distribution on the average age of college graduates, the mean is 22.5 and the standard deviation is 3.5. With a sample size of 100, the standard error of the mean is 0.35; if we increase the sample size to 200, and the mean and standard deviation remain the same, the standard error of the mean decreases to 0.25. If the sample size remains at 100 but the standard deviation increases to 5.5, the standard error of the mean increases to 0.55. These data are summarized in Table 5–4.

The data in Exhibit 5–6 further illustrate these principles. Using SPSS, three simple random samples were drawn from our population of 100 nursing home residents in which the mean is 76.4 and the standard deviation is 6.24.

Table 5–4 Statistical Comparisons by Sample Size

	Sample 1	*Sample 2*	*Sample 3*
N	100	200	100
Mean	22.5	22.5	22.5
SD	3.5	3.5	5.5
SE	0.35	0.25	0.55

Exhibit 5–6 Three Random Samples of Ages of Nursing Home Residents

Random Sample 1		*Random Sample 2*		*Random Sample 3*	
Age	*N*	*Age*	*N*	*Age*	*N*
66	1	65	3	65	1
67	1	66	1	66	3
68	1	70	2	65	2
70	1	72	2	68	1
73	1	73	1	69	2
75	1	75	3	70	2
78	1	76	1	71	3
80	1	77	3	72	1
81	1	78	1	73	2
82	1	80	4	74	5
86	1	81	4	75	2
89	1	82	2	76	2
90	1	83	1	77	3
		84	1	78	4
		89	1	79	2
		90	2	80	3
		91	1	81	3
				82	2
				83	1
				84	1
				85	1
				88	1
				90	1
Total	**13**		**33**		**48**
Mean	77.3		77.7		75.5
S.E. Mean	2.285		1.242		0.874
Variance	67.90		50.92		36.64
SD	8.24		7.14		6.05

Population Mean = 76.4

The sample size in sample 1 is 13 and the mean, 77.3, is 0.9 units above the population mean. In sample 2, the sample size is 33, and the mean, 77.7, is 1.3 units above the population mean. The standard deviation for sample 1 ($s = 8.24$) is greater than that for sample 2 ($s = 7.14$), indicating more variation and a greater standard error for sample 1 over sample 2—8.24 versus 7.14, respectively.

However, when we increase the sample size to 48, the observed mean, 75.5, and the standard deviation, 6.05, are less than in either of the previous two samples; there is also a corresponding decrease in the standard error, which is 0.874 for sample 3.

CONFIDENCE INTERVALS

The sample mean, $\overline{X}$, is a point estimate of the population mean, μ. With the additional information of the standard error of the mean, $s_{\overline{x}}$, and our knowledge of the normal curve, we can estimate the limits within which the true population mean probably lies. This is called a **confidence interval** on μ, and gives a range of values that might reasonably contain the true population mean. The confidence interval is represented as

$$a <= \mu <= b.$$

Using the data from sample 1 in Exhibit 5–6, we can add and subtract the standard error of the mean from the sample mean:

$$\text{point estimate} \pm \text{error} = \overline{X}_1 \pm s_{\overline{x}_1} = 77.3 \pm 2.285$$
$$75.0 \leq \mu \leq 79.6$$

The results indicate that the true population mean falls within ± 1 standard error on each side of the sample mean. This is interpreted as meaning that if we draw a number of samples from the population, 68% of these sample means will fall between 75.0 and 79.6. From this we can infer that we are 68% confident that the population mean lies within these limits.

But when developing confidence intervals, we generally want to be more confident about the sample statistic. Generally, the confidence intervals are set at 95%. As noted earlier in this chapter, 95% of the area under the standard normal curve lies between $+1.96$ and -1.96 standard deviations of the mean. The 95% confidence interval (CI_{95}) may be constructed as follows:

$$CI_{95} = \overline{X} \pm 1.96(s_{\overline{x}})$$

Substituting into our formula the CI_{95} for sample 1 in Exhibit 5–6 yields

$$CI_{95} = 77.3 \pm 1.96(2.285) = 77.3 \pm 4.48$$

Thus, we can say that we are 95% confidant that our true population mean lies between 72.8 and 81.8. This confidence interval is expressed as:

$$72.8. \leq \mu \leq 81.8$$

As you can see from the population data in Exhibit 5–6, the true population mean, 76.4, lies within this band. Note that by increasing the confidence interval from 68% to 95% we have increased the range within which the true mean may fall. Increasing the size of the confidence interval increases our confidence that the true population mean lies within that interval.

Correspondingly, we can compute the CI_{95} for samples 2 and 3:

Sample 2

$$CI_{95} = 77.7 \pm 1.96(1.242) = 77.7 \pm 2.4343$$

$$75.3 \leq \mu \leq 80.1$$

Sample 3

$$CI_{95} = 75.5 \pm 1.96(0.874) = 75.5 \pm 1.713$$

$$73.8 \leq \mu \leq 77.2$$

Sample 1 has the greatest variation, as indicated by the standard deviation; thus, it also has the greatest standard error and the widest confidence band.

The general formula for calculating the confidence interval is

$$CI_{\mu} = \bar{X} \pm \alpha(s_{\bar{x}})$$

where α is the confidence coefficient. For the normal distribution, the 95% confidence interval, the confidence coefficient is equal to 1.96; for a confidence interval of 99%, the confidence coefficient is equal to 2.58 ($\pm$2.58 standard deviations from the mean).

SAMPLING METHODS

Before we can make inferences about a population, we must select a sample. Samples should be as representative of the underlying population as possible. If the sample is representative, inferences made from the sample about the population will be correct. There are two general sampling techniques: probability and nonprobability sampling. In probability sampling, each member of a population has a known probability of being selected for the sample. Nonprobability samples are those in which members of a sample are deliberately selected for a specified purpose. One example is the selection of patients admitted to the emergency department in January to study the effects of a new anticoagulant. In nonprobability sampling, generalization of results to a population is extremely limited. Nonprobability sampling is often used in conducting clinical trials.

We will limit our discussion to several methods of probability sampling: **simple random sampling**, **stratified random sampling**, **systematic sampling**, and **cluster sampling**. In

simple random sampling, each member of a population has an equal chance of being selected for inclusion in the sample. The selection of one member of the population has no influence on the selection of another. The drawing of numbers in state lotteries is an example of a simple random sample.

Occasionally we are interested in studying different strata within a population. A stratum is a variable by which the population can be subdivided. Common examples in health care include dividing the human population by age or sex, or categorizing acute care facilities by size. Basically, we group the population into subcategories. In stratified random sampling, we want to draw a sample so that each stratum within the population is proportionately represented in the sample. Before drawing the sample, we must divide the population into the strata we are interested in studying and then randomly draw the appropriate sample from each stratum. If we are studying a problem where sex is an important variable, we will want a stratified sample that is composed of 50% men and 50% women.

In systematic sampling, we select every *k*th member of a population from a list—such as selecting the records of every 5th patient discharged for closed medical record review. If we want to select a sample of 10 from a population of 100, we will select every 10th name from the discharge list. Caution must be exercised with systematic sampling if the list is ordered in any way. For example, if a list of university students is listed by class rank, the resulting sample may not be representative of the population. In systematic sampling, every member of the population does not have an equal chance of being selected for the sample. Selection for inclusion in the sample is dependent upon the first member selected for inclusion in the sample.

In cluster sampling, the sampling units are groups rather than individuals. For example, if we want to survey physicians in the state of Ohio, we can define the population as physicians practicing in acute care facilities in the state of Ohio. The units to be sampled are hospitals—clusters; then, from each hospital included in the sample, a sample of physicians is surveyed. This is two-stage sampling. We first randomly select hospitals, and then randomly select physicians that practice at the hospitals.

HYPOTHESIS TESTING AND STATISTICAL SIGNIFICANCE

With statistical inference, we are interested in making generalizations about a particular population from a sample drawn from that population. When we generalize, we are describing the population from our sample statistics. Often, however, we are interested in determining whether two population means are different with respect to a given variable, such as age, length of stay (LOS), and/or total charges. To determine whether two means are different, we must first develop a hypothesis and perform a statistical test to determine if the observed differences between the means are statistically significant. We want to know whether the observed difference between the group means is greater than what would be expected by chance alone.

To understand the concept of statistical significance, consider the following example. Let's say that the researcher is interested in determining whether physician A's practice profile is superior to physician B's, as indicated by the LOSs of their respective patients. The

researcher will evaluate the differences between practice profiles of these two physicians on the basis of the patients' average length of stay (ALOS) for physicians A and B. In addition, the researcher will want to know if the "observed" difference between the two mean LOSs is due to something other than measurement error. The observed differences in the patients' mean LOSs could be due to the following reasons:

1. The practice profile of physician A is actually superior to the practice profile of physician B.
2. Some confounding factor that was not controlled in any way, such as the age or type of patient, may account for the difference.
3. Random variation could account for the difference.

Only after the second and third reasons have been ruled out can we say that the practice profile of physician A is superior to the practice profile of physician B. To rule out reason 2, we have to design a study that does not permit any extraneous factors that might bias the comparison. To rule out reason 3, we test for statistical significance.

But before we select a statistical test for comparing the two means, we must develop a hypothesis for statistical testing. For tests involving the comparison of two or more groups, the null hypothesis states that there is no difference between the population means from which the two samples were drawn. The **null hypothesis** is consistent with the idea that the observed difference between the means of two or more groups is due to random variation in the data. The null hypothesis is expressed as

$$H_0: \mu_A = \mu_B$$

The interpretation is that the population mean for group A is equal to the population mean for group B.

After we have developed the null hypothesis, we must state the alternative hypothesis. The **alternative hypothesis** states what our theory is, or what we expect to happen as a result of the statistical test. The alternative hypothesis may take one of several forms:

1. $H_A: \mu_A \neq \mu_B$
2. $H_A: \mu_A < \mu_B$
3. $H_A: \mu_A > \mu_B$

In the first example, we are stating that we expect that the two population means will not be equal. We are interested only in whether the observed differences between the two population means are significant. In the latter two examples, we are stating a direction in which we expect the population means to differ. In the second example, we are stating that we expect the mean for population A to be significantly less than the mean for population B. In the third example, we are stating that we expect the mean for population A to be greater than the mean for population B.

For our comparison of the practice profiles of the two physicians, our alternative hypothesis will take the form of the second example. This could indicate that physician A was superior to physician B with regard to LOS.

LEVEL OF SIGNIFICANCE

To determine if physician A is more effective than physician B, we select an appropriate test of statistical significance and establish an appropriate alpha level, such as $\alpha = 0.05$ or 0.01. The alpha level is the maximum probability of rejecting the null hypothesis when it is true. This is referred to as a **type I error**. If alpha is set at 0.05, we run a 5% risk of error when we reject the null hypothesis—that is, when we state that the means of the two groups are different. A **type II error** occurs when we accept the null hypothesis when it is false.

The null hypothesis is rejected only if the sample results are so different from the hypothesis that the probability of such a difference occurring by chance alone is very low or insignificant. The lower the significance level—for example, $\alpha = 0.01$—the more the sample data must depart from the null hypothesis to be statistically significant. An alpha set at 0.01 is considered to be more strict than an alpha set at 0.05—that is, it is more difficult to reject the null hypothesis when alpha is set at 0.01.

When we reject the null hypothesis, we actually support what we believe to be true. Rejecting the null hypothesis supports our theory. Failure to reject the null hypothesis does not mean that the null hypothesis is true; it only means that we did not prove that the observed difference between the means of the two groups was statistically significant beyond a reasonable doubt. The level of significance refers to the probability of making a type I error. For a type I error, the level of significance is designated by α (alpha). The level of significance is usually set at 0.01 or 0.05. For small sample sizes, alpha is usually set at 0.05; for large sample sizes, alpha is usually set at 0.01. This is because it is easier to achieve statistical significance with large samples. If the significance for α is set at 0.05 and the null hypothesis is rejected, the probability of a type I error is 5%. In the long run, with the drawing of multiple samples from the same population, the rejection of a true null hypothesis will occur 5% of the time. Conversely, the sample data will justify accepting the true null hypothesis 95% of the time.

The probability of committing a type II error is designated by β (beta). Type II errors occur only when we incorrectly fail to reject the null hypothesis. Because the level of significance is set to reduce the probability of type I error, the probability of a type II error is increased. However, the probability of making a type II error decreases as the sample size increases. Table 5–5 shows the probabilities of type I and type II errors.

Table 5-5 Probabilities of Type I and Type II Errors

Action	*H_0 is True*	*H_0 is False*
Reject H_0	Type I error α	Correct $1 - \beta$
Accept H_0	Correct $1 - \alpha$	Type II error β

THE *p* VALUE

The **level of significance** is determined before conducting the statistical test of significance, or *a priori*. The ***p* value** is obtained from the statistical test of significance and indicates the probability that the observed difference between the means could have been obtained by chance alone, given random variation and a single test of the null hypothesis. If the obtained *p* value is 0.03, the correct interpretation is that the probability of obtaining the test statistic at least as extreme as the one calculated is 3%. That is, only 3% of all possible samples will produce a test statistic as extreme as the calculated test statistic if the null hypothesis is true. If the *p* value obtained from the statistical test is less than or equal to the preset alpha level, the result is considered sufficiently rare so that the null hypothesis is rejected.

Exhibit 5–7 provides an example of a statistical test for the difference between two population means using Excel. In the example, the *z* test for the difference between two population means was conducted to determine if the ALOS varied by sex for a particular diagnosis-related group (DRG). In the example, the LOSs for males and females discharged from DRG XXX were compared. The mean LOS for males is 4.26 days, and the mean LOS for women is 4.95 days. Prior to conducting the *z* test, the null and alternative hypotheses were stated and the alpha level was set:

Exhibit 5–7 *z* Test for the Difference Between Population Means, Excel Output

	Male	*Female*
Mean	4.26	4.95
Known Variance	9.60	9.00
Observations	134.00	151.00
Hypothesized Mean Difference	0.50	
z	−3.27	
P(*Z*<=*z*) one-tail	0.00	
z Critical one-tail	1.64	
P(*Z*<=*z*) two-tail	0.00	
z Critical two-tail	1.96	

The null hypothesis states that there is no difference between the ALOS for patients by sex for DRG XXX; the alternative hypothesis states that there is a difference between the ALOS of the patients by sex for DRG XXX.

The Excel output provides the calculated z statistic and its corresponding *p* values, which indicate whether the means of the two groups are significantly different from each other. (Exhibit 5–8 provides an explanation of each row in the Excel output.) In Exhibit 5–7, the calculated value of *z* is −3.27. In comparing the LOSs for the two groups, the procedure automatically transformed the LOSs for the two groups to *z* scores. Excel provides two other

Exhibit 5–8 Explanation of Excel Output, *z* Test for the Difference Between Two Population Means

Mean: The arithmetic mean for each group, male and female
Known Variance: The population variance for each group must be provided in order to conduct the *z* test. This will be discussed in Chapter 6.
Observations: The sample size for each group.
Hypothesized Mean Difference: When conducting the *z* test, the investigator must specify the difference between the two means that he/she believes to be important. No difference was specified for this example.
***Z*:** The **calculated** value of *z* (−3.2733) as a result of conducting the statistical test.
***p*(*Z*<=*z*) one-tail:** The *p* value. The probability that the **critical** value of *z* (1.64) is less than or equal to the **calculated** value of *z* (−3.2733) for a one-tailed test.
***z* critical one-tail:** The **critical** value of *z* (1.64) for a one-tailed test when $\alpha = 0.05$. It should be interpreted as ±1.64 depending on the direction of the test.
***p*(*Z*<=*z*) two-tail:** The *p* value. The probability that the **critical** value of *z* (1.96) is less than or equal to the **calculated** value of *z* (−3.2733) for a two-tailed test.
***z* critical two-tail:** The **critical** value of *z* (1.96) for a two-tailed test when $\alpha = 0.05$.

z values: 1.96 and 1.64. These values are interpreted as ±1.96 and ±1.64. These are the critical values of *z* for both one- and two-tailed tests when alpha is set at 0.05. Recall that in the standard normal distribution, 95% of the observations fall between +1.96 and −1.96 standard deviations of the mean. This means that 5% of the observations fall outside ±1.96 standard deviations. The remaining 5% of the observations are divided between the two tails of the standard normal distribution—2.5% in the left tail and 2.5% in the right tail. The critical *z* value for a one-tailed test is ±1.64. In a one-tailed test, we are interested in whether one population mean is greater than or less than the other population mean. When alpha equals 0.05, this is interpreted as 95% of observations falling above or below 1.64, depending on the direction of the test. The remaining 5% of the observations are located in either the right tail or the left tail of the standard normal distribution. The calculated value of *z*, −3.27, must equal or exceed the critical value of *z* if we are to reject the null hypothesis.

The Excel output provides two *p* values, one for a one-tailed test and one for a nondirectional test, or two-tailed test. Our alternative hypothesis states that we are interested in whether the mean LOSs for males and females for DRG XXX are significantly different. We will use the *p* value for the two-tailed test. The *p* value is 0.0011, which is less than the preset alpha level of 0.05. The interpretation of the *p* value is that the probability of obtaining a *z* statistic as extreme as −3.27, given the corresponding sample size, is 0.11%, or less than 1.0%. Since the *p* value is less than the previously stated alpha level, we reject the null hypothesis and conclude that the mean LOSs for males and females are significantly different from each other. We will discuss one- and two-tailed tests in more detail in Chapter 6.

Remember that the *p* value is not the level of significance; the level of significance, or alpha level, is set prior to conducting the statistical test. The *p* value is obtained as a *result* of the statistical test. The *p* value is the probability of obtaining the resultant test statistic when

all possible samples are drawn. The *p* value is a statistic that indicates how rare the particular sample is, whereas the level of significance is an independent criterion for evaluating the sample result and is in no way dependent upon that particular result.

When our statistical analysis results in a nonsignificant difference, we should evaluate the sample size. When the sample size is small, sampling error is likely to be large, and this often leads to a nonsignificant test result even when the observed difference is caused by a real effect. There is no way to determine whether a nonsignificant difference is the result of the small sample size or whether the null hypothesis is correct. It is for this reason that when our statistical test is not significant, we should almost always regard it as inconclusive rather than as an indication of no effect. On the other hand, very large samples are very likely to result in statistical significance. With large samples, the alpha level is set at 0.01, which requires very strong evidence to reject the null hypothesis. Even with an alpha as strict as 0.01, one must judge the practical implications of the findings. Yes, the test may result in statistical significance, but does this difference have any practical application? When working with data, one cannot rely solely on the results of statistical tests; the knowledge and judgment of the researcher play a vital role in the interpretation of statistical procedures. We will discuss sample size in relation to type I and type II errors in greater detail in a later chapter.

Calculating Sample Size

Procedures for calculating the sample size necessarily vary depending on the type of statistical test to be used and the type of research study to be conducted. In general, the size of the sample (n) is based on

1. the size of the population from which the sample is to be drawn.
2. the desired alpha level, which controls for type I error.
3. the choice on the bounds on the error of the estimate; i.e., how close the sample estimate is to be to the true population value.

Sample size procedures may control for type I error or for both type I and type II errors. The procedures that we will discuss control only for type I error.

Sample sizes will be large when the underlying population is large, when variation within the population is great, when the designated alpha level is strict ($\alpha = 0.01$), and when the interval on the error of the sample estimate is narrow. Various procedures for calculating sample size are presented in Exhibits 5–9 through 5–14.

Exhibit 5–9 Sample Size Calculation to Estimate Proportion or Rate When Population *N* is Known

Problem 1: Calculation of size of a simple random sample to estimate a proportion or rate when the size of the total population is known.

$$\text{Sample size} = n/[n(B^2) + 1]$$

Where B is the bound on the error of the estimate.

Example: There are 600 students in the School of Allied Medical Professions at XYZ University. We want to know what proportion of the student body is male within ±0.5% of the true population proportion. If we want the population proportion to be within $^1/_2$ of 1.0% of the true population proportion, the bound on the error of the estimate is 1.0%(0.5% × 2).

$$\begin{aligned}\text{Sample size} &= n/[n(B^2) + 1]\\ &= 600/[600(0.01^2) + 1]\\ &= 600/1.06\\ &= 566.04, \text{ or } 566 \text{ cases}\end{aligned}$$

Why do we need a sample size of 566 cases, which would constitute 94% of the population? Because we based our calculations on a very narrow bound on the estimate: ±0.5%. We are not willing to tolerate much error when making inferences regarding the true population proportion from the sample mean. If we increase the bound on the estimate to ±2.5%, we have:

$$\begin{aligned}\text{Sample size} &= n/[n(B^2) + 1]\\ &= 600/[600(0.05^2) + 1]\\ &= 600/2.5\\ &= 240 \text{ cases}\end{aligned}$$

Increasing the interval around our sample estimate decreases the sample size required by 58%.

Exhibit 5–10 Sample Size Calculation to Estimate Proportion or Rate When Population *N* is Unknown

Problem 2: Calculation of size of a simple random sample to estimate a proportion or rate when the size of the total population is unknown.

$$\text{Sample size} = [(1.96^2)(pq)]/B^2$$

Where p is the estimated proportion or rate, q is the 1 − estimated proportion or rate, and B is the bound on the error of the estimate (1.96 varies depending on the desired confidence level sought in estimating the proportion or rate).

Example: We want to conduct a study of members of the American Health Information Management Association (AHIMA). We know that in AHIMA, 10% of the population is male. What is the size of the sample needed if we want to be sure we have enough men in our sample within ±1.0% of the true population proportion?

$$\begin{aligned}\text{Sample size} &= [(1.96^2)(pq)]/B^2\\ &= \{(1.96^2)[(0.1)(0.9)]\}/0.02^2\\ &= [3.8416(0.09)]/0.0004\\ &= 864.36, \text{ or } 864 \text{ cases}\end{aligned}$$

Again, we ask why we need a such a large sample size. We need it primarily because we placed such a narrow bound on our estimate. If we increase the bound on the estimate to ±2.5%, the resultant sample size reduces to 138.

$$\begin{aligned}\text{Sample size} &= [(1.96^2)(pq)]/B^2\\ &= \{(1.96^2)[(0.1)(0.9)]\}/0.05^2\\ &= [3.8416(0.09)]/0.0025\\ &= 138.3, \text{ or } 138 \text{ cases}\end{aligned}$$

Increasing the interval around our sample estimate decreases the sample size required by 84%. By increasing the bound on the estimate, we also risk not including enough men in our study sample since they comprise only 10% of the population. We must sometimes oversample to include enough cases when members of a stratum are underrepresented in a population.

Exhibit 5–11 Sample Size Calculation to Estimate Mean When Population *N* is Known

Problem 3: Calculation of size of a simple random sample to estimate the mean when the size of the total population is known.

$$\text{Sample size} = \frac{n(\text{SD})^2}{n - 1[\text{B}^2/1.96^2] + (\text{SD})^2}$$

where SD equals the standard deviation, which can be estimated by dividing the range by 4 (R/4).

Example: We want to select a sample to study the mean number of days in the surgical intensive care unit (SICU) for patients have coronary artery bypass graft (CABG) surgery. Last year, 400 patients were discharged; for these patients, the number of SICU days ranged from 2 to 9. We want to estimate the mean number of days with ±0.25 days of the true population mean.

$$\text{SD} = \text{R}/4 = 9 - 2 = 7/4 = 1.75$$

$$\text{Sample size} = \frac{n(\text{SD})^2}{n - 1[\text{B}^2/1.96^2] + (\text{SD})^2}$$

$$= \frac{400(1.75)^2}{399[(0.5^2)/1.96^2] + 1.75^2}$$

$$= 1{,}225/\{[399(0.065)] + 3.0625\}$$

$$= 1{,}225/28.9975$$

$$= 42.25, \text{ or } 42 \text{ cases}$$

Exhibit 5–12 Sample Size Calculation to Estimate Mean When Population *N* is Unknown

Problem 4: Calculation of size of a simple random sample to estimate the mean when the size of the total population is unknown.

$$\text{Sample size} = [1.96^2(\text{SD})^2]/\text{B}^2$$

Where SD is the standard deviation, which can be estimated by dividing the range by 4 (R/4).
Example: We want to select a sample to study the mean number of surgical intensive care unit (SICU) days for patients having coronary artery bypass graft (CABG) surgery. The CABG benchmarking literature indicates that the number of SICU days ranges from 2 days to 11 days. We want to estimate the mean number of days within ± 0.25 days of the true population mean.

$$\text{SD} = R/4 = 11 - 2 = 9/4 = 2.25$$

$$\text{Sample size} = [1.96^2(\text{SD})^2]/\text{B}^2$$

$$= [1.96^2(2.25)^2]/0.5^2$$

$$= 19.4481/.25$$

$$= 77.79, \text{ or } 78 \text{ cases}$$

Exhibit 5–13 Stratified Sample Size Calculation to Estimate Mean When Population *N* is Known and *n* in Each Stratum is Known

Problem 5: Calculation of size of stratified random sample to estimate the mean when the size of the total population is known and the *n* in each stratum is known.

$$\text{Sample size} = q/\{[(N)(B^2)/1.96^2] + (q/n)\}$$

Where $q = \Sigma[(n)(\text{SD})]^2$ for each stratum, *n* is the population size of each stratum, *N* is the total population size, SD is the standard deviation, and *B* is the bound on the error of estimate.

Example: We want to select a stratified random sample to study the mean drug charges for congestive heart failure (CHF) patients treated during the previous year. The population consists of 370 patients stratified by three physicians. We want the mean charge estimate to be within ± $25 of the true population mean. Information related to physicians A, B, and C appears below:

	Phys A	*Phys B*	*Phys C*
No. of patients	120	160	90
p of patients	0.324	0.433	0.243
Range of drug charges	$1,000	$900	$852
SD (*R*/4)	$250	$225	$213

First calculate *q*:

$$= \Sigma[(n)(\text{SD})]^2$$
$$= [(120)(2500]^2 = 900{,}000{,}000$$
$$= [(160)(225)]^2 = 1{,}296{,}000{,}000$$
$$= [(90)(213)]^2 = 367{,}488{,}900$$
$$2{,}563{,}488{,}900$$

$$\text{Sample size} = q/\{[(N)(B^2)/1.96^2] + (q/n)\}$$
$$= 2{,}563{,}488{,}900/\{[(370)(1.96^2)] + (2{,}563{,}488{,}900/370)\}$$
$$= 370 \text{ total cases (rounded)}$$

For each stratum:

Physician A = 370 × 0.324 = 120 cases
Physician B = 370 × 0.433 = 160 cases
Physician C = 370 × 0.243 = 90 cases

Exhibit 5–14 Sample Size Calculation to Estimate Difference Between Two Means

Problem 6: Calculation of size of a simple random sample to estimate the difference between two means at the 95% confidence level.

Sample size = $\{1.96^2[(SD_1)^2 + (SD_2)^2]\}/B^2$

Where SD_1 = standard deviation of group 1, SD_2 = standard deviation of group 2, and B^2 = size of the difference to be detected.

Example: We want to select a sample to determine if the difference between the mean average length of stay (ALOS) between two physicians is at least ± 0.25 days. The standard deviation for ALOS for the patients of physician A is 1.98; for physician B, it is 1.52.

Sample size = $\{1.96^2[(SD_1)^2 + (SD_2)^2]\}/B^2$
$=\{1.96^2[(1.98^2 + 1.52^2)]\}/0.5^2$
$=23.94/.25$
$= 95.74$, or 96 cases per group

USING COMPUTER SOFTWARE TO SOLVE PROBLEMS

Many computer programs are available that can assist you in solving statistical problems. The advantage of using electronic spreadsheets or a dedicated statistical package is that the data can be entered directly, and databases can be designed to help evaluate these problems on a timelier basis. If you recall from Chapter 3, the more timely the data, the more valuable they are in the decision-making process. The health information manager should become proficient not only in data collection, but also in the analysis of data so that the data are useful to health care providers, planners, and researchers. Examples have been limited to the use of SPSS, but electronic spreadsheets such as Excel may also be used.

Selection of statistical computer software is a matter of choice. An example of using Excel for descriptive statistics is presented in Exhibit 5–15. The output displayed is provided automatically be Excel when one is using the "descriptive statistics" option. Dedicated statistical packages, such as SPSS, are often easier to use in that less manipulation of the data is required. Also, dedicated statistical packages can offer the user more choices in the type of statistical tests that are available, and most include both parametric and nonparametric procedures. Excel does not include options for nonparametric procedures.

Exhibit 5–15 Excel Output for Descriptive Statistics on Age

	Age
Mean	77.733
Standard Error	1.086
Median	78.5
Mode	74
Standard Deviation	5.948
Sample Variance	35.375
Kurtosis	−0.436
Skewness	−0.100
Range	24
Minimum	66
Maximum	90
Sum	2332
Count	30
Confidence Level (95.0%)	2.2209

CONCLUSION

The normal and standard normal distributions are theoretical distributions used for testing statistical problems in health care, since many naturally occurring phenomena follow the normal distribution. The normal distribution is actually a family of distributions in which the population mean can take on any value. The normal distribution is a symmetrical, bell-shaped distribution where 50% of the observations fall above the mean and 50% of the observations fall below the mean. In the normal distribution, the mean, median, and mode are equal.

To make comparisons between distributions, the normal distribution may be standardized. The standard normal distribution has many of the same properties as the normal distribution, except that in the standard normal distribution, the mean is equal to 0 and the standard deviation is 1. There is only one standard normal distribution. In the standard normal distribution, 68% of the observations fall between +1.0 and −1.0 standard deviations of the mean, and 95% of the observations fall between +1.96 and −1.96 standard deviations of the mean.

The standard normal distribution is important in statistical inference. The standard normal distribution can be used to make comparisons between populations. In statistical testing, we are interested in whether one population is the same as or differs from another population on a variable of interest.

When making comparisons between populations, we must first draw a sample from the population. Probability sampling is the preferred sampling technique. Types of probability sampling include simple random sampling, stratified random sampling, systematic sam-

pling, and cluster sampling. It is important to take care when drawing samples so that we have some level of confidence when making inferences from the sample to the underlying population. In general, the larger the sample size, the more confidence we have in our results.

With statistical inference, we are interested in making generalizations about our population parameters from sample statistics. For example, if we are interested in comparing two population means for significant differences, we must set up a hypothesis. For statistical testing we set up two hypotheses—the null and alternative hypotheses. The null hypothesis states that there is no difference between the population parameters of interest; the alternative hypothesis states that there is a statistically significant difference in the population parameters of interest. After formulating our hypotheses, we must set an alpha level for rejection of the null hypothesis. For large samples, alpha is usually set at 0.01 because it is easy to achieve statistical significance with large samples. Conversely, for small samples, alpha is usually set at 0.05 because it is more difficult to achieve statistical significance with smaller samples. The alpha level also indicates the probability of making a type I error. In a type I error, we reject the null hypothesis when it is true. A type II error is made when we fail to reject a null hypothesis that is false.

ADDITIONAL RESOURCES

Jekel, J., et al. 1996. *Epidemiology, biostatistics, and preventive medicine*. Philadelphia: W.B. Saunders.

Schott, S. 1990. *Statistics for health professionals*. Philadelphia: W.B. Saunders.

Stockburger, D.W. 1998. Introductory statistics: Concepts, models and applications. The normal curve. http:// www.psychstat.umsu.edu/introbook/sbk13m.html.

U.S. Department of Health and Human Services, Public Health Service. 1992. *Principles of epidemiology: An introduction to applied epidemiology and biostatistics*. Atlanta, GA: USDHHS.

Appendix 5–A

Exercises for Solving Problems

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
2. Compare the normal distribution with the standard normal distribution.
3. What is the difference between the standard deviation and the standard normal deviate?
4. You have been analyzing discharges by sex from DRG 462, Rehabilitation, for Critical Care Hospital. Specifically, you are interested in determining if there is a difference in average age by sex. Review the data in Table 5–A–1, and answer the questions that follow.

Table 5–A–1 Mean Age of Patients by Sex, DRG 462, Rehabilitation, Critical Care Hospital, 2004

	Female	*Male*	*Total*
N	80	88	168
Mean	55.467	53.947	54.671
Standard Deviation	17.5344	19.1981	18.3856
Standard Error	1.9604	2.0465	1.4185
Lower bound 95% CI	51.565	49.879	51.871
Upper bound 95% CI	59.370	58.015	57.472
Minimum age	19.4	17.5	17.5
Maximum age	89.1	97.6	97.6

   a. State the null and alternative hypotheses; state the a priori alpha level.
   b. What is the average age for the entire group? For men? For women?
   c. Why is the standard error larger for the male group than for the female group?
   d. Why is the standard error for the total group smaller than either of the standard errors for the male or female groups?
   e. What is the 95% confidence interval for men? For women? For the entire group?
   f. What is your interpretation of the 95% confidence interval?

5. You have been analyzing hospital discharges from DRG 15, Transient Ischemic Attack and Precerebral Occlusions. The average length of stay (ALOS) for patients discharged from DRG 15 is 2.2 days. The national length of stay for DRG 15 is 4.1 days. You are interested in determining whether the hospital's length of stay for DRG 15 is significantly different from the national ALOS.
   a. State the null and alternative hypotheses.
   b. Set the alpha level.
6. You are interested in studying patient average length of stay (ALOS) for three physicians. You want to determine if there is a significant difference between the ALOSs of the three physicians.
   a. State the null and alternative hypotheses.
   b. Set the alpha level.
7. Explain the differences between the alpha level and the p value.
8. What are type I and type II errors? What factors contribute to making either a type I or type II error?
9. The mean length of stay for patients discharged from DRG 005, Extracranial Vascular Procedures, is 3.33. The standard deviation for the group is 3.18, and the number of patients discharged is 21. Calculate the 95% confidence interval for the mean length of stay.

MULTIPLE CHOICE

1. In a normal distribution, 68% of the observations fall within:
   a. $\pm 1\ \sigma$ of the mean
   b. $\pm 2\ \sigma$ of the mean
   c. $\pm 3\ \sigma$ of the mean
   d. $\pm 1\frac{1}{2}\ \sigma$ of the mean
2. In a normal distribution, 32% of the scores fall outside:
   a. $\pm 1\ \sigma$ of the mean
   b. $\pm 2\ \sigma$ of the mean
   c. $\pm 3\ \sigma$ of the mean
   d. $\pm 1\frac{1}{2}\ \sigma$ of the mean
3. The normal distribution is:
   a. continuous
   b. a family of distributions
   c. symmetrical about the mean
   d. all of the above

4. Which of the following is not a characteristic of the normal curve?
   a. It is unimodal.
   b. It is a discrete distribution.
   c. It is asymptotic to the *x*-axis.
   d. The mean may take on any value.

5. The lengths of stay for DRG 123 were standardized so that comparison could be made across hospitals. What percentage of the lengths of stay would have a *z* value greater than or equal to 1.00?
   a. 32%
   b. 16%
   c. 8%
   d. 4%

6. In a standard normal distribution, what percentage of the lengths of stay for DRG 123 would fall above $z = -1.96$?
   a. 99%
   b. 97.5%
   c. 95%
   d. 68%

7. If a distribution has a long tail to the right, it is:
   a. bimodal
   b. abnormal
   c. positively skewed
   d. negatively skewed

8. In a standard normal distribution, the mean:
   a. is 0.00
   b. is -1.00
   c. is $+1.00$
   d. may take on any value

9. A normal distribution has a mean of 20 and a standard deviation of 5. Ninety-five percent of the scores fall between:
   a. 15 and 25
   b. 10 and 30
   c. 5 and 35
   d. not enough information provided

10. The standard deviation of a distribution is 24; the sample size is 9. The standard error of the mean is:
    a. 3
    b. 8
    c. 75
    d. 225
    e. not enough information provided

11. A condition that is fundamental to statistical inference is:
    a. random sampling
    b. that the population is normally distributed
    c. that the mean of the population is known
    d. (a) and (b)
    e. all of the above

12. The null hypothesis is a statement that is:
    a. probably true
    b. considered to be false until proven true
    c. evaluated statistically as either true or false
    d. all of the above

13. When $\alpha = 0.05$, the null hypothesis will be:
    a. rejected 5% of the time
    b. rejected 5% of the time when it is true
    c. accepted 5% of the time
    d. accepted 5% of the time when it is false

14. A type I error occurs when:
    a. we reject the null hypothesis when it is true
    b. we reject the null hypothesis when it is false
    c. we accept the null hypothesis when it is true
    d. we accept the null hypothesis when it is false

15. A type II error occurs when we:
    a. use a two-tailed test when a one-tailed test is more appropriate
    b. use a one-tailed test when a two-tailed test is more appropriate
    c. reject the null hypothesis when it is true
    d. accept the null hypothesis when it is false

16. If we change the alpha level for a statistical test from 0.05 to 0.01, we are:
    a. increasing the risk of making a type I error
    b. decreasing the risk of making a type I error
    c. decreasing the risk of making a type II error
    d. increasing the probability of finding statistical significance

17. In general, large sample sizes:
    a. reduce the risk of type I error
    b. reduce the risk of type II error
    c. make it easier to achieve statistical significance
    d. (a) and (c)
    e. all of the above

18. The *p* value is the:
    a. power of a statistical test
    b. probability that the null hypothesis is true
    c. probability of making a type II error
    d. probability of getting a result as extreme as the one observed if the null hypothesis is true

19. We have drawn a simple random sample of 100 patients who were discharged from Critical Care Hospital in January. Of all the patients discharged during January, 55% were women and 45% were men. To match the population, our sample should contain:
    a. 45 men and 55 women
    b. 50 men and 50 women
    c. 55 men and 45 women
    d. The ratio of men to women in the sample size is not important.

20. You are assisting a physician who is conducting a study on the number of cancer cases at Critical Care Hospital. You jointly decide to take a 5% random sample of the estimated 20,000 charts. This is an example of:
    a. cluster sampling
    b. simple random sampling
    c. stratified random sampling
    d. two-stage random sampling

21. The physician now decides to draw his 5% sample of cancer cases by selecting every 20th chart by medical record number. This is an example of:
    a. random cluster sampling
    b. two-stage random sampling
    c. systematic sampling
    d. stratified random sampling

22. You decide to study coding quality by randomly selecting hospitals in your state. From each of the hospitals selected, you review coded charts of randomly selected coders. This is an example of:
    a. cluster sampling
    b. simple random sampling
    c. stratified random sampling
    d. two-stage random sampling

PROBLEMS

1. Review the data on length of stay that appear in Table 5–A–2 to answer the questions below.

a. using a microcomputer statistical package, calculate the average length of stay for the entire group, for men, and for women.
b. You are interested in determining if there is a difference in the average length of stay by sex. State the null and alternative hypotheses and state the a priori alpha level.
c. Calculate the standard error of the mean length of stay for the entire group, for men, and for women.
d. Calculate the 95% confidence interval for length of stay for the entire group, for men, and for women.

Table 5–A–2 Critical Care Hospital, Length of Stay of Patients by Sex, DRG 127, Heart Failure and Shock

LOS Days	*Female*	*Male*	*Total*
1	1	5	6
2	1	5	6
3	3	13	16
4	4	1	5
5	0	5	5
6	3	1	4
7	0	1	1
8	2	2	4
10	1	3	4
11	1	2	3
13	1	0	1
14	0	1	1
15	1	0	1
16	0	1	1
17	0	1	1
27	1	0	1
36	1	0	1
Total	20	41	61

CHAPTER 6

Hypothesis Testing of the Difference Between Two Population Means

KEY TERMS

Hypothesis testing
One-tailed test
Two-tailed test
Noncritical region
Region of rejection
Critical region
z test
  z test for comparing two independent population means
  z test for comparing two population proportions
Effect
t test
  one-sample *t* test
  t test for comparison of two independent sample means
  paired *t* test
Degrees of freedom (df)

LEARNING OBJECTIVES

At the conclusion of this chapter, you should be able to:

1. Define key terms.
2. Calculate one- and two-tailed tests of significance for one and two independent samples using *z*.
3. Use *z* for comparing two population proportions.
4. Compare and contrast the normal distribution and *t* distribution.
5. Calculate one- and two-tailed tests of significance for one and two independent samples using *t*.
6. Conduct a paired-sample *t* test.
7. Use statistical software to calculate the various *t* tests.

In our discussion of **hypothesis testing** of the difference between population means, we will be dealing with examples where population parameters are both known and unknown. If population parameters are known and we have sufficiently large samples ($N \geq 30$), the standard normal distribution is used as the basis for statistical decision making in the form of a *z* test. If we are comparing means when population parameters are unknown, and if our samples are smaller, Student's *t* test is used as the basis for statistical decision making. Our discussion will first focus on the use of the *z* test and the standard normal distribution, and then consider the *t* test and the *t* distribution for comparing sample means to population means, comparing the means of two populations, and comparing pre- and posttest scores of matched pairs.

THE STANDARD NORMAL DISTRIBUTION AND THE *z* TEST FOR COMPARING POPULATION MEANS

Before we conduct a statistical test such as the *z* test, we must develop a hypothesis for the test. We select a statistical test based on our research questions, which are developed into statistical hypotheses. A statistical hypothesis may involve comparing a sample mean to a population mean or comparing two or more sample means drawn from two populations. The hypothesis may be either one-tailed (directional) or two-tailed (nondirectional). In a **one-tailed test**, we are seeking to determine if our sample mean, $\overline{X}$, is significantly greater or less than the population parameter μ. If we are interested in determining whether our sample $\overline{X}$ is significantly greater than the population parameter μ, we look for a critical *z* value in the positive tail of the distribution. Conversely, if we are interested in determining whether our sample $\overline{X}$ is significantly less than the population parameter μ, we look for a critical *z* value in the negative tail of the distribution. In a one-tailed test with an alpha of 0.05, we look for statistical significance in either the upper or lower tail of the distribution.

In a nondirectional or **two-tailed test**, we are interested only in determining whether the sample mean, $\overline{X}$, and the population parameter, μ, are significantly different from each other; the direction of the inequality is not an issue. In this case, both tails of the *z* distribution are used in the statistical decision making. A two-tailed test divides alpha in half, placing 0.025 in each tail. That is, when alpha is set at 0.05, we look for statistical significance in either the positive or negative tails of the standard normal distribution.

The null and alternative hypotheses may take the following forms:

$$H_{0A}\mu_1 = 5.5 \text{ or } H_{0A}: \mu_1 = \mu_2$$
$$H_A\mu_1 \neq 5.5 \text{ or } H_A: \mu_1 \neq \mu_2 \text{ (two-tailed tests)}$$
$$H_A\mu_1 < 5.5 \text{ or } H_A: \mu_1 < \mu_2 \text{ (one-tailed tests)}$$
$$H_A\mu_1 > 5.5 \text{ or } H_A: \mu_1 > \mu_2 \text{ (one-tailed tests)}$$

From our discussion in Chapter 5, we know from the central limit theorem that we can expect that 95% of the sampling means will fall between +1.96 and −1.96 standard deviations of the true mean, and that 99% of the sampling means will fall between +2.58 and

−2.58 standard deviations of the true mean. For a two-tailed test, these standard deviations correspond to the significance levels of 0.05 and 0.01, respectively. In a one-tailed test, $z = \pm 1.64$ when $\alpha = 0.05$, and $z = \pm 2.33$ when $\alpha = 0.01$. A one-tailed test is considered more "robust" than a two-tailed test because statistical significance is easier to achieve. In a one-tailed test when $\alpha = 0.05$, ± 1.65 is required to achieve statistical significance; in a two-tailed test, ± 1.96 is required for statistical significance. The space between the critical values of *z* is called the **noncritical region**; the space outside or equal to the critical values of *z* is called the **region of rejection**. The critical values for the standard normal distribution are displayed in Table 6–1.

Table 6–1 Critical Values for Standard Normal Distribution

	Alpha	*Critical Value z*
Two-tailed test	0.05	± 1.95
	0.01	± 2.58
One-tailed test	0.05	± 1.65
	0.01	± 2.33

A one-tailed test is considered more "robust" than a two-tailed test because it is easier to achieve statistical significance. When conducting a statistical test, we "calculate" the value of *z* and compare it to the critical value of *z*. On a one-tailed test when $\alpha = 0.05$, a *z* of ± 1.65 is required to achieve statistical significance, so our calculated value of *z* must fall outside ± 1.65 in order to be statistically significant. In a two-tailed test, a *z* of ± 1.96 is required for statistical significance, so our calculated value of *z* must fall outside ± 1.96 in order to be statistically significant. In a two-tailed test, the space between ± 1.96 when $\alpha = 0.05$ is called the **noncritical region** or the area where we do not reject the null hypothesis. Thus, if the calculated value of *z* falls between ± 1.96, we fail to reject the null hypothesis. On the other hand, the space outside or beyond $\pm$ 1.96 when $\alpha = 0.05$ is called the **critical region**. If the calculated value of *z* is greater than or equal to $+1.96$ or less than or equal to -1.96, we reject the null hypothesis.

When we are conducting either a one-tailed or a two-tailed *z* test, the population parameters μ and σ are known. In our first example, we will conduct a two-tailed test in which we are interested in whether the hospital's average length of stay (ALOS) for patients discharged from hypothetical DRG XXX is different from the national ALOS. The 2003 national mean length of stay (LOS) (μ) for DRG XXX is 6.3 days, and the hypothetical standard deviation is 2.44 (σ). (We are using a hypothetical standard deviation because the actual population parameter is not available.) At Critical Care Hospital in 2003, 104 patients were discharged from DRG XXX with an ALOS of 5.49 days and a standard deviation equal to 3.44. In determining whether the observed difference in the ALOSs is statistically significant, we need to first state the null and alternative hypotheses and alpha level:

$$H_0: \mu_1 = \mu_2$$
$$H_A: \mu_1 \neq \mu_2$$
$$\alpha = 0.05$$

The null hypothesis states that the hospital's ALOS is equal to the national ALOS for DRG XXX. The alternative hypothesis states that the hospital ALOS for DRG XXX is significantly different from the national ALOS. Since the alternative hypothesis is one of inequality, we are conducting a two-tailed test. The calculated value of z must fall outside or be equal to the critical value of z, ± 1.96. The z value is calculated as follows:

$$\begin{aligned} z &= \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \\ &= \frac{5.49 - 6.30}{2.44/\sqrt{104}} \\ &= \frac{-0.81}{(2.44/10.19)} \\ &= -3.39 \end{aligned}$$

The results would be reported as:

$$Z_{calc} = -3.39$$
$$Z_{crit} = -1.96$$

Since our calculated value is negative, we refer to the negative tail of the distribution. The calculated value of z, -3.39, falls outside the critical value of z, -1.96, when $\alpha = 0.05$. We therefore reject the null and conclude that the ALOS at Critical Care Hospital is significantly different from the national ALOS for DRG XXX.

In our second example, we will conduct a one-tailed test. The quality improvement team at Critical Care Hospital is interested in determining if the administration of a new antibiotic has had any effect in reducing the ALOS for patients discharged from DRG XXX. (An **effect** is a change in one variable—in this case, the ALOS—that may be associated with another variable—in this case, the antibiotic.) In 2003, the ALOS for the 104 patients discharged from DRG XXX at Critical Care Hospital was 5.49 days, and the standard deviation was 3.44. Specifically, we want to know if the antibiotic resulted in significant reduction of the ALOS.

We will first construct our null and alternative hypotheses. In the null hypothesis, we are assuming that the antibiotic will have no effect. If we reject the null hypothesis on the basis of our statistical test, we are saying that the probability of getting such a mean value by random chance is too small for us to believe that the null hypothesis is true. If we fail to reject the null hypothesis, we are stating that the probability associated with the observed mean is large, that such samples are common, and that there is not enough evidence that the antibiotic was effective in reducing the ALOS.

The null and alternative hypotheses are

$$H_0: \mu_1 = 5.49$$
$$H_A: \mu_1 < 5.49$$
$$\alpha = 0.05$$

The null hypothesis states that the hospital ALOS is equal to 5.49, the hospital's ALOS for DRG XXX in 2003. The alternative hypothesis states that the hospital ALOS is significantly less than the 5.49 days.

In this one-tailed test, we are interested in determining if the ALOS for DRG XXX is significantly less than the ALOS prior to administration of the new antibiotic. So we will look for statistical significance in the negative end of the tail. For a one-tailed test in which alpha is set at 0.05, the critical value of z is -1.64, so the calculated value of z must be less than or equal to -1.64 in order for the results to be statistically significant.

Now that we have stated the null and alternative hypotheses, we need to determine the sample size necessary to provide us with results in which we can be confident. Using the formula in the previous chapter in Exhibit 5–11, we will calculate the sample size, with a bound on the estimate equal to 0.25. The bound on the estimate indicates that we want to be within ± 0.25 of our true population mean. The sample size is calculated as follows:

$$\frac{N(\mathrm{SD}^2)}{[N - 1(B^2/1.96^2)] + \mathrm{SD}^2}$$
$$= \frac{104(3.44^2)}{[103(0.5^2/1.96^2)] + 3.44^2}$$
$$= 66.39$$
$$= 66$$

The sample size for our study is 66; the size of the standard deviation indicates that there is quite a bit of variation in the LOS for the 2003 discharges. Thus, for the amount of precision required in the study, a comparatively large sample is required. The descriptive statistics relating to our sample appear in Exhibit 6–1. The sample mean for the 66 patients who make up the sample who were treated after implementation of the antibiotic is 4.71.

Exhibit 6–1 Descriptive Statistics for Sample from DRG 089

		LOS	*Valid N (listwise)*
N	Statistic	66	66
Minimum	Statistic	1.00	
Maximum	Statistic	11.00	
Mean	Statistic	4.7121	
Std. Deviation	Statistic	2.6181	
Skewness	Statistic	.601	
	Std. Error	.295	
Kurtosis	Statistic	−.608	
	Std. Error	.582	

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$
$$= \frac{4.721 - 5.49}{2.44/\sqrt{66}}$$
$$= \frac{-0.78}{(2.44/8.12)}$$
$$= -2.59$$

Our calculated value of z is -2.59 and falls in the region of rejection, since the critical value of z for a one-tailed test is -1.64 when alpha equals 0.05. Because the calculated value is less than or equal to the critical value of z, we reject the null and conclude that the new antibiotic may have had an effect on reducing LOS for patients discharged from DRG XXX in 2004. Note that this example is used for illustrative purposes only. Other factors may be at work that account for the reduction in the ALOS.

We can find the critical value of z by referring to a statistical table. Appendix B contains a table for the z scores in the standard normal distribution (Table B-1). The table contains three columns:

Z	*Cum p*	*Tail p*
0.00	0.5000	0.5000
0.01	0.5040	0.4960
..	..	..
..	..	..
3.90	1.000	0.0000

The first column contains the values of z from 0.00 to 3.90; these should be read as both positive and negative, since the normal distribution is symmetrical. As you recall, the z score, 0.00, is the mean or center of the distribution. The second column is the cumulative probability of z from the lower end to the distribution to the location of z in the standard normal distribution. The third column is the tail probability, or the area beyond the location of z in the remaining portion of the distribution. For each value of z, the sum of these two probabilities is equal to 1.00.To locate the exact probability of a calculated value of z, such as -3.39, locate 3.39 in the z column. The third column indicates that the exact p value for a one-tailed test is 0.0003. This is the portion of the standard normal distribution that is in one tail of the distribution beyond the z value $+-3.39$. If we are conducting a two-tailed test, we must double the p value to 0.0006.

THE *z* TEST FOR COMPARING TWO POPULATION PROPORTIONS

Often in quality improvement activities or various types of medical research, we are interested in comparing population proportions rather than population means. For example, we

may be interested in comparing the proportion of patients in medical and surgical intensive care who acquire nosocomial infections, or comparing the proportion of patients who survive five years after experiencing two different surgical procedures. We can use a variation of *z* for comparing two population means or comparing two population proportions. The *z* test for comparing two population proportions may be either directional or nondirectional. Just as when we compare two population means using *z*, we assume that the two samples are selected independently and randomly from their respective populations and that the samples are normally distributed.

The null and alternative hypotheses may take the following forms:

$$H_0: p_1 = p_2$$
$$H_A: p_1 \neq p_2 \text{ (two-tailed tests)}$$
$$H_A: p_1 < p_2 \text{ or } p_1 > p_2 \text{ (one-tailed tests)}$$

The formula for calculating *z* for comparing two proportions is

$$z = \frac{(p_1 - p_2) - [1/2(1/n_1 + 1/n_2)]}{\sqrt{pq[(1/n_1) + (1/n_2)]}}$$

where p is the proportion of p_1 and p_2 when considered together as one sample and q is $(1 - p)$.

Consider the hypothetical example where we are comparing five-year survival rates following surgery for breast cancer. In group 1, 100 patients were followed after undergoing lumpectomy. In group 2, 100 patients were followed after undergoing mastectomy of the affected breast. In group 1, 80% of the patients were still alive after five years; in group 2, 85% were still alive after five years. The research question is whether there is a significant difference in the five-year survival rates between the two groups. The null and alternative hypotheses for a nondirectional test are

$$H_0: p_1 = p_2$$
$$H_A: p_1 \neq p_2$$
$$\alpha = 0.05$$

The null hypothesis states that the proportion of patients who survived in group 1 is equal to the proportion who survived in group 2. The alternative hypothesis states that the proportions of patients who survived in each group are not equal. The test we will conduct is a nondirectional test for the difference between two population proportions. In order for the test to be statistically significant, the calculated value of *z* must fall outside or equal the critical value of *z*, ± 1.96.

For a nondirectional test for an alpha of 0.05, the critical value of *z* is ± 1.96, and *z* is calculated as

$$
\begin{aligned}
z &= \frac{(p_1 - p_2) - [1/2(1/n_1 + 1/n_2)]}{\sqrt{pq[(1/n_1) + /n_2)]}} \\
&= \frac{(0.8 - 0.85)[1/2(1/100 + 1.100)]}{\sqrt{(0.825)(0.175)(1/100 + 1/100)}} \\
&= \frac{(-0.05) - [1/2(0.01 + 0.01)]}{\sqrt{0.144(0.01 + 0.01)}} \\
&= -0.06/0.054 \\
&= -1.12
\end{aligned}
$$

Since our calculated value of z, -1.12, does not fall in the region of rejection, we fail to reject the null and conclude that it appears that the five-year survival rates following these two different surgical procedures are not significantly different. In other words, the type of surgical procedure had no effect on the five-year survival rate for breast cancer.

In the above example, the z calculations were performed with the assistance of a hand-held calculator. Neither SPSS, version 12.0, nor Excel provides a procedure for conducting a z test for the difference between two population proportions.

THE *t* TEST

Thus far in this text, we have discussed only the normal and standard normal distributions. There are other distributions from which statistical inferences can be made, one of which is the t distribution, sometimes referred to as Student's t. Student's t is named for William Gosset, who published under the pseudonym of Student. He was the first to describe this family of distributions. We conduct a ***t* test** when the population parameter σ is unknown. The σ is estimated from the sample statistic s. Before conducting the actual t test, we will compare the standard normal distribution with the t distribution. The t distribution is used for statistical testing when population parameters are unknown and/or when the sample size is small. The definition of small varies; some researchers state that a sample size of less than 500 is small, while others consider a sample size of less than 90 small. Others have used the various forms of the t test with sample sizes of less than 30. With very large samples (e.g., $N \geq 1{,}000$), the t distribution and the normal distribution are approximately the same. In fact, the standard normal curve is a special case of the t distribution when $\text{df} = \infty$.

Both the standard normal and t distributions are symmetrical about a mean of zero. Like the normal distribution, the t distribution is actually a family of distributions based on sample size. This additional parameter is referred to as **degrees of freedom** (df) and is calculated by subtracting 1 from the sample size ($\text{df} = N - 1$). Exhibit 6–2 provides an explanation of degrees of freedom. The normal distribution is bell-shaped; the shape of the t distribution is related to the number of degrees of freedom. A distribution with a small

number of degrees of freedom is flatter; this results in a greater area in the tails of the distribution. Because the *t* distribution is so spread out, it is more difficult to achieve statistical significance with a small sample size.

Exhibit 6–2 Degrees of Freedom

The term *degrees of freedom* refers to the number of values that are free to vary after certain restrictions have been placed on the data. For example, in any set of interval level data, we can sum the data to get a total. Assume that for a given frequency distribution, the sum is 100 and there are 10 cases ($N = 10$). If we arbitrarily assign a value of 10 to the first case, 5 to the second case, 15 to the third, 8 to the fourth, 2 to the fifth, 18 to the sixth, 23 to the seventh, 4 to the eighth, and 6 to the ninth, the cumulated sum will be:

$$10 + 5 + 15 + 8 + 2 + 18 + 23 + 4 + 6 = 91$$

The first nine values sum to 91. For all 10 cases to sum to 100, the 10th case must be equal to 9. Nine scores were arbitrarily set before the 10th was determined. Nine scores were free to vary; thus, we have nine degrees of freedom. Degrees of freedom are the number of elements in a set that can be arbitrarily defined before the rest of the elements in the set are determined.

In the *t* distribution when $\alpha = 0.05$, the critical value required to achieve statistical significance varies with the size of the sample. With the *t* distribution for a sample size of 10 ($df = 9$) and for a two-tailed test, a value outside or equal to ± 2.262 must be obtained to achieve statistical significance, as compared to ± 1.96 in the standard normal distribution. In the *t* distribution with nine degrees of freedom, 95% of the observations fall between -2.262 and $+2.262$ standard deviations of the mean. If the size of the sample is increased from 10 to 20 ($df = 19$), the critical value required to achieve statistical significance decreases to ± 2.093. Thus, the critical value needed to achieve statistical significance decreases as the sample size increases. As the sample size increases, the *t* distribution becomes less broad and flat and approaches the bell shape of the normal curve. With a sample size of 500, the critical value for the *t* distribution, when alpha is set at 0.05, is ± 1.965, similar to the critical value of ± 1.96, when alpha is set at 0.05, for the normal distribution. To test a hypothesis using the *t* distribution, we compare the calculated value of *t* to the critical value of that is contained in the *t* table (Appendix B, Table B–2). Remember that the *t* distribution is actually a family of distributions where the tabled value of *t* is dependent upon the number of degrees of freedom in the sample, where $df = N - 1$.

The assumptions for the *t* test are that the samples are drawn randomly and independently from their respective populations and that they are normally distributed. In addition, it is assumed that the population variances of the two groups are approximately equal.

Two-tailed *t* Test

In a nondirectional *t* test, our question is just whether our sample mean is significantly different from the population mean, regardless of direction. In this case, both tails of the *t* distribution are used in the statistical decision making. A two-tailed *t* test divides alpha in half, placing half in each tail. That is, when alpha is set at 0.05, 2.5% of the area under both the upper and lower tails of the curve is considered when deciding whether to accept the null hypothesis.

Given this information about the *t* test and the *t* distribution, let us work through the same LOS problem that we used with the one-tailed *z* test. To restate our question of interest, we are interested in determining whether the hospital's ALOS for DRG XXX is significantly different from the national ALOS for DRG XXX. The national ALOS is 6.3 days; our hospital mean LOS is 5.49, and the standard deviation is 3.44. There were 104 discharges from DRG XXX. The first step in the process is to state our null and alternate hypotheses and to set the alpha level at which we will reject our null hypothesis:

$$H_0: \mu_1 = \mu_2$$
$$H_A: \mu_1 \neq \mu_2$$
$$\alpha = 0.05$$

As with the nondirectional *z* test, the null hypothesis states that the hospital ALOS for DRG XXX is equal to the national ALOS for DRG XXX. The alternative hypothesis states that the ALOS for the hospital is significantly different from the national ALOS for DRG XXX. Since we are interested in determining only whether our hospital mean is different from the national mean, the *t* test is nondirectional, or a two-tailed test. Next, we determine the critical region for rejection of the null hypothesis. For a sample size of 104 (df = 103), the tabled *t* or critical $t(t_{0.05})$ value is approximately ± 1.98; since H_A is nondirectional, the critical region consists of all values of $t \geq 1.984$ or ≤ 1.984.

To locate the critical value of *t*, we refer to the tabled critical values of the *t* distribution in Appendix B, Table B–2. The first column lists the degrees of freedom, and the remaining columns identify the critical values of *t* for *a priori* alpha levels for both one- and two-tailed tests. Since the table does not list all possible degrees of freedom, we select the row for df = 100 for our problem. For a two-tailed test with alpha set at 0.05, the critical value of *t* is 1.984. Therefore, our region of rejection for the calculated value of *t* must equal or fall outside ± 1.984.

Using the formula for a **one-sample *t* test**, we can now calculate *t*:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{N}}$$
$$t = \frac{5.49 - 6.30}{3.44/\sqrt{104}}$$
$$= \frac{-0.81}{3.44/10.19}$$
$$= -2.399$$

The results would be reported as

$$t_{calc} = -2.399$$
$$t_{crit.05} = -1.984$$

Since our t_{calc} falls in the region of rejection, we reject the null and conclude that our hospital LOS is significantly different from the national ALOS for DRG XXX.

Note that the formula for *t* takes the same general form as that for *z*. The only difference is that in *z*, the population parameter σ is used in the calculation, whereas in *t*, the population σ is estimated from the sample statistic *s*. To calculate the CI_{95}, we use the same procedures as presented in Chapter 5, except that we use the critical value of t_{05}, which for 103 df is 1.98.

$$\begin{aligned} CI_{95} &= \bar{X} \pm t_{\alpha}(s/\sqrt{N}) \\ &= 5.49 \pm 1.98(3.44/\sqrt{104}) \\ &= 5.49 \pm 1.98(3.44/10.19) \\ &= 5.49 \pm 1.98(0.34) \\ &= 5.49 \pm 0.668 \\ &[4.82,6.16] \end{aligned}$$

Thus, we are 95% confident that the true population mean lies between 4.82 and 6.16. We can use SPSS to calculate the one-sample *t* test. When requesting the one-sample *t* test, the population parameter (i.e., 6.3) to which the sample mean is being compared must be specified. The output for the one-sample *t* test appears in Exhibit 6–3 and an explanation of the SPSS output appears in Exhibit 6–4.

Exhibit 6–3 SPSS Output for One-Sample *t* Test

One-Sample Statistics

	LOS
N	104
Mean	5.4904
Std. Deviation	3.44159
Std. Error Mean	.33748

One-Sample Test

			LOS
Test Value = 6.3	*t*		−2.399
	df		103
	Sig. (2-tailed)		.018
	Mean Difference		−.80962
	95% Confidence Interval of the Difference	Lower	−1.4789
		Upper	−.1403

Exhibit 6–4 Output for One-Sample *t* Test

Test Value	*Population Parameter to Which the Sample is Compared*
t	The calculated value of *t*.
df	Degrees of freedom; for the one-sample *t* test the degrees of freedom are equal to $(n - 1)$
Sig. two-tailed	For a two-tailed or nondirectional statistical test, the *p* value for the calculated value of *t*
Mean difference	The **actual** difference between the two population means; the hospital mean is subtracted from the national mean
95% confidence interval of the difference	The interval that covers the true difference between the two population means

✓ To Obtain a One-Sample *t* Test Using SPSS:

- From the menus, choose:
  Analyze
  →Compare Means
  →One Sample *t* Test
- Select one or more variables to be tested against the hypothesized value. Enter a numeric text value against which each sample mean is compared.

Note that the calculated *t* for our SPSS output is the same as that calculated with the assistance of a hand-held calculator. (When minor differences occur, they are most likely due to rounding.) The significance level or *p* value is 0.018, which is less than our previously established alpha level of 0.05. SPSS reports results in terms of significance—what we previously described as the *p* value. The critical value of *t* for the predetermined alpha level is not reported. The 95% confidence interval of the difference between the means is also provided. The 95% confidence interval of the mean is interpreted as meaning that we are 95% confident that the interval [−0.1403, −1.4789] covers the true difference in the LOS between the hospital mean and the national mean for DRG XXX. It is calculated as

$$CI_{95} = (\overline{X}_1 - \overline{X}_2) \pm t_\alpha[s\sqrt{(1/n_1) + (1/n_2)}]$$

One-Tailed *t* Test

Using the same information for the one-tailed *z* test, we will conduct a one-tailed *t* test. Recall that the quality improvement team is interested in determining if the administration of a new antibiotic resulted in a decrease in the ALOS. As before, the null and alternative hypotheses are

$$H_0: \mu_1 = 5.49$$
$$H_A: \mu_1 < 5.49$$
$$\alpha = 0.05$$

From the data in Exhibit 6–1, we know that the mean after administration of the antibiotic is 4.7121. Our previous calculations indicate that the required sample size is 66. Because this is a one-tailed test, the critical *t* value for an alpha of 0.05 and for 65 degrees of freedom is approximately −1.664. (We had to approximate the critical *t* value because our *t* table does not include the critical values of *t* for all possible degrees of freedom.) Our table provides the critical values of *t* for 60 and 80 degrees of freedom. So a conservative estimate of the critical *t* for 80 df was selected, −1.664. This is not much different from the critical value of *t* for 60 df, −1.671.

$$t = \frac{\bar{X} - \mu}{s/\sqrt{N}}$$
$$t = \frac{4.71 - 5.49}{2.62/\sqrt{66}}$$
$$= \frac{-0.78}{2.62/8.124}$$
$$= -2.41$$

Since our t_{calc}, −2.41, is less than or equal to −1.664, we reject the null hypothesis and state that it appears that the new antibiotic may have been effective in reducing the hospital ALOS for DRG XXX. The 95% CI for the mean would be calculated as

$$CI_{95} = \bar{X} \pm t_\alpha(s/\sqrt{N})$$
$$= 4.71 \pm 1.67(2.62/\sqrt{66})$$
$$= 5.71 \pm 1.67(0.322)$$
$$= 4.71 \pm 0.54$$
$$[4.17,5.25]$$

We are therefore 95% confident that the interval from 4.17 to 5.25 covers the true LOS for our hospital patients after administration of the antibiotic.

When we use SPSS for the one-tailed *t* test, the output is the same as that for the two-tailed test, as displayed in Exhibit 6–5. The significance of the *t* statistic or *p* value for a two-sided test is reported. If we are conducting a one-tailed test, we must divide the *p* value when reporting the results. For a one-tailed *t* test, the *p* value becomes 0.0095 (0.019/2), illustrating that it is easier to achieve statistical significance with a one-tailed test than with a two-tailed test.

Exhibit 6–5 SPSS Output for One-Sample *t* Test

One-Sample Statistics

	LOS
N	66
Mean	4.7121
Std. Deviation	2.61807
Std. Error Mean	.32226

One-Sample Test

			LOS
Test Value = 5.49	*t*		−2.414
	df		65
	Sig. (2-tailed)		.019
	Mean Difference		−.77788
	95% Confidence Interval of the Difference	Lower	−1.4215
		Upper	−.1343

The *t* Test for Comparing Two Independent Sample Means

Sometimes we are interested in comparing means from two independent samples—for example, comparing average charges by DRG or average charges by physician. In these examples, we would draw independent random samples from their respective populations. The null hypothesis would state that there was no difference in the population means for the two groups; the alternative hypothesis would be that there was a difference between the two population means:

$$H_0: \mu_1 = \mu_2$$
$$H_A: \mu_1 \neq \mu_2$$
$$\alpha = 0.05$$

The formula for the ***t* test for comparison of two independent sample means** is

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_p\sqrt{(1/n_1) + (1/n_2)}}$$

where s_p is called the pooled standard deviation. The pooled standard deviation is an average of the sample variances – n_1 and n_2. The pooled standard deviation is found by

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Degrees of freedom for the *t* test for two independent samples are equal to $n_1 + n_2 - 2$. Exhibit 6–6 outlines the steps for solving the two-sample *t* test. The null hypothesis states that the mean charges for Dr. Sparenocost and Dr. Spendtheleast are equal; the alternative hypothesis states that the mean charges for Drs. Sparenocost and Spendtheleast are significantly different. We are conducting a one-tailed, nondirectional *t* test. The SPSS output for a two-sample *t* test appears in Exhibit 6–7.

Exhibit 6–6 Calculation of *t* Test for Two Independent Sample Means

At Critical Care Hospital, the average charge for Dr. Sparenocost (Physician #1550) is $18,130 (rounded); the average charge for Dr. Spendtheleast (Physician #1510) is $7,049. The question of interest is whether the average charges for Dr. Sparenocost are significantly different from those for Dr. Spendtheleast.

1. State the null and alternative hypotheses, and set the alpha level.
   H_0: $\mu_1 = \mu_2$
   H_1: $\mu_1 \neq \mu_2$
   $\alpha = 0.05$
2. Determine the region of rejection for a two-tailed two-sample *t* test where df $= n_1 + n_2 - 2 = (116 + 101) - 2 = 215$, $t_{crit} = \pm 1.96$.
3. Calculate the sample statistics:

	Dr. Sparenocost (#1550)	Spendtheleast (#1510)
Mean	$14,694	$5,203
S.D.	$24,685	$7,843
N	101	116

The pooled standard deviation:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{(101 - 1)24{,}685^2 + (116 - 1)7{,}843^2}{101 + 116 - 1}}$$

$$= 17{,}785.4$$

Calculation of *t*:

$$= \frac{\$5{,}203 - \$14{,}694}{17{,}785.4\sqrt{(1/116) + (1/101)}}$$

$$= \frac{-9{,}491}{2{,}420.5}$$

$$= -3.92$$

4. Conclusion: $t_{calc} = -3.92$; $t_{calc} \geq t_{crit}$ (df $= 215$) $= \pm 1.96$. We reject the null hypothesis and conclude that it appears that the average charges for Dr. Sparenocost are significantly different from the average charges for Dr. Spendtheleast.

Exhibit 6–7 SPSS Output for Two-Sample *t* Test

Group Statistics

	Charges	
	Physician	
	1510	*1550*
N	116	101
Mean	5203.1953	14694.444
Std. Deviation	7843.4506	24685.3201
Std. Error Mean	728.24611	2456.28124

Independent Samples Test

			Charges	
			Equal variances assumed	*Equal variances not assumed*
Levene's Test for Equality of Variances	F		6.397	
	Sig.		.012	
t-test for Equality of Means	*t*		−3.921	−3.705
	df		215	117.563
	Sig. (2-tailed)		.000	.000
	Mean Difference		−9491.24862	−9491.24862
	Std. Error Difference		2420.53525	2561.96407
	95% Confidence Interval of the Difference	Lower	−14262.26671	−14564.83023
		Upper	−4720.23053	−4417.66700

✓ To Obtain an Independent Samples *t* Test Using SPSS:

- From the menus, choose
  Analyze
  →Compare Means
  →Independent Samples *t* Test
- Select one or more quantitative variables. A separate *t* test is computed for each variable.
- Select a single grouping variable, and click "Define Groups" to specify two codes for the groups you want to compare.

The results indicate that the calculated value of *t*, −3.921, is greater than the critical value of *t*, −1.96. We therefore reject the null hypothesis and conclude that the mean charges for each doctor are significantly different from each other. Why is it important to know if the

charges of different physicians are significantly different from one another? The results could help us determine if one physician is more cost-effective than another in the delivery of health care services. Some possible questions that could be explored after obtaining this result include: Does one physician order more diagnostic tests than the other? Is the ALOS for the patients of one physician less than that for the other physician? If so, why? Answers to these and other questions can help us to educate physicians on the variations in their practice patterns. This may lead to more cost-effective delivery of health care services. It is important to remember that we should be careful in drawing any definitive conclusions before conducting further investigation. In the example, the mean charges for one physician were significantly different from the mean charges for the other. When significant results are achieved, the data analyst must conduct further investigation to determine if there is an explanation for the difference. Perhaps the patients of one physician are older and sicker than the other physician's patients. It would also be important to compare charges between all physicians who treat similar patients.

The results of our hand calculations are the same as the SPSS results. We are interested in the results in the column labeled "Equal Variances Assumed." SPSS automatically provides the results of Levene's test, which compares the equality of the variances between the groups. One of the assumptions of the *t* test is that the variances of the populations from which the two samples are drawn are equal. Since the result of the test is not significant, the variances of the two samples are assumed to be equal.

Excel may also be used for conducting the *t* test for comparing the means of two independent samples. The Excel output appears in Exhibit 6–8. The interpretation of the output is the same as that outline in Exhibit 5–8, with the exception of the "pooled variance." The pooled variance is the pooled standard deviation squared that appears in Exhibit 6–6.

Exhibit 6–8 Excel Output for Two-Sample *t* Test, Assuming Equal Variances

	1510	*1550*
Mean	5203.195	14694.444
Variance	61519718.6060	609365068.3906
Observations	116	101
Pooled Variance	316331509.203	
Hypothesized Mean Difference	0	
df	215	
t Stat	−3.921	
P(*T*<= *t*) one-tail	0.000	
t Critical one-tail	1.652	
P(*T*<= *t*) two-tail	0.000	
t Critical two-tail	1.971	

Paired *t* Test

In a **paired *t* test**, we are comparing the means of two samples that have been drawn from a single population. In a paired *t* test, the pair may be composed of two individuals who are matched on a set of characteristics such as height and weight, or the individual may be self-paired—serving as his or her own control. In this test, we are trying to determine if there is a difference in the means between the individuals who make up the pair or in the difference between "before" and "after" observations when one individual is serving as his or her control. The null and alternative hypotheses for the paired *t* test are

$$H_0: \overline{D} = 0$$
$$H_{A_1}: \overline{D} \neq 0$$

for a two-tailed test or

$$H_0: \overline{D} > 0$$
$$H_{A_2}: \overline{D} < 0$$

for a one-tailed test.

In the paired *t* test, *t* is calculated as

$$t = \overline{d}/(s_{\overline{d}})$$

where

$$\overline{d} = \sum D/N$$

and

$$s_{\overline{d}} = \sqrt{\sum d^2/N(N-1)}$$

and

$$\sum d^2 = \sum D^2 - (\sum D)^2/N$$

D is the difference in the observations before and after treatment for each individual in the study, or it is the difference in the observations between the experimental and control groups. $\overline{D}$ or $\overline{d}$ is the observed mean difference between the "before" and "after" observations or the mean observed difference between the experimental and control groups.

Let's consider an example where we are interested in improving the attitude of health information management (HIM) students toward statistics. There are 10 students in the statistics class; each student is given an attitude assessment prior to watching a video on the role of the HIM professional as a data analyst in health care. Following the video, the students are given a second attitude assessment. The null and alternative hypotheses are

$$H_0: \overline{D} = 0$$
$$H_{A_1}: \overline{D} < 0$$
$$\alpha = 0.05$$

The null hypothesis states that the difference between the pre- and postvideo assessment cores equals zero. The alternative hypothesis states that the postvideo assessment scores will be less than zero. The "before" and "after" data appear in Table 6–2.

Table 6–2 Pre- and Post-Attitude Assessment of HIM Students

Student	*Before* X_1	*After* X_2	*Difference* D	D^2
1	2	28	−3	9
2	23	19	4	16
3	30	34	−4	16
4	7	10	−3	9
5	3	6	−3	9
6	22	26	−4	16
7	12	13	−1	1
8	30	47	−17	289
9	5	16	−11	121
10	14	9	5	25
Total	171	208	−37	511

This is a one-directional *t* test. Why do we expect the average difference to be less than zero? Because we expect the postvideo attitude assessment score to be greater than the prevideo attitude assessment score. When subtracting the postvideo assessment score from the preassessment score, the expected result should be negative. The critical value of *t* for a one-direction test with nine degrees of freedom, $\alpha = 0.05$, is -1.833 (Appendix B, Table B–2). For the paired *t* test, degrees of freedom are equal to $n - 1$ where n is the number of pairs.

Before we can calculate *t*, we must determine the sum of the squares of the difference score.

$$\begin{aligned}\sum d^2 &= \sum D^2 - [(\sum D)^2/n] \\ &= 511 - [(-37)^2/10 \\ &= 511 - 136.9 \\ &= 374.10\end{aligned}$$

The standard error of the difference is

$$\begin{aligned}s_{\bar{d}} &= \sqrt{\sum d^2/n(n-1)} \\ &= \sqrt{374.10/10(9)} \\ &= 2.039\end{aligned}$$

and

$$\begin{aligned}\bar{d} &= \sum D/n \\ &= -37/10 \\ &= -3.7\end{aligned}$$

$$\begin{aligned}t &= \bar{d}/(s_{\bar{d}}) \\ &= -3.7/2.039 \\ &= -1.815\end{aligned}$$

Since the calculated value of t, -1.815, does not fall in the region of rejection for the critical value of t when $\alpha = 0.05$, -1.833, we fail to reject the null and conclude that there is not enough evidence to indicate that the video improved the attitude of HIM students regarding statistics.

We can use SPSS to calculate the paired t test; the results appear in Exhibit 6–9. Note that the SPSS calculated t matches our "hand-calculated" t. For a two-tailed test, the exact p value for a t calc of -1.815 (df $= 9$) is 0.103; for a one-tailed test, the p value is halved to 0.052.

✓ To Obtain a Paired-Sample *t* Test Using SPSS:

- From the menus, choose:
  Analyze
  →Compare means
  →Paired Samples *t* Test
- Select a pair of variables as follows:
  →Click each of two variables. The first variable appears in the "Current Selections" group as "variable 1," and the second appears as "variable 2."
  →After you have selected a pair of variables, click the arrow button to move the pair into the "Paired Variables" list. You may select more pairs of variables. To remove a pair of variables from the "Paired Variables" list, select a pair in the list and click the arrow button.

We may also use Excel to calculate a paired sample t test. The Excel Output appears in Exhibit 6–10. The Excel output includes the Pearson r correlation coefficient, which is $+0.87$ (see Chapter 8).

Exhibit 6–9 SPSS Output for Paired Sample *t* Test

Paired Samples Statistics

	Pair 1	
	Prevideo	*Postvideo*
Mean	17.1000	20.8000
N	10	10
Std. Deviation	10.2029	12.9168
Std. Error Mean	3.2265	4.0847

Paired Samples Test

			Pair 1 Prevideo–Postvideo
Paired Differences	Mean		−3.7000
	Std. Deviation		6.4472
	Std. Error Mean		2.0388
	95% Confidence Interval of the Difference	Lower	−8.3121
		Upper	.9121
t			−1.815
df			9
Sig. (2-tailed)			.103

Exhibit 6–10 Excel Output for Paired Sample *t* Test

	Prevideo	*Postvideo*
Mean	17.1	20.8
Variance	104.1	166.844
Observations	10	10
Pearson Correlation	0.870	
Hypothesized Mean Difference	0	
df	9	
t Stat	−1.815	
P(*T*<= *t*) one-tail	0.051	
t Critical one-tail	1.833	
P(*T*<= *t*) one-tail	0.103	
t Critical two-tail	2.262	

CONCLUSION

In this chapter, we have reviewed statistical procedures for testing hypotheses of the differences between means. We can use the *z* distribution for testing hypotheses involving one and two independent samples. To use the *z* distribution, we must assume that the samples are independent and are normally distributed and the sample size must be greater than 30. The population parameters μ and σ must be known if we are to use *z*.

When population parameters are not known, we can use the *t* distribution to test hypotheses of differences between population means. We can useto compare a sample mean with a population mean or to compare the mean of two samples drawn independently from two populations, and we can use the paired *t* test for comparing means between matched pairs. In using both *t* and *z*, we are restricted to comparing means of two samples.

ADDITIONAL RESOURCES

Hall, H.I. 1998. The *z* test. *Quality Resource* 16, no. 5: 7.

Jekel, J.F. et al. 1996. *Biostatistics, epidemiology, and preventive medicine*. Philadelphia: W.B. Saunders.

Katz, D.L. 1997. *Biostatistics, epidemiology, and preventive medicine review*. Philadelphia: W.B. Saunders.

Appendix 6–A

Exercises for Solving Problems

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
2. Compare and contrast the *z* distribution with the *t* distribution.
3. Describe situations in which we would use one-tailed tests; describe situations in which we would use two-tailed tests.
4. What assumptions must be met when using *z* and *t* for hypothesis testing? Why are these assumptions not always strictly followed?
5. In hypothesis testing, what is meant by the term effect?
6. Exhibit 6–9 displays the results of a paired *t* test in which we were interested in determining if the attitudes of HIM students toward statistics changed after viewing a video. We failed to reject the null hypothesis in this situation. What are some reasons for our failure to reject the null in this situation?

MULTIPLE CHOICE

1. It is easier for a statistical test to achieve statistical significance when we conduct a:
   a. nondirectional test
   b. one-tailed test
   c. two-tailed test
   d. all of the above
2. If a statistical test is significant at the 0.01 level, it is:
   a. also significant at the 0.05 level
   b. not significant at the 0.05 level
   c. also significant at the 0.001 level
   d. also significant at the 0.0001 level

For questions 3 through 6, refer to the problem below:

We are conducting a study of the age of nursing home patients in the county. The null and alternative hypotheses are

$$H_0: \mu = 80$$
$$H_A: \mu \neq 80$$

We have used the *z* test for evaluating our results when the critical values of *z* are ± 1.96 where $\alpha = 0.05$, and ± 2.58 where $\alpha = 0.01$.

3. In this case, the null hypothesis is:
   a. directional
   b. nondirectional
   c. a one-tailed *z* test
   d. a two-tailed *t* test

4. If the calculated *z* is $+2.30$, we would:
   a. accept the H_0 at $\alpha = 0.05$ but reject the H_0 at $\alpha = 0.01$
   b. accept the H_0 at $\alpha = 0.05$ and accept the H_0 at $\alpha = 0.01$
   c. reject the H_0 at $\alpha = 0.05$ but accept the H_0 at $\alpha = 0.01$
   d. reject the H_0 at $\alpha = 0.05$ and reject the H_0 at $\alpha = 0.01$

5. If the calculated value of *z* is $+1.80$, we would:
   a. accept the H_0 at $\alpha = 0.05$ but reject the H_0 at $\alpha = 0.01$
   b. accept the H_0 at $\alpha = 0.05$ and accept the H_0 at $\alpha = 0.01$
   c. reject the H_0 at $\alpha = 0.05$ but accept the H_0 at $\alpha = 0.01$
   d. reject the H_0 at $\alpha = 0.05$ and reject the H_0 at $\alpha = 0.01$

6. If the calculated value of *z* is -2.80, we would:
   a. accept the H_0 at $\alpha = 0.05$ but reject the H_0 at $\alpha = 0.01$
   b. accept the H_0 at $\alpha = 0.05$ and accept the H_0 at $\alpha = 0.01$
   c. reject the H_0 at $\alpha = 0.05$ but accept the H_0 at $\alpha = 0.01$
   d. reject the H_0 at $\alpha = 0.05$ and reject the H_0 at $\alpha = 0.01$

7. Which of the following is not needed to transform a score to a *z* score?
   a. mean
   b. variance
   c. raw score
   d. standard deviation

8. You believe that the hospital's average length of stay for DRG XXX is significantly less than the national average for the same DRG. Which of the following statistical tests would be most appropriate for answering this question?
   a. single-sample directional *t* test
   b. single-sample nondirectional *t* test

c. *t* test for two independent samples
d. any of the above

9. In the above problem, the sample size is 100. The number of degrees of freedom for evaluating the statistical significance of the test result is:
   a. 99
   b. 98
   c. 97
   d. not enough information provided

PROBLEMS

1. As an HIM DRG analyst, you are interested in comparing the mean length of stay (LOS) for Critical Care Hospital and the national mean for DRG 002, Craniotomy, Age Greater Than 17 without CC. The hospital mean LOS is 4.17, and the standard deviation is 2.57. The national average LOS for DRG 005 is 5.2 days. The summary data for Critical Care Hospital appear in Table 6–A–1 and Exhibit 6–A–1.

Table 6–A–1 Frequency Distribution for Length of Stay, DRG 002 in 2004 at Critical Care Hospital (SPSS Output)

		Frequency	*Percent*	*Valid Percent*	*Cumulative Percent*
Valid	1	2	8.7	8.7	8.7
	2	4	17.4	17.4	26.1
	3	6	26.1	26.1	52.2
	4	2	8.7	8.7	60.9
	5	4	17.4	17.4	78.3
	6	2	8.7	8.7	87.0
	7	1	4.3	4.3	91.3
	10	1	4.3	4.3	95.7
	11	1	4.3	4.3	100.0
	Total	23	100.0	100.0	

Exhibit 6–A–1 Length of Stay Descriptive Statistics for DRG 002 in 2003 at Critical Care Hospital (SPSS Output)

N	Valid	23
	Missing	0
Mean		4.17
Std. Error of Mean		.536
Median		3.00
Mode		3
Std. Deviation		2.570

a. State the null and alternative hypotheses and the *a priori* alpha level.
b. Calculate the difference between the hospital mean and the national mean using the one-sample *t* test. What is the number of degrees of freedom? What is the resultant *t* statistic? Is it statistically significant?
c. What are your conclusions?

2. For DRG 410, Chemotherapy without Acute Leukemia as Secondary Diagnosis, you want to determine if there is a difference in the length of stay by two payors—Managed Care and Medicare. The summary data appear in Tables 6–A–2 and 6–A–3.

Table 6–A–2 Frequency Distribution of Length of Stay by Payer, DRG 410, at Critical Care Hospital in 2004 (SPSS Output)

LOS * Payor Crosstabulation

Count

		Payor		
		Managed Care	*Medicare*	*Total*
LOS	1	12	3	15
	2	5	16	21
	3	1	5	6
	4	13	8	21
	5	10	5	15
	6	1	0	1
	7	1	0	1
	8	2	1	3
	9	1	1	2
	13	1	0	1
	14	1	0	1
	15	2	1	3
	23	1	0	1
	54	0	1	1
Total		51	41	92

Table 6–A–3 Mean and Standard Deviation for Length of Stay by Payer, DRG 410, at Critical Care Hospital in 2004 (SPSS Output)

Payor	*Mean*	*N*	*Std. Deviation*
Managed Care	4.80	51	4.382
Medicare	4.71	41	8.280
Total	4.76	92	6.379

a. State the null and alternative hypotheses and the *a priori* alpha level.
b. Use the *t* test for two independent sample means to determine if the observed difference between the two means is statistically significant. What is the number of degrees of freedom, the resultant *t* statistic, and the significance level?
c. What are your conclusions?

3. You have been monitoring the lengths of stay for two of your physicians who discharge the most patients from DRG 410, Chemotherapy without Acute Leukemia as a Secondary Diagnosis. The relevant statistics appear in Tables 6–A–4 and 6–A–5. You specifically want to know if the observed difference in the lengths of stay for physicians 1460 and 8210 is statistically significant.
   a. State the null and alternative hypotheses and the *a priori* alpha level.
   b. Use the *t* test for two independent sample means to determine if the observed difference between the two means is statistically significant. What are the number of degrees of freedom, the resultant *t* statistic, and the significance level?
   c. What are your conclusions?

Table 6–A–4 Frequency Distribution of Length of Stay by Physician, DRG 410, in 2004 at Critical Care Hospital (SPSS Output)

LOS * Physician Crosstabulation

		Physician		
		1460	*8210*	*Total*
LOS	2	13	21	34
	3	1	1	2
	4	1	0	1
	5	1	1	2
	13	1	0	1
	14	0	1	1
	15	1	0	1
	23	1	0	1
Total		19	24	43

Table 6–A–5 Mean and Standard Deviation for Length of Stay by Physician, DRG 410, in 2004 at Critical Care Hospital (SPSS Output)

Report

Physician	*Mean*	*N*	*Std. Deviation*
1460	4.68	19	5.812
8210	2.67	24	2.496
Total	3.56	43	4.350

CHAPTER 7

Analysis Of Variance

KEY TERMS

Analysis of Variance (ANOVA)
- Sum of squares between (SSB)
- Sum of squares within (SSW)
- Total sum of squares (TSS)

Post hoc procedures
- Tukey's honest significant difference (HSD) test
- Scheffé test

LEARNING OBJECTIVES

At the conclusion of this chapter, you should be able to:

1. Define key terms.
2. Use ANOVA to calculate the differences between two or more sample means.
3. Use statistical software to conduct ANOVA procedures.
4. When the ANOVA procedure results in a significant *F* statistic, use *post hoc* procedures to determine which means are significantly different.
5. Explain the purpose of post hoc tests.
6. Relate the concept of statistical power to beta error and sample size.

In Chapter 6, we discussed procedures for comparing two population means. But what do we do when we want to compare means when more than two groups are involved? If, for example, we had three populations, A, B, and C, for which we wanted to compare sample means, we could conduct multiple *t* tests by comparing A and B, A and C, and B and C. But this could become quite tedious: If we had 10 groups for which we wanted to compare means, we would have to make 45 comparisons. Also, in making these multiple comparisons, we increase the probability of making a type I error—rejecting the null hypothesis when it is true. The **analysis of variance (ANOVA)** procedure is used when we want to

compare the observed differences between two or more means from two or more independent samples.

ANALYSIS OF VARIANCE

As the name implies, ANOVA deals with variances rather than standard deviations and standard errors. In statistical terminology, variation refers to the sum of the squared deviations from the mean and is often referred to as the sum of squares: $\sum(X - \overline{X})^2$ or $\sum x^2$. When divided by the appropriate degrees of freedom, it is referred to as the variance: $\sum(X - \overline{X})^2/(N - 1)$. The goal of the ANOVA procedure is to explain the total variation in a study. To do this, we must obtain two independent estimates of variance, one based on the variability between groups **(sum of squares between, SSB)** and the other based on the variability within groups **(sum of squares within, SSW).** These combined (SSB + SSW) equal the **total sum of squares (TSS).**

To conduct the ANOVA procedure, the dependent variable must be continuous and/or be at least at the interval or ratio level of measurement. The samples should be drawn independently and randomly and be normally distributed. However, in cases where large samples are drawn, this assumption may be relaxed because of the central limit theorem. The variances of the group populations should be approximately equal. The independent variable is discrete or categorical.

The *F* distribution is used to test the difference between the two variance estimates. (ANOVA is often referred to as the *F* test, which is derived from the name of the individual who developed it, Sir Ronald Fisher.) Like the *t* distribution, the *F* distribution is a family of distributions based on the number of degrees of freedom associated with the variance estimates. The *F* distribution is a positively skewed distribution, so the calculated and critical values of *F* are positive.

The *F* ratio is obtained by dividing the mean of the between-group variance estimate (SSB) by the mean of the within-group variance estimate (SSW). If the obtained value of *F* equals or exceeds the critical value of *F*, the null hypothesis is rejected, and we conclude that the observed means differ more than what would be expected by chance alone. When conducting the ANOVA procedure, the alternative hypothesis is always one of inequality or nondirectional.

We will limit our discussion to the one-way ANOVA procedure. In the one-way ANOVA, we want to determine the effect of only one factor, or independent variable (IV), on the dependent variable (DV): for example, the effect of sex (IV) on length of stay (DV). The one-way ANOVA procedure for analyzing two groups is an extension of the *t* test for comparing two sample means as $t^2 = F$. In the case of two independent samples, the ANOVA procedure is used more often because it is considered more powerful in complex experimental research designs. The sum of squares is the basic concept in the ANOVA procedure. We have already learned to calculate the sum of squares in computing the variance. The raw score formula for the sum of squares is

$$\sum x^2 = \sum X^2 - (\sum X)^2/N$$

which is equal to $\sum (X - \bar{X})^2$, as previously described in Chapter 4.

The TSS can be broken down into the sum of squares within groups (SSW) and the sum of squares between groups (SSB). Thus, the basic ANOVA model is

$$\text{TSS} = \text{SSB} + \text{SSW}$$

The TSS represents the variation of all the observations around the grand mean. The grand mean is the mean of the samples when they are combined or treated as one. The grand mean is represented by $\bar{\bar{X}}$. The SSB represents the variation of the sample means around the grand mean, and the SSW represents the variation of the independent observations in each sample around their respective means. We will now conduct a simple ANOVA procedure to determine the significance of the difference between two sample means. In Table 7–1, the length of stay (LOS) for patients discharged from the medical intensive care unit (MICU) and surgical intensive care unit (SICU) are displayed. For this problem, we are interested in determining whether there is a significant difference between the average LOSs (ALOSs) for the SICU (X) and the MICU (Y). As with the *t* test, the first step is to state the null hypothesis and set the alpha level:

$$H_0: \mu_1 = \mu_2$$
$$H_A: \mu_1 \neq \mu_2$$
$$\alpha = 0.05$$

Table 7–1 Patient Lengths of Stay (LOSs) in Surgical and Medical Intensive Care Units (SICU and MICU), July 1–July 14, 20xx

Patient LOS SICU (X_1)	X_1^2	*Patient LOS MICU (X_2)*	X_2^2
21	441	9	81
19	361	10	100
18	324	20	400
13	169	14	196
15	225	18	324
20	400	5	25
22	484	8	64
25	625	11	121
17	289	12	144
10	100	13	169
$\Sigma X_1 = 180$	$\Sigma X_1^2 = 3{,}418$	$\Sigma X_2 = 120$	$\Sigma X_2^2 = 1{,}624$
$\bar{X}_1 = 18$		$\bar{X}_2 = 12$	
		$\bar{\bar{X}} = 15$	

The null hypothesis states that there is no significant difference in the ALOSs for patients discharged from the MICU and the SICU. The alternative hypothesis states that the observed difference in the ALOSs for MICU and SICU is significant. In ANOVA, the alternative hypothesis is always nondirectional since the F distribution is a positive distribution.

In conducting the ANOVA procedure, we first calculate the TSS, which is the sum of the squares for the two groups treated as one, we use the following formula:

$$\begin{aligned}\sum x^2 &= \sum X^2 - (\sum X)^2/N \\ &= (3{,}418 + 1{,}1624) - [(180 + 120)^2/20] \\ &= 542\end{aligned}$$

Next, we find the sum of squares within each group (SSW). For the SSW, each group is considered separately. The sum of squares for the SICU is

$$\begin{aligned}\sum x^2 &= \sum X_1^2 - [(\sum X_1)^2/N] \\ &= 3{,}418 - [(180)^2/10] \\ &= 178\end{aligned}$$

The sum of squares for the MICU is

$$\begin{aligned}\sum x^2 &= \sum X_2^2 - [(\sum X_2)^2/N] \\ &= 1{,}624 - (120)^2/10 \\ &= 184\end{aligned}$$

So the total SSW is 178 + 184 = 362.

Since TSS = SSB + SSW, we could subtract the SSW from the TSS to obtain the SSB. But to serve as a check on our calculations, we will directly calculate the SSB. The SSB is a measure of the variation of the group means about the combined mean. When the group means do not differ from each other, the SSB will be equal to zero. The greater the variation between the group means, the larger the SSB will be. The size of the SSB tells us how large the effect of the independent variable is on the dependent variable. In our example, the independent variable is type of care unit, and the dependent variable is the patient's LOS. The SSB is calculated as follows:

$$SSB = \sum n_i(\overline{X}_i - \overline{\overline{X}})^2$$

where n_i is the number of observations in each group, $\overline{\overline{X}}$ is the overall mean, or grand mean, and $\overline{X}_i$ is the mean for each group. From Table 7–1, we know that the mean for the SICU is 18 days, the mean for the MICU is 12 days, and the grand mean is 15 days. The calculation for the SSB is

$$\begin{aligned} SSB &= \sum n_i(\bar{X}_i - \bar{\bar{X}})^2 \\ &= 10(18 - 15)^2 + 10(12 - 15)^2 \\ &= 180 \end{aligned}$$

Thus, in terms of our ANOVA model, we now have

$$\begin{aligned} \text{TSS} &= \text{SSB} + \text{SSW} \\ 542 &= 362 + 180 \end{aligned}$$

Each of these sums of squares has a specified number of degrees of freedom. Since the TSS refers to the two groups as one, the degrees of freedom is equal to $n - 1$. The degrees of freedom for the SSW is equal to $n_i - 1$, where n_i is the number of observations in each group. But since we have more than one group, the number of degrees of freedom for the SSW will be equal to $k(n_i - 1)$, where k is equal to the number of groups. This latter formula applies only when the sample size for each group is the same. The number of degrees of freedom for the SSB is equal to $k - 1$, where k is the number of groups. In summary, the degrees of freedom for the F ratio are:

Source of Variation	*Degrees of Freedon*
SSB	$k - 1$
SSW	$k(n_i - 1)$, only when all sample sizes are equal
TSS	$n - 1$

After the sum of squares for each source of variation has been calculated, the data are summarized in an ANOVA table, as in Table 7–2. The components of the table are:

Source of Variation	*df*	*Mean Square*	*F*
SSB	$k - 1$	SBB/df	$\frac{\text{Means square SSB}}{\text{Means square SSW}}$
SSW	$k(n - 1)$	SSW/df	
TSS	$n - 1$		

A new column appears in Table 7–2 that we have not yet discussed—the mean square column. In the F test, we are comparing the SSB mean square to the SSW mean square. The mean squares are obtained by dividing the sum of squares for the SSB and SSW by their corresponding degrees of freedom. The SSB mean square is an estimate of the common

Table 7–2 ANOVA Table for Patient Length of Stay in Medical and Surgical Intensive Care Units

Source	*SS*	*df*	*Mean Square*	*F*
SSB	180	1	180.00	8.95
SSW	362	18	20.11	
TSS	542	19		

population variance that is independent of the variation in the group means; that is, how much the group means differ from the overall mean. This is the effect on each observation from belonging to that particular group and is due to the effect of the treatment. If the group means tend to cluster around the grand mean, there is no treatment effect.

The SSW mean square is an estimate of the population variance that is independent of the variance within groups; that is, how much the observations within each group are spread out around the group mean. Variation within groups is considered error; this is because if each subject within a group is treated the same, the expected outcome for each member of the group should be similar. That is, the differences in observations within a group cannot be due to differential treatment. Wide variation within a group would indicate that there was no relationship between the independent variable (care unit) and the expected outcome, so any effect could not be attributed to the independent variable. When we reject the null, we are stating that the between-group variation is greater than the within-group variation. To obtain the value of *F*, the SSB mean square is divided by the SSW mean square:

$$F = \frac{\text{Mean square between groups}}{\text{Mean square within groups}}$$

If the population means are equal, indicating no effect, *F* will equal 1. If they differ, the mean square SSB will be greater than the mean square SSW, and *F* will be greater than 1.

Recall that the *F* distribution is based on the degrees of freedom associated with the between-group variance estimate and the within-group variance estimate. In our example, there are two groups, so the degrees of freedom for SSB are equal to $k - 1$, or $2 - 1 = 1$. The within-group degrees of freedom are equal to $k(n_i - 1)$ or $2(10 - 1) = 18$.

To determine whether the *F* value is significant at our preset alpha level (0.05), we refer to the *F* table (Appendix B, Table B–3). In the *F* table, the degrees of freedom for the numerator are represented in the columns, and the degrees of freedom for the denominator are represented in the rows. To locate the critical value of *F*, we find the cell where the column and row for the designated degrees of freedom intersect. For 1 and 18 degrees of freedom, the critical value of *F* at .05 is 4.41. Since our calculated *F* ratio, 8.95, is larger than the critical value of 4.41, we reject the null hypothesis and conclude that it appears that the observed difference in the ALOSs for the MICU and the SICU is statistically significant.

When the value of F is significant, we conclude that the groups under study differ significantly from each other; that is, the groups show more variation than what can be attributed to random sampling from populations with a common population mean. The greater the effect, the larger the obtained F ratio. From the information contained in the ANOVA table, we can compute a correlation ratio, eta^2:

$$\begin{aligned} eta^2 &= SSB/TSS = 180/542 \\ &= 180/542 \\ &= 0.33 \end{aligned}$$

In our example, the correlation ratio is 0.33, or 33%. The correlation ratio explains how much variation in the dependent variable is explained by the independent variable. So we would state that 33% of the variation in length of stay is explained by the type of care unit.

To calculate this simple ANOVA procedure using SPSS, select "Compare Means" from the "Analyze" drop-down menu. Then select "Means." In the "Means" dialog box, type in "LOS" as the dependent variable and "care unit" as the factor. Click "Options" for the desired descriptive statistics and request the ANOVA table and eta^2. The SPSS output appears in Exhibit 7–1.

Exhibit 7–1 SPSS Output for "Compare Means"

Descriptives

Length of Stay

	SICU	*MICU*	*Total*
Mean	18.0000	12.0000	15.0000
N	10	10	20
Std. Deviation	4.44722	4.52155	5.34100
Std. Error of Mean	1.40633	1.42984	1.19428
Minimum	10.00	5.00	5.00
Maximum	25.00	20.00	25.00

ANOVA

Length of Stay

	Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Between Groups (Combined)	180.000	1	180.000	8.950	.008
Within Groups	362.000	18	20.111		
Total	542.000	19			

Measures of Association

	Eta	*Eta Squared*
Length of Stay * Care unit	.576	.332

In the dialog box, click "Options" to select a wide range of descriptive statistics. In the report section of Exhibit 7–1, the mean, standard deviation, variance, minimum and maximum values (range), and standard error of the mean are reported. The ANOVA table that appears in Exhibit 7–1 contains the information that we previously discussed except that the exact *p* value, 0.008, is provided. Eta and eta^2 are the same as what we calculated previously.

✓ To Obtain a One-Way Analysis of Variance Using SPSS:

- From the menus, choose:
  - →Analyze
    - →Compare means
      - →Select one or more dependent variables
      - →Select a single independent factor variable
      - →Click "Options" for descriptive statistics and ANOVA table

We can also use Excel to conduct the one-way ANOVA procedure; the results appear in Exhibit 7–2. The information provided is similar to the ANOVA output for SPSS. However, Excel also displays the critical value of *F*, whereas SPSS does not.

Exhibit 7–2 Excel Output for One-Way ANOVA Procedure

Groups	*Count*	*Sum*	*Average*	*Variance*
SICU	10	180	18	19.778
MICU	10	120	12	20.444

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	180	1	180	8.950	0.008	4.414
Within Groups	362	18	20.111			
Total	542	19				

ANOVA IN THE THREE-SAMPLE CASE

Determining if there is a difference between two sample means is rather straightforward. But what do we do when we have more than two sample means? How do we know which means are actually different from one another? When we are working with more than two samples, we could have a situation where only two of the three means were different from each other. Follow-up procedures, called post hoc tests , must be conducted to determine which means are significantly different from one another. We will look at two post hoc tests: the **Scheffé test** and **Tukey's honest significant difference (HSD) test**.

The Scheffé test performs simultaneous joint pairwise comparisons for all possible pairwise combinations of means. The Scheffé test uses the *F* distribution for testing the significance of the mean differences and is the most conservative method of making *post hoc* multiple comparisons. The advantage of the Scheffé test is that it can be used when the *n*'s in each group being compared are equal or unequal.

The **Tukey HSD** test uses the Studentized range statistic to make all of the pairwise comparisons between group means.

Let us now review an example where we wish to compare three sample means. Three physicians were compared in regard to the hospital LOS of their respective patients following a minor surgical procedure without complications. A sample of eight medical records were selected for each physician; the LOSs appear in Table 7–3.

Table 7–3 Patient Length of Stay by Physician

A	A^2	B	B^2	C	C^2
4	16	4	16	5	25
5	25	5	25	3	9
5	25	4	16	3	9
4	16	3	9	3	9
6	36	4	16	3	9
6	36	5	25	3	9
4	16	3	9	4	16
5	25	3	9	5	25
$\Sigma A = 36$	$\Sigma A^2 = 195$	$\Sigma B = 31$	$\Sigma B^2 = 125$	$\Sigma C = 29$	$\Sigma C^2 = 111$
$\overline{A} = 4.875$		$\overline{B} = 3.875$		$\overline{C} = 3.625$	

For this problem, we are interested in determining whether there is a significant difference in the ALOS for the patients of these three physicians. The null and alternative hypotheses and alpha level are

$$H_0: \mu_A = \mu_B = \mu_C$$
$$H_A: \mu_A \neq \mu_B \neq \mu_C$$
$$\alpha = 0.05$$

The null hypothesis states that the ALOSs for the patients of the three physicians are equal. The alternative hypothesis states that the ALOSs of the patients of the three physicians are not equal. To determine the TSS, we treat the three groups as one group.

$$\begin{aligned}\sum x^2 &= \sum X^2 - (\sum X)^2/N \\ &= (195 - 125 - 111) - [(39 + 31 + 29)^2/24] \\ &= 22.625\end{aligned}$$

Next, we find the sum of squares within each group (SSW). For the SSW, each group is considered separately. The SSW for physician A is:

$$\begin{aligned}\sum x^2 &= \sum X^2 - [(\sum X)^2/N] \\ &= 195 - [(39)^2/8] \\ &= 4.875\end{aligned}$$

The sum of squares for Physician B is

$$\begin{aligned}\sum x^2 &= \sum X^2 - [(\sum X)^2/N] \\ &= 125 - [(31)^2/8] \\ &= 4.875\end{aligned}$$

The sum of squares for Physician C is

$$\begin{aligned}\sum x^2 &= \sum X^2 - [(\sum X)^2/N] \\ &= 111 - (29)^2/8 \\ &= 5.875\end{aligned}$$

So the total SSW is 4.875 + 4.875 + 5.875 = 15.625.
After calculation of the overall mean, 4.125, the SSB is:

$$\begin{aligned}SSB &= \sum n_i (\overline{X}_i - \overline{\overline{X}})^2 \\ &= 8(4.875 - 4.125)^2 + 8(3.875 - 4.125)^2 + 8(3.625 - 4.125)^2 \\ &= 4.5 + 0.5 + 2.0 \\ &= 7.0\end{aligned}$$

Thus, in terms of our ANOVA model, we now have

$$\begin{aligned}\text{TSS} &= \text{SSB} + \text{SSW} \\ 22.625 &= 7.0 + 15.625\end{aligned}$$

The data are summarized in the ANOVA Table 7–4.

The critical F value for 2 and 21 degrees of freedom when $\alpha = 0.05$ is 3.47. Since our calculated F is greater than the critical F, we reject the null hypothesis and conclude that it appears that patient LOS does vary by physician. But with three groups of physicians, we do not know if all three means are different from each other or if only two means are different from

Table 7–4 ANOVA Table for Physicians A, B, and C

Source	*SS*	*df*	*Mean Square*	*F*
SSB	7.0	2	3.5	4.73
SSW	15.625	21	0.74	
SST	22.625	23		

each other. We must now perform ***post hoc*** **procedures** to determine where these differences lie. Calculations for the Tukey and Scheffé post hoc procedures appear in Exhibit 7–3.

Exhibit 7–3 Tukey HSD and Scheffé's Test

Tukey HSD

$$\text{HSD} = q(a)\sqrt{\text{MS}_{\text{SSW}}\ /n}$$

Where MS_{SSW} is the mean square of the SSW, a is the number of means to be compared, n is the number in each group, and q is the df associated with MS_{ssw}.

In our example, the df associated with MS_{ssw} is 21. From the distribution of the Studentized range statistic (Appendix B, Table B–4) for comparison of three means and 21 degrees of freedom, the critical value of t where $\alpha = 0.05$ is approximately 3.55. Thus:

$$3.55\sqrt{0.74/8} = 1.08$$

The difference between any two means must be at least 1.08.
Therefore:

$$\overline{A} = \overline{B}$$
$$\overline{B} = \overline{C}$$
$$\overline{A} \neq \overline{C}$$

The average length of stay for the patients of physician A is significantly different from the average length of stay for patients of physician C.

Scheffé Test

In the Scheffé test, F must be computed to make the comparison.

$$F = (\overline{X}_1 - \overline{X}_2)^2/[\text{MS}_{\text{SSW}}(n_1 + n_2)]/n_1n_2$$

For means A and B,

$$F = \frac{(4.875 - 3.875)^2}{0.74(8 + 8)/64}$$
$$= 5.41$$

For means A and C,

$$F = \frac{(4.875 - 3.625)^2}{0.74(8 + 8)/64}$$
$$= 8.45$$

For means B and C,

$$F = \frac{(3.875 - 3.625)^2}{0.74(8 + 8)/64}$$
$$= 0.34$$

To obtain the critical value of F, we multiply $(k - 1)$ by the critical value for F in the original ANOVA procedure. The number of groups in our analysis is three, so $k - 1 = 2$; and for 2 and 21 degrees of freedom, F_{crit} for comparison purposes is 6.94 (2×3.47).
For the differences between the two means to be statistically significant, the F's calculated from the above formula must exceed 6.94. In our example, the calculated F exceeds the critical value for F for only one comparison—means A and C, Therefore,

$$\overline{A} = \overline{B}$$
$$\overline{B} = \overline{C}$$
$$\overline{A} \neq \overline{C}$$

The average length of stay for the patients of physician A is significantly different from the average length of stay for patients of physician C.

Calculating the correlation coefficient eta^2:

$$\text{eta}^2 = \text{SSB/TSS} = 7.415/22.625$$
$$= 0.32$$

Eta^2 indicates that 32% of the variation in LOS is related to the physician. There could be a multitude of reasons for the patients of a particular physician having a significantly longer LOS, on average, than the other physicians. The physician may treat older patients or sicker patients, or physician A may have practice patterns that are different from those of physicians B and C. Whatever the reason, it should not be ascribed without investigation.

The results of the SPSS ANOVA procedure for the three means are displayed in Exhibit 7–4, and the results of the post hoc procedures appear in Exhibit 7–5. For both the Tukey HSD and the Scheffé, the ALOSs for physicians A and C are significantly different. The SPSS output in Exhibit 7–4 also displays homogeneous subsets by type of post hoc procedure. For both the Tukey HSD and the Scheffé test, the ALOSs for physicians B and C and physicians B and A are not significantly different. The results of the Excel ANOVA procedure are displayed in Exhibit 7–6. Excel does not provide post hoc procedures for analyzing differences between means for more than two groups.

Exhibit 7–4 SPSS Output for Comparing Group Means of More Than Two Groups

Descriptives

Length of Stay

		A	*B*	*C*	*Total*
N		8	8	8	24
Mean		4.8750	3.8750	3.6250	4.1250
Std. Deviation		.83452	.83452	.91613	.99181
Std. Error		.29505	.29505	.32390	.20245
95% Confidence Interval for Mean	Lower Bound	4.1773	3.1773	2.8591	3.7062
	Upper Bound	5.5727	4.5727	4.3909	4.5438
Minimum		4.00	3.00	3.00	3.00
Maximum		6.00	5.00	5.00	6.00

ANOVA

Length of Stay

	Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Between Groups	7.000	2	3.500	4.704	.021
Within Groups	15.625	21	.744		
Total	22.625	23			

Exhibit 7–5 SPSS Output for Post Hoc Procedures Comparing Group Means of More Than Two Groups

Multiple Comparisons

Dependent Variable: Length of Stay

	(I) Physician	(J) Physician	Mean Difference (I − J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	A	B	1.00000	.43129	.075	−.0871	2.0871
		C	1.25000(*)	.43129	.022	.1629	2.3371
	B	A	−1.00000	.43129	.075	−2.0871	.0871
		C	.25000	.43129	.832	−.8371	1.3371
	C	A	−1.25000(*)	.43129	.022	−2.3371	−.1629
		B	−.25000	.43129	.832	−1.3371	.8371
Scheffé	A	B	1.00000	.43129	.091	−.1357	2.1357
		C	1.25000(*)	.43129	.029	.1143	2.3857
	B	A	−1.00000	.43129	.091	−2.1357	.1357
		C	.25000	.43129	.846	−.8857	1.3857
	C	A	−1.25000(*)	.43129	.029	−2.3857	−.1143
		B	−.25000	.43129	.846	−1.3857	.8857
LSD	A	B	1.00000(*)	.43129	.031	.1031	1.8969
		C	1.25000(*)	.43129	.009	.3531	2.1469
	B	A	−1.00000(*)	.43129	.031	−1.8969	−.1031
		C	.25000	.43129	.568	−.6469	1.1469
	C	A	−1.25000(*)	.43129	.009	−2.1469	−.3531
		B	−.25000	.43129	.568	−1.1469	.6469

*The mean difference is significant at the .05 level.

Length of Stay

	Physician	N	Subset for alpha = .05	
			1	2
Tukey HSD[a]	C	8	3.6250	
	B	8	3.8750	3.8750
	A	8		4.8750
	Sig.		.832	.075
Scheffé[a]	C	8	3.6250	
	B	8	3.8750	3.8750
	A	8		4.8750
	Sig.		.846	.091

Means for groups in homogeneous subsets are displayed.
[a]Uses Harmonic Mean Sample Size = 8.000.

Exhibit 7–6 Excel Output for One-Way ANOVA for More Than Two Groups

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
A	8	39	4.875	0.696
B	8	31	3.875	0.696
C	8	29	3.625	0.839

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	7	2	3.5	4.704	0.021	3.467
Within Groups	15.625	21	0.744			
Total	22.625	23				

STATISTICAL POWER

In the previous chapter, we discussed the effect that sample size has on the achievement of statistical significance. We also discussed sample size in relation to type I and type II errors. Recall that type I error is the probability of rejecting the null hypothesis when it is true; type II error is the probability of accepting the null hypothesis when it is false. These concepts may be represented as

$$\alpha = \Pr(\text{rejecting } H_0 | H_0 \text{ is true})$$
$$\beta = \Pr(\text{accepting } H_0 | H_0 \text{ is false})$$

Ordinarily, in statistical testing, we control for type I error when we set the alpha level. However, the more strict the alpha level, the more the probability of making a type II error increases. To get around this problem, we often increase the sample size because this reduces type II error. Recall that the standard error of the mean is a function of sample size as demonstrated by

$$SE_{\bar{x}} = \sigma/\sqrt{n}$$
$$SE_{\bar{x}} = \sigma/\sqrt{100} = 1/10\sigma$$
$$SE_{\bar{x}} = \sigma/\sqrt{400} = 1/20\sigma$$

Therefore, we can say that increasing the sample size decreases sampling error and lowers the probability of committing a type II error.

Another way to avoid making a type II error is to conduct a statistical power analysis. Power analysis helps us decide (1) how large the sample size must be for accurate and reliable statistical judgments, and (2) how likely it is that our sample test statistic will detect effects for a given sample size.

We make many compromises when selecting the appropriate sample size. If our resultant decisions are to be accurate, the sample size must be large enough so that sampling error is small. But we do not want to choose sample sizes larger than needed because they are more expensive and time consuming to implement and administer. On the other hand, with small sample sizes, the results may be so imprecise as to be rendered useless.

Power analysis can help us select an appropriate sample size. The power of a statistical test is the ability of the test to reject the null hypothesis given that the null is false. This is represented as

$$\beta = \Pr(\text{rejecting } H_0 | H_0 \text{ is false}) = 1 - \beta$$

In other words, statistical power $(1 - \beta)$ is the probability of obtaining a statistically significant difference when one actually exists. The power of a statistical test is the complement of a type II error. Calculating statistical power for every possible test is beyond the scope of this text. However, we will review one example in which we calculate sample size controlling for type I error only and a second example in which we control for both type I and type II errors.

In most research, a beta error of 20% is set. This is equal to a z_β of 0.84. In most cases where $\alpha = 0.05$, for a two-tailed test, z_α is 1.96. In our example, we want to use a *t* test for two independent samples to determine if the mean difference between the LOSs for two special care units, MICU and SICU (Table 7–1), is at least 2 days. The standard deviation for the two groups together, as indicated in Exhibit 7–1, is 5.34. The sample size required, controlling for alpha error only, is

$$\begin{aligned} N &= (z_\alpha)^2 \times 2 \times (s)^2/\bar{d}^2 \\ &= (1.96)^2 \times 2 \times (5.34)^2/(2)^2 \\ &= 3.8416 \times 2 \times 28.5156/4 \\ &= 219.09/4 \\ &= 54.77, \textit{or } 55 \text{ cases} \end{aligned}$$

The result, 55 cases, is the sample size required for one sample; however, since we want to compare the ALOS for two samples, we multiply 55 times 2, which equals 110 cases. Or we could state that 55 cases are required in each sample, when we are interested in determining if the difference is at least two days and controlling for alpha error only. When we want to control for both alpha and beta error together, the formula becomes

$$\begin{aligned} N &= (z_\alpha + z_\beta)^2 \times 2 \times (s)^2/\bar{d}^2 \\ &= (1.96 + .84)^2 \times 2 \times (5.34)^2/(2)^2 \\ &= (7.84 \times 2 \times 28.5156)/4 \\ &= 447.125/4 \\ &= 111.78, \textit{or } 112 \text{ cases} \end{aligned}$$

Note that when we control for both alpha and beta errors together, the size of one sample doubles. This illustrates how larger sample sizes help us control for beta error. If we wanted to detect for even smaller differences between LOSs, such as one day, an even larger sample size would be required. We can readily see this by examining the mean difference expected in the formula above: the denominator would be changed from 4 (2^2) to 1 (1^2). If we are interested in detecting a difference of only one day, 447 cases would be required in each sample (447.125/1).

For other types of statistical problems, the statistical website of the University of California at Los Angeles has an online calculator for determining the power of a statistical test (www.stat.ucla.edu).

CONCLUSION

When we are interested in comparing means of two or more samples, we use ANOVA procedures. ANOVA also requires that the samples be randomly drawn, normally distributed, and independent of each other. When more than two samples are involved, and statistical significance is found so that the null hypothesis is rejected, we must conduct post hoc procedures to determine which sample means differ. The Tukey HSD and the Scheffé test were conducted to determine which of the three means differed from one another.

Last, we discussed the concept of statistical power. Statistical power helps us control for beta error in conducting our research. One way to control for beta error is to select large samples. Another is to use various formulas to determine the sample size appropriate for the type of statistical test that we use. If we wish to detect small differences between the means of the population under study, larger sample sizes will be required; if we wish to detect difference between sample means that may be somewhat larger, we can generally get by with a smaller sample size. If the sample size is excessively large, we run the risk of detecting a difference smaller than what is important, thereby not only wasting time and resources but making interpretation of the results more problematic.

ADDITIONAL RESOURCES

Hall, H.I. 1998. The *z* test. *Quality Resource* 16, no. 5: 7.

Katz, D.L. 1997. *Biostatistics, epidemiology, and preventive medicine review*. Philadelphia: W.B. Saunders.

Stockburger, D.W. 1998. Introductory statistics: Concepts, models and applications. ANOVA. www.psychstat.umsu.edu/introbook.

Appendix 7–A

Exercises for Solving Problems

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
2. Explain each component of the ANOVA model. How is the *F* ratio obtained?
3. To conduct the ANOVA procedure, the dependent variable must fall upon which scale of measurement? The grouping variable or independent variable falls upon which scale of measurement?
4. When conducting ANOVA to compare the means of three groups, under what conditions would we reject the null hypothesis? What conclusion would be drawn?
5. What is the purpose of conducting post hoc procedures?

MULTIPLE CHOICE

1. The CEO of Critical Care Hospital wants to compare average charges for congestive heart failure patients with those of two other acute care facilities in the community. Which of the following statistical tests would be most appropriate for answering the question?
   a. one-sample *t* test
   b. *t* test for two independent samples
   c. one-way ANOVA
   d. any of the above

For questions 2 through 4, refer to the following:
You are using ANOVA to compare the average age of patients discharged from two DRGs. In DRG XXX, there are 16 patients with an average age of 45. In DRG YYY, there are 20 patients with an average age of 50.

2. In the ANOVA table, the number of degrees of freedom for the SSB is:
   a. 3
   b. 2
   c. 1
   d. none of the above
3. The number of degrees of freedom for the SSW is:
   a. 34
   b. 35
   c. 36
   d. none of the above
4. The number of degrees of freedom for the SST is:
   a. 34
   b. 35
   c. 36
   d. none of the above
5. The ANOVA procedure may be used:
   a. with large sample sizes
   b. with small sample sizes
   c. when comparing the means of four groups
   d. all of the above
6. In the ANOVA procedure, we reject the null hypothesis when the calculated value of F is:
   a. zero
   b. greater than the critical value of F
   c. less than or equal to the critical value of F
   d. equal to the critical value of F
   e. b and d
   f. all of the above
7. If we fail to reject the null in the ANOVA procedure:
   a. the treatment had no effect on the subgroups under study
   b. the observed differences between the group means are statistically significant
   c. the observed differences between the group means are not statistically significant
   d. a and c
   e. all of the above
8. In the ANOVA procedure, the variations of the observation around their respective group means is an indication of:
   a. within-group variation
   b. between-group variation
   c. variation among the combined groups
   d. all of the above

9. We have conducted an ANOVA procedure in which we compared three population means. The calculated F equals 4.704, and the critical value of F is 3.467. Under this circumstance, we would:
   a. reject the null hypothesis
   b. conduct post hoc procedures
   c. fail to reject the null hypothesis
   d. a and b

10. You want to use the ANOVA procedure to compare the average length of stay of the patients of two physicians. The ANOVA procedure has the most power when:
    a. the two sample sizes are equal
    b. one sample is twice the size of the second sample
    c. one sample is three times the size of the second sample
    d. sample size has no effect on power

PROBLEMS

1. You have been analyzing hospital discharges from DRG 14, Intracranial Hemorrhage and Stroke with Infarction. You want to know if there is a difference in the average age of men and women discharged from DRG 14. The frequency distribution for discharges by sex appears in Table 7–A–1. You have decided to use the ANOVA procedure to calculate your results.
   a. State the null and alternative hypotheses and the alpha level that you will use.
   b. What is the mean age for men? What is the mean age for women?
   c. What is the calculated value of F? Is it statistically significant?
   d. What is your conclusion?

2. You also want to know if there is a difference in the average length of stay (ALOS) by gender for patients discharged from DRG 14. The frequency distribution for ALOS by gender appears in Table 7–A–2. You have decided to use the ANOVA procedure to calculate your results.
   a. State the null and alternative hypotheses and the alpha level that you will use.
   b. What is the ALOS for men? What is the ALOS for women?
   c. What is the calculated value of F? Is it statistically significant?
   d. What is your conclusion?

Table 7–A–1 Frequency Distribution of Age at Discharge by Gender, DRG 14, in 2004 at Critical Care Hospital (SPSS Output)

Age * Gender Crosstabulation

Count

		Gender		
		Female	*Male*	*Total*
Age	22	0	1	1
	23	0	1	1
	26	0	1	1
	39	1	0	1
	44	0	1	1
	46	1	0	1
	47	0	1	1
	49	1	0	1
	50	1	0	1
	52	0	1	1
	55	0	1	1
	56	0	1	1
	57	1	0	1
	57	0	1	1
	57	0	1	1
	58	1	0	1
	58	1	0	1
	60	0	1	1
	60	0	1	1
	64	0	1	1
	68	1	0	1
	68	1	0	1
	70	0	1	1
	70	0	1	1
	71	0	1	1
	71	0	1	1
	72	1	0	1
	72	0	1	1
	73	0	1	1
	75	0	1	1
	76	1	0	1
	77	1	0	1
	77	1	0	1
	78	0	1	1
	83	0	1	1
	84	0	1	1
	86	1	0	1
	86	1	0	1
Total		15	23	38

Table 7–A–2 Frequency Distribution of ALOS at Discharge by Gender, DRG 14, in 2004 at Critical Care Hospital (SPSS Output)

LOS * Gender Crosstabulation

Count

		Gender		
		Female	*Male*	*Total*
LOS	1	0	3	3
	2	4	7	11
	3	5	6	11
	4	1	0	1
	5	2	3	5
	6	1	2	3
	7	1	0	1
	12	1	1	2
	16	0	1	1
Total		15	23	38

3. Physicians 2170 and 8060 have the most patients discharged from DRG 14. You want to know if there is a difference in the average age and length of stay of patients of these two physicians. The frequency distribution for age and length of stay for these two physicians appears in Tables 7–A–3 and 7–A–4. You have decided to use the ANOVA procedure to calculate your results.
   a. State the null and alternative hypotheses and the alpha level that you will use.
   b. What is the average age for patients of physician 2170? What is the average age of patients for physician 8060?
   c. What is the average length of stay for patients of physician 2170? What is the average length of patients for physician 8060?
   d. What is the calculated value of *F* for each variable? Are they statistically significant?
   e. What is your conclusion? What factors may be influencing your results?

Table 7–A–3 Frequency Distribution of Age at Discharge, DRG 14, Physicians 2170 and 8060, in 2004 at Critical Care Hospital (SPSS Output)

Age * Physician Crosstabulation

Count

		Physician		
		2170	*8060*	*Total*
Age	26	0	1	1
	44	1	0	1
	46	0	1	1
	50	1	0	1
	52	1	0	1
	56	1	0	1
	57	0	1	1
	57	0	1	1
	57	0	1	1
	68	1	0	1
	68	0	1	1
	70	1	0	1
	75	1	0	1
	76	1	0	1
	77	0	1	1
	84	1	0	1
	86	0	1	1
Total		9	8	17

Table 7–A–4 Frequency Distribution of Length of Stay, Physicians 2170 and 8060, DRG 14, in 2004 at Critical Care Hospital (SPSS Output)

LOS * Physician Crosstabulation

Count

		Physician		
		2170	*8060*	*Total*
LOS	2	3	1	4
	3	4	3	7
	5	1	1	2
	6	0	1	1
	7	1	0	1
	12	0	2	2
Total		9	8	17

CHAPTER 8

Correlation and Linear Regression

KEY TERMS

Pearson *r* correlation coefficient
Coefficient of determination *r*
Scatter diagram
- Slope
- Intercept
- Line of best fit
- Regression line
- Standard error of the estimate s_{yx}

Multiple regression
Multicollinearity

LEARNING OBJECTIVES

Upon completion of this chapter, you should be able to:

1. Define key terms.
2. Define the Pearson *r* product moment correlation coefficient.
3. Construct scatter diagrams for variables *X* and *Y*.
4. Construct scatter diagrams using microcomputer statistical software.
5. Interpret the Pearson *r*.
6. Compare the Pearson *r* with the coefficient of determination.
7. Explain "line of best fit" in linear regression.
8. Explain "slope" and "intercept" in the regression model.
9. Construct linear regression models using microcomputer statistical software.
10. Conduct hypothesis testing for the Pearson *r* and linear regression.
11. Interpret regression models for given situations.
12. Differentiate between simple regression and multiple regression.
13. Explain multicollinearity.

Often we are interested in determining relationships between variables such as age and length of stay (LOS), age and survival time, or type of diet and cholesterol levels. To determine the extent to which two variables are related, we can calculate a correlation coefficient. There are many types of correlation coefficients, but we will limit our discussion to the Pearson *r*, or, more formally, the **Pearson *r* correlation coefficient**. The Pearson *r* is a measure of the linear relationship between two variables; it is used when both variables under study fall on the interval or ratio scale of measurement. There are other measures of association for variables that are either nominal or ordinal; these will be discussed in Chapters 9 and 10.

CHARACTERISTICS OF PEARSON *r*

To calculate the Pearson *r*, measures must be taken on two variables, *X* and *Y*—for example, height (*X*) and weight (*Y*). These measures are taken in pairs for each member of a sample randomly drawn from a population. The values of the calculated Pearson *r* range from −1.00 to +1.00. A correlation coefficient of −1.00 indicates that the two variables have a perfect negative relationship; a correlation coefficient of +1.00 indicates that the two variables have a perfect positive relationship; and a correlation coefficient of 0.0 indicates that there is no relationship between the two variables.

A positive relationship between the two variables means that as the measures on one variable increase, so do the measures on the second variable and, conversely, that as the measures on one variable decrease, so do the measures on the second variable. In other words, the measures on each variable move in the same direction. Thus, we can say that there is a direct relationship between the two variables.

A negative relationship means that the observations for the two variables are moving in opposite directions. As measures tend to increase on one variable, they tend to decrease on the second variable. In this situation, we say that there is an inverse relationship between the two variables.

The underlying assumption for the Pearson *r* is that the relationship between the two variables is linear. Since relationships between variables are not always linear, one should construct a scatter diagram or scatter plot to assess the type of relationship that exists between the two variables. We can construct a **scatter diagram** by plotting one variable, *X*, on the abscissa (horizontal axis) of a graph and plotting the second variable, *Y*, on the ordinate (vertical axis). If the points appear to approximate a straight line, then the two variables are linearly related, and it is appropriate to use the Pearson *r*. Scatter diagrams displaying positive and negative linear relationships appear in Figure 8–1; a scatter diagram indicating no relationship is also displayed.

To further illustrate, let's consider a health information management (HIM) example in which we are interested in determining if there is a relationship between age (*X*) and total charges for DRG 336, Transurethral Prostatectomy with CC. The scatter diagram of age (*X*) and total charges (*Y*) appears in Figure 8–2.

Examination of the scatter diagram indicates that the relationship between the two variables is negative. Imagine a line drawn through the plots from the upper left corner down to

Figure 8–1 Scatter Diagram of Linear Relationships

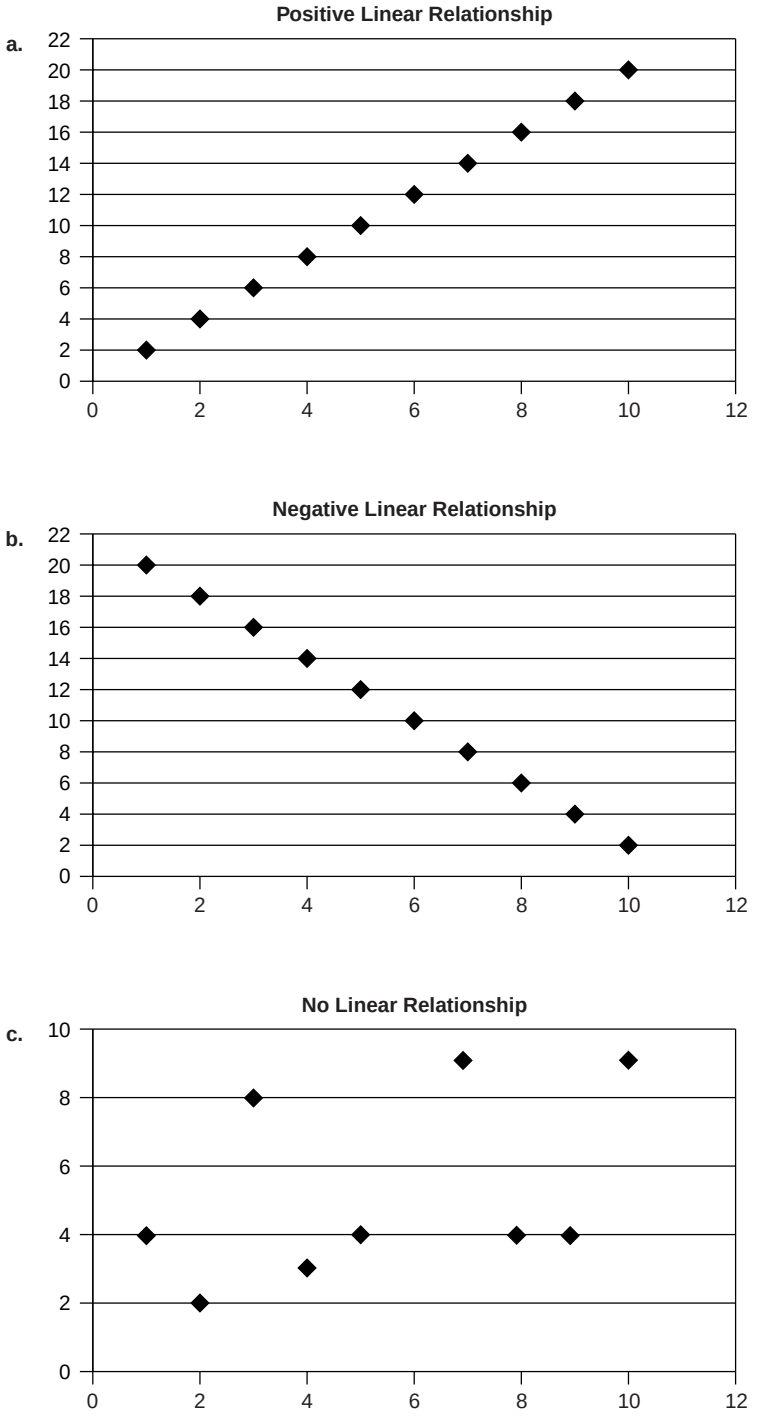
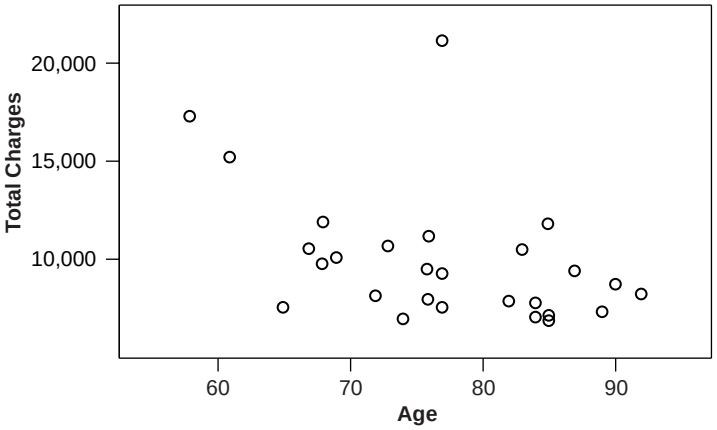


Figure 8–2 Scatter Diagram, Age and Total Charges for DRG 336 (SPSS Output)



the lower right. The diagram indicates that as age increases, total charges decrease. Therefore, the relationship between the two variables is considered negative. The means and standard deviations for age and charges for DRG 336 appear in Exhibit 8–1. The SPSS output for the Pearson *r* verifies that the relationship is negative (Exhibit 8–2). The Pearson *r* is −0.471, which is statistically significant ($p = 0.011$). The results indicate a moderate negative correlation between our two variables, age and total charges. The SPSS output displays the Pearson *r* for each variable with itself—for example, the Pearson *r* for "total charges" with "total charges" is 1.00. The Excel output for the Pearson *r* appears in Exhibit 8–3. Excel does not provide the *p* value for the Pearson *r*.

Exhibit 8–1 Means and Standard Deviations, Age and Total Charges, DRG 336 (SPSS Output)

		AGE	*Total Charges*
N	Valid	28	28
	Missing	0	0
Mean		77.4286	9803.7500
Median		77.0000	9003.5000
Mode		76.00[a]	6899.00[a]
Std. Deviation		9.11827	3326.68467
Minimum		58.00	6899.00
Maximum		92.00	21093.00

[a] Multiple modes exist. The smallest value is shown.

Exhibit 8–2 SPSS Output for Pearson *r*, Age and Total Charges, DRG 336

		AGE	*Total Charges*
AGE	Pearson Correlation	1	−.471*
	Sig. (2-tailed)		.011
	N	28	28
Total Charges	Pearson Correlation	−.471*	1
	Sig. (2-tailed)	.011	
	N	28	28

*Correlation is significant at the 0.05 level (2-tailed).

Exhibit 8–3 Excel Output for Pearson *r*, Age and Total Charges, DRG 336

	Age	*Totchg*
Age	1	
Totchg	−0.47111	1

This relationship seems somewhat paradoxical. Why do total charges tend to decrease with age? Most would expect the opposite to occur. We will consider this situation again later in the chapter.

✓ To Obtain a Simple Scatter Diagram Using SPSS:

- From the menus, choose:
  →Graphs
  →Scatter
- Select the icon for "Simple."
- Select "Define."
- Select a variable for the *x*-axis and a variable for the *y*-axis.

These variables must fall on the interval or ratio scale of measurement.

CALCULATION OF THE PEARSON *r*

The Pearson *r* is a sample of the true population correlation value, which is denoted by the Greek symbol ρ, and is subject to sampling variation. When calculating the Pearson *r*, we

are interested in testing the null hypothesis that $\rho = 0$; that is, the true population correlation is zero. A value of $\rho = 0$ indicates that there is no linear relationship between the two variables of interest. A significant *r* indicates that there is a relationship between the two variables of interest. Just as with the *t* and *F* tests, the obtained value of *r* is compared to a critical value of *r* to determine its statistical significance. For the Pearson *r*, we may conduct either a directional or nondirectional test.

The null and alternative hypotheses are

$$H_0: \rho = 0$$
$$H_A: \rho \neq 0 \text{ (two-tailed alternative)}$$
$$H_A: \rho < 0 \text{ or } \rho > 0 \text{ (one-tailed alternative)}$$

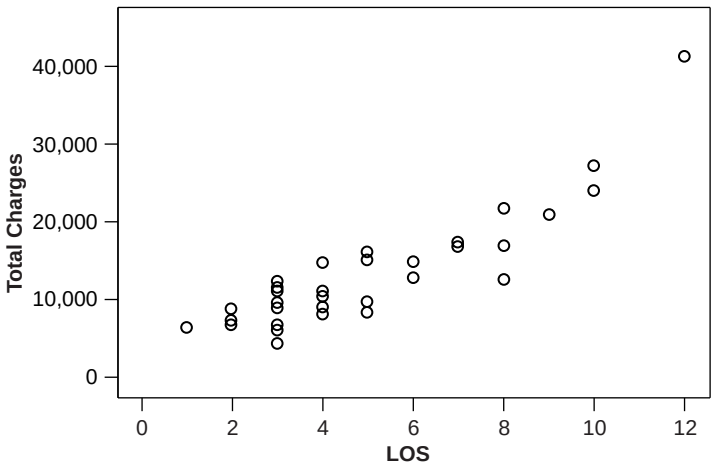
The formula for calculating the Pearson *r* is

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

Let us now follow the procedure for calculating the Pearson *r* for the two variables, LOS and total charges for DRG 087, Pulmonary Edema and Respiratory Failure. It should be obvious that the longer one stays in the hospital, the greater the charges, so one would expect a positive correlation to result.

We will first construct a scatter diagram to assess the relationship between the two variables. We want to determine whether our assumption of linearity is tenable. The scatter diagram appears in Figure 8–3.

Figure 8–3 Scatter Diagram, Length of Stay (LOS) and Total Charges DRG 087



Examination of the scatter diagram indicates a linear relationship between LOS and total charges. As the LOS increases, so do total charges. We would expect the resultant Pearson *r* to be positive.

The null and alternative hypotheses are

$$H_0: \rho = 0$$
$$H_A: \rho > 0$$
$$\alpha = 0.05$$

The null hypothesis states that there is no relationship between LOS and total charges. The alternative hypothesis states that there is a positive relationship between LOS and total charges, so we will be conducting a one-tailed test. We are using a one-tailed test because we expect that the relationship between the two variables will be positive—that is, LOS and total charges will move in the same direction. All cases discharged from DRG 087 are presented in Table 8–1. Microsoft Excel 200 was used to prepare the data required for calculating the Pearson *r*. Before we can calculate the Pearson *r*, we must first calculate Σx^2, Σy^2, and Σxy (from Table 8–1):

$$\begin{aligned}\sum x^2 &= \sum X^2 - \left[\sum (X)^2 / n\right] \\ &= 1{,}125 - [(173)^2 / 34] \\ &= 1{,}125 - (29{,}929 / 34) \\ &= 1{,}125 - 880.26 \\ &= 244.74\end{aligned}$$

$$\begin{aligned}\sum y^2 &= \sum Y^2 - \left[\sum (Y)^2 / n\right] \\ &= 7{,}915{,}583{,}426 - [(457{,}442)^2 / 34] \\ &= 7{,}915{,}583{,}436 - (209{,}253{,}183{,}364 / 34) \\ &= 7{,}915{,}583{,}436 - 6{,}154{,}505{,}393.06 \\ &= 1{,}761{,}078{,}032.94\end{aligned}$$

$$\begin{aligned}\sum xy &= \sum XY - \left[\sum (X)(Y) / n\right] \\ &= 2{,}905{,}363 - [(173)(457{,}442) / 34] \\ &= 2{,}905{,}363 - (79{,}137{,}466 / 34) \\ &= 2{,}905{,}363 - 2{,}327{,}572.53 \\ &= 577{,}790.47\end{aligned}$$

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$
$$= 577{,}790.47 / \sqrt{(9{,}244.74)(1{,}761{,}078{,}032.94)}$$
$$= 577{,}790.47 / \sqrt{431{,}006{,}237{,}781.74}$$
$$= 577{,}790.47 / 656{,}510.65$$
$$= 0.88$$

Table 8–1 Total Charges and Length of Stay (LOS) for Patients Discharged from DRG 087

Patient	*LOS* (*X*)	*Total Charges* (*Y*)	*LOS*^{2} (X^2)	*Total Charges*^{2} (Y^2)	*LOS x Charges XY*
1	1	$6,507	1	$42,341,049	$6,507
2	2	$8,771	4	$76,930,441	$17,542
3	2	$6,971	4	$48,594,841	$13,942
4	2	$7,405	4	$54,834,025	$14,810
5	3	$11,290	9	$127,464,100	$33,870
6	3	$8,944	9	$79,995,136	$26,832
7	3	$11,133	9	$123,943,689	$33,399
8	3	$4,304	9	$18,524,416	$12,912
9	3	$6,702	9	$44,916,804	$20,106
10	3	$12,143	9	$147,452,449	$36,429
11	3	$5,867	9	$34,421,689	$17,601
12	3	$11,061	9	$122,345,721	$33,183
13	3	$9,494	9	$90,136,036	$28,482
14	4	$10,920	16	$119,246,400	$43,680
15	4	$14,917	16	$222,516,889	$59,668
16	4	$8,222	16	$67,601,284	$32,888
17	4	$10,566	16	$111,640,356	$42,264
18	4	$9,389	16	$88,153,321	$37,556
19	5	$9,660	25	$93,315,600	$48,300
20	5	$15,106	25	$228,191,236	$75,530
21	5	$16,289	25	$265,331,521	$81,445
22	5	$8,285	25	$68,641,225	$41,425
23	6	$12,893	36	$166,229,449	$77,358
24	6	$14,840	36	$220,225,600	$89,040
25	7	$17,375	49	$301,890,625	$121,625
26	7	$16,925	49	$286,455,625	$118,475
27	8	$16,892	64	$285,339,664	$135,136
28	8	$12,462	64	$155,301,444	$99,696
29	8	$16,955	64	$287,472,025	$135,640
30	8	$21,754	64	$473,236,516	$174,032
31	9	$20,830	81	$433,888,900	$187,470
32	10	$23,915	100	$571,927,225	$239,150
33	10	$27,245	100	$742,290,025	$272,450
34	12	$41,410	144	$1,714,788,100	$496,920
Total	173	$457,442	1,125	$7,915,583,426	$2,905,363

The calculated Pearson *r* is +0.88, indicating a strong positive relationship. Referring to Table B–5 in Appendix B, we find that the critical value for *r* for 32 degrees of freedom, $\alpha = 0.05$ (df = $n - 2$, where n = number of pairs), and for a one-tailed test is 0.287. Since our calculated value for *r* exceeds the critical value, we reject the null hypothesis and conclude that the relationship between LOS and total charges is statistically significant. The SPSS descriptive statistics and results of Pearson *r* appear in Exhibit 8–4. The SPSS calculated Pearson *r* matches the results we obtained by using the hand-held calculator. The Excel output for the Pearson *r* appears in Exhibit 8–5.

Exhibit 8–4 SPSS Output for Pearson *r*, DRG 087, Total Charges and Length of Stay

Descriptive Statistics

	Mean	*Std. Deviation*
Total Charges	13454.1765	7305.20369
LOS	5.0882	2.72327
N	34	34

Correlations

		Total Charges	*LOS*
Total Charges	Pearson Correlation	1	.880**
	Sig. (2-tailed)	.	.000
	N	34	34
LOS	Pearson Correlation	.880**	1
	Sig. (2-tailed)	.000	.
	N	34	34

**Correlation is significant at the 0.01 level (2-tailed).

Exhibit 8–5 Excel Output for Pearson *r*, DRG 087, Total Charges and Length of Stay

	Length of Stay	*Total Charges*
Column 1	1	
Column 2	0.880	1

✓ To Obtain Pearson *r* Using SPSS:

- From the menus, choose
  →Analyze
  →Correlate
  →Bivariate
- Select two or more numeric variables.

It is important to remember that two variables' correlation with one another does not necessarily imply causality. We cannot assume that *X* causes *Y* or vice versa. In this example,

we cannot state that long LOSs cause high charges. We can only state that there is a strong relationship between the two variables. The two variables have a high correlation because they vary together in some systematic way.

In the previous example, where we were considering the relationship between age and total charges, it would not be logical to conclude that old age causes lower charges. Intuitively, this does not make sense. There must be some other variable at work that results in lower charges for older people.

Just as sample size plays a role in statistical significance when determining the difference between population means, it is also true when calculating the Pearson *r*. Small values of *r* may be statistically significant when there are many observations, while large values of *r* may not be statistically significant when there are a few observations.

From the Pearson *r*, we can calculate the **coefficient of determination r^2**. The r^2 tells us how much of the variation in *Y* is accounted for by the *X* variable. In our LOS (*X*) and total charges (*Y*) example, $r^2 = 0.774$ (0.88^2). So we conclude that 77.4% of the variation in total charges for DRG 087 is explained by the patient's LOS. The r^2 is a better measure of assessing the strength of a relationship between the two variables, *X* and *Y*, than *r*. The Pearson *r* alone can be used to make it seem as if the relationship between the two variables is much greater than it actually is. For example, a Pearson *r* equal to 0.50 appears to indicate a fairly strong relationship between two variables, when in fact only 25% (0.50^2) of the variance is accounted for by the two variables together. We will meet the coefficient of determination again in our discussion of linear regression.

INTRODUCTION TO LINEAR REGRESSION

In the previous section, we learned that two variables may have a linear relationship as designated by the Pearson product moment correlation coefficient. We also learned that correlation does not imply causality. Just because two variables, *X* and Y, have a high correlation with each other, we cannot assume that *X* causes *Y* or vice versa. For example, we would not state that height causes weight. But we can use this information in other ways. If the relationship between two variables is sufficiently large, we can predict the value of one variable from another. The objective of linear regression is to estimate the value of one variable that corresponds to the value of the other variable. In linear regression, we are trying to construct a mathematical model that explains the relationship between two variables.

Linear regression requires a pair of observations (*X* and *Y*) for each subject. The *Y* variable is usually designated as the dependent variable (DV), and the *X* variable is designated as the independent variable (IV). Our goal, then, is to predict *Y* from a given value of *X*. In our LOS (*X*) and total charges (*Y*) correlation problem discussed previously, the Pearson *r* is equal to 0.88, and the coefficient of determination, r^2, is equal to 0.774, indicating that 77.4% of the variation in total charges is explained by the variable LOS. Thus, there is a strong relationship between the two variables, and it is appropriate to develop a regression model that predicts total charges from LOS.

Just as with the Pearson *r*, in linear regression, a major assumption is that the two variables under consideration are linearly related. That is, a straight line can be used to describe the relationship between the two variables. A scatter diagram should be constructed to assess the relationship between the two variables. If the relationship appears to be linear, it can be described by a straight line. From high school algebra, you may recall that the general form for a straight line is

$$Y = a + bX$$

where *b i*s the slope of the line and *a* is the point where the line intercepts the *y*-axis. The slope represents the average change in *Y* that is associated with a change in *X*. The steeper the slope, the greater the change in *Y* that is associated with a change in *X*, and the stronger the relationship between the two variables of interest. The point at which *a* intercepts or crosses the *y*-axis is an estimate of the average value of *Y* when *X* is equal to zero.

For any two points, it is easy to determine the equation for the straight line. However, if there are three or more points, it is not possible to find a straight line that goes through all points simultaneously unless the correlation is a perfect ± 1.0. Thus, in linear regression we find the line that "best fits" all the points. The **line of best fit** is called **the regression line**.

The equation for the straight line indicates that for each observation of *X*, only one *Y* value is possible. This indicates that the measurement is precise—that is, without error. However, in reality, most relationship studies are inexact. And as you may recall from Chapter 3, error is integral to the measurement process. So the regression equation is more realistically expressed as

$$Y = a + bX + e$$

where *e* represents error. The error term acknowledges that the prediction equation does not perfectly predict *Y*. Thus, for a given *X*, there may be more than one *Y*. So the slope (*b*) indicates the average change in *Y* associated with *X*.

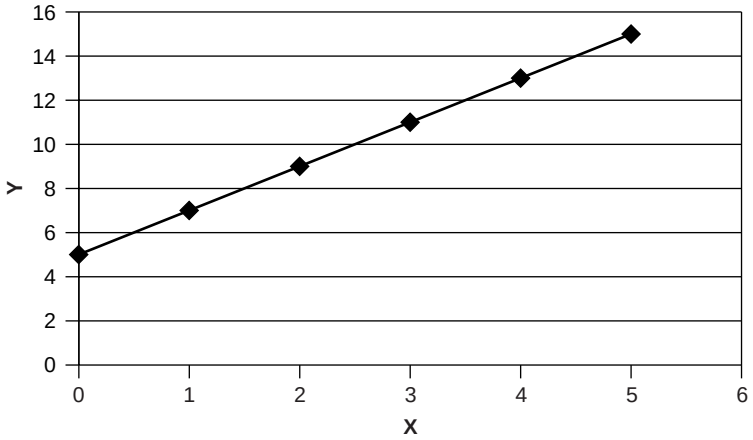
To illustrate these principles, consider the following data set for *X* and *Y*:

X	*Y*
0	5
1	7
2	9
3	11
4	13
5	15

In this data set, for each value of X there is only one value for Y. The relationship between the two variables is perfect, as shown in the scatter diagram in Figure 8–4. The regression line perfectly fits the X, Y data points. And the corresponding regression equation is $Y = 5 + 2X$. Note that the regression line crosses the y-axis at 5—the average value of Y when X is equal to zero.

$$\begin{aligned} Y &= 5 + 2X \\ &= 5 + 2(0) \\ &= 5 \end{aligned}$$

Figure 8–4 Scatter Diagram, X and Y



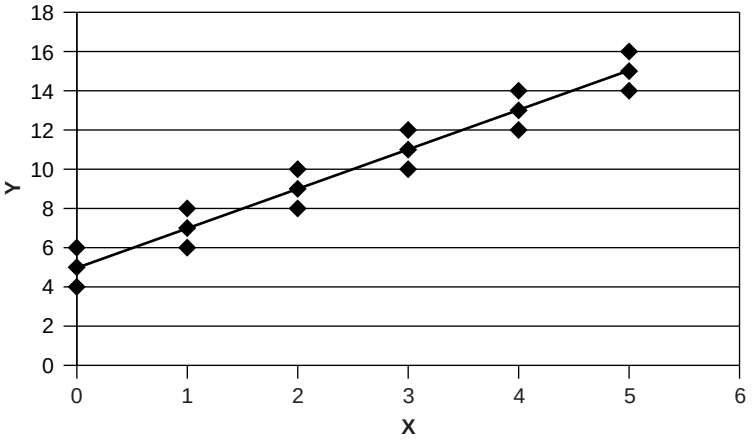
However, such perfection is rarely encountered in health care data analysis. Most often, we find the line that "best fits" all the points in the regression problems. As an example, consider the following data set in the following table:

X	Y	X	Y
0	4	3	10
0	5	3	11
0	6	3	12
1	6	4	12
1	7	4	13
1	8	4	14
2	8	5	14
2	9	5	15
2	10	5	16

In examining the data set, you can see that a given X variable does not take on the same value for Y each time. Therefore, the regression line will not perfectly fit all of the (X, Y) data points. The regression line that appears in the scatter diagram in Figure 8–5 is the line that best fits all of the data points. The distance between the data points and the regression line represents the error term in the regression equation. The distances of the observations from the line of best fit are represented as

$$d_i = Y_i - \hat{Y}$$

Figure 8–5 Scatter Diagram of Line of Best Fit



where $\hat{Y}$ pronounced "y-hat") is the predicted value of Y from X. Since the distance from the regression line may be either positive or negative, we compute the sum of the squared deviations from the regression line to measure the overall fitness of the line:

$$\sum d_i^2 = \sum (Y_i - \hat{Y})^2$$

or, when expressed as the error term,

$$\text{SSE} = \sum (Y_i - \hat{Y})^2$$

Error is the difference between the observed value of Y and the predicted value of Y $(\hat{Y})$. It is the amount of variation that cannot be accounted for in the regression model. The predicted value of Y $(\hat{Y})$ is the mean of the population of possible Ys for a given X. In Figure 8–5, you can see that the regression line falls through the means of the observed values of Y for a given X. Exhibit 8–6 shows the descriptive statistics and correlations for X and Y.

Exhibit 8–6 Pearson *r* Correlation Coefficients for *X* and *Y* (SPSS Output)

Descriptive Statistics

	X	*Y*
Mean	2.5000	10.0000
Std. Deviation	1.75734	3.61370
N	18	18

Correlations

		X	*Y*
X	Pearson Correlation	1	.973**
	Sig. (2-tailed)	.	.000
	N	18	18
Y	Pearson Correlation	.973**	1
	Sig. (2-tailed)	.000	.
	N	18	18

**Correlation is significant at the 0.01 level (2-tailed).

For some cases, we can set up the regression equation to predict either *Y* from *X* or *X* from *Y*. In other cases, it makes more sense to designate one variable as the independent variable (*X*) and the other variable as the dependent variable (*Y*). By convention, the DV is plotted on the *y*-axis and the IV on the *x*-axis. The equation for the sample regression line is written as

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1$$

where $\hat{Y}$ is the estimated value of *Y* given by the population regression line, $\hat{\beta}_0$ (pronounced "beta naught") is a constant that indicates where the regression line "intercepts" the *y*-axis and that estimates the average value of *Y* when $X = 0$, and $\hat{\beta}_1$ (pronounced "beta sub one") is the slope estimate that indicates the average change in *Y* associated with a change in *X*. Both $\hat{\beta}_0$ and $\hat{\beta}_1$ are referred to as the population regression coefficients and may vary from sample to sample.

The **slope** of the regression line ($\hat{\beta}_1$) indicates how steeply the regression line rises or falls. If the slope has an upward slant, the slope is positive and indicates that the correlation between *X* and *Y* is positive. If the slope has a downward slant, the slope is negative and indicates that the correlation between *X* and *Y* is negative (Figure 8–1).

To develop the regression equation, we need to find the values for the regression coefficients, $\hat{\beta}_0$ and $\hat{\beta}_1$. As stated earlier, we can either regress *Y* from *X* or *X* from *Y*. To solve for $\hat{\beta}_0$ and $\hat{\beta}_1$ when *Y* is regressed from *X*, we have:

$$\beta_{1yx} = \sum xy / \sum x^2$$

and

$$\beta_{0yx} = \overline{Y} - \beta_1 \overline{X}$$

and to regress *X* from *Y*, we have

$$\beta_{1xy} = \sum xy / \sum y^2$$

and

$$\beta_{0xy} = \overline{X} - \beta_1 \overline{Y}$$

To illustrate, we will use some hypothetical height and weight data of nine patients. These data appear in Table 8–2. The scatter diagram and the descriptive statistics and correlations for height and weight appear in Figure 8–6 and Exhibit 8–7, respectively. The scatter diagram indicates a positive linear relationship.

Table 8–2 Height and Weight Measurements in Nine Patients

Patient	*Height* (*X*) (*in* inches)	*Weight* (*Y*)	X^2	Y^2	XY
1	60	135	3,600	18,225	8,100
2	60	120	3,600	14,400	7,200
3	62	140	3,844	19,600	8,680
4	62	130	3,844	16,900	8,060
5	62	135	3,844	18,225	8,370
6	64	145	4,096	21,025	9,280
7	66	150	4,356	22,500	9,900
8	68	150	4,624	22,500	10,200
9	68	160	4,624	25,600	10,880
Total	572	1,265	36,432	178,975	80,670
	$\overline{X} = 63.6$	$\overline{Y} = 140.6$	$\Sigma X^2 = 36{,}432$	$\Sigma Y^2 = 178{,}975$	$\Sigma XY = 80{,}670$
	$s_x = 3.13$	$s_y = 12.1$	$\Sigma x^2 = 78.2$	$\Sigma y^2 = 1{,}172.2$	$\Sigma xy = 272.2$

Figure 8–6 Scatter Diagram, Height and Weight

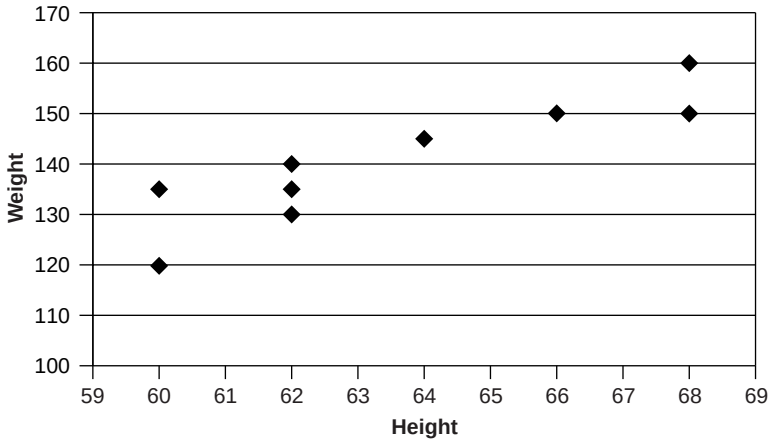


Exhibit 8–7 Descriptive Statistics and Correlation for Height and Weight (SPSS Output)

Descriptive Statistics

	Height	*Weight*
Mean	63.5556	140.5556
Std. Deviation	3.12694	12.10487
N	9	9

Correlations

		Height	*Weight*
Height	Pearson Correlation	1	.899**
	Sig. (2-tailed)	.	.001
	N	9	9
Weight	Pearson Correlation	.899**	1
	Sig. (2-tailed)	.001	.
	N	9	9

**Correlation is significant at the 0.01 level (2-tailed).

Using the data in Table 8-2, we will regress *Y* (weight) from *X* (height):

$$\begin{aligned}\beta_{1yx} &= \sum xy / \sum x^2 \\ &= 272.2 / 78.2 \\ &= 3.48\end{aligned}$$

$$\begin{aligned}\beta_{0yx} &= \overline{Y} - \beta_1 \overline{X} \\ &= 140.6 - 3.48(63.6) \\ &= 140.6 - 221.328 \\ &= -80.7\end{aligned}$$

Thus, to predict weight from height, our regression equation is

$$\hat{Y} = -80.7 + 3.48X$$

Alternatively, we can also regress *X* from *Y* as follows:

$$\begin{aligned}\beta_{1xy} &= \sum xy / \sum y^2 \\ &= 272.2 / 1{,}172.2 \\ &= 0.23\end{aligned}$$

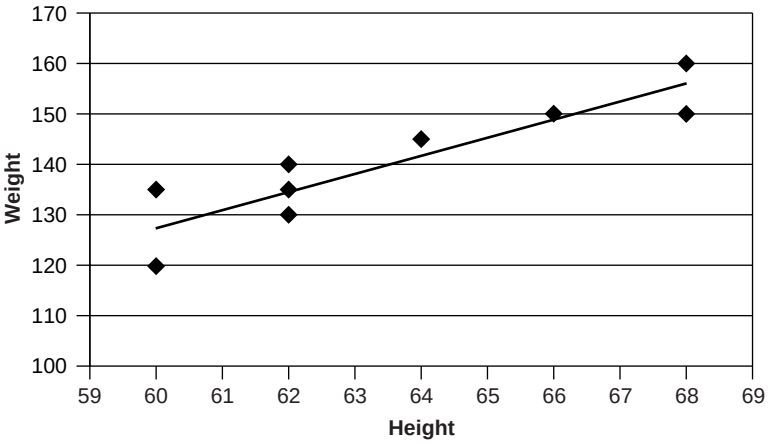
$$\begin{aligned}\beta_{0xy} &= \overline{X} - \beta_1 \overline{Y} \\ &= 63.6 - (0.23)140.6 \\ &= 31.3\end{aligned}$$

So, to predict height from weight, the regression equation is

$$\hat{X} = 31.3 + 23Y$$

We will now reproduce the scatter diagram that appears in Figure 8–6 with the addition of the regression line (Figure 8–7).

Figure 8–7 Scatter Diagram and Regression Line—Height and Weight



As we have already discussed, because regression deals with prediction, there is error. It is unlikely that the predicted value of Y will correspond exactly to the actual value of Y for a given value of X. Not all of the (X, Y) plots in Figure 8–7 fall on the regression line. For example, if a woman is 5′5″ (65 inches), we do not expect that her weight will be exactly 145.5 pounds, as predicted by the regression equation

$$\begin{aligned}\hat{Y} &= -80.7 + 3.48X \\ &= -80.7 + 3.48(65) \\ &= 145.5\end{aligned}$$

The predicted value is an estimate of the average weight of individuals who are of that height. As you may recall, the average is the best estimate or the most typical value in a distribution. In this case, it is the best estimate of a person's weight. If the correlation between the two variables is low, there will be considerable variation of the actual values around the predicted values, and if the correlation is high, the actual values will cluster more closely around the predicted values. Only when the correlation is a perfect ± 1.0 (termed unity) will the actual value match the predicted values.

If there is a great deal of scatter in the observed values of Y around the regression line, the predicted values of Y based on the regression equation will not be very close to the observed values of Y. The standard error of the estimate is a measure of scatter or spread of the observed values of Y around the corresponding values estimated from the regression equation.

Just as we calculate the standard error of the mean, we can calculate the **standard error of the estimate**, designated s_{yx}. The standard error of the estimate measures the scatter of the observed values of Y around the predicted values of Y. To calculate the standard error of the estimate,

$$s_{yx} = \sqrt{\text{SSE} / n - 2}$$

where we have already seen that

$$\text{SSE} = \sum (Y_i - \hat{Y})^2$$

But since it is rather cumbersome to calculate the SSE using this formula, we will use the alternative (the standard deviation for Y appears in Table 8–2):

$$\begin{aligned} s_{yx} &= s_y \sqrt{1 - r^2} \\ &= 12.1 \sqrt{1 - 0.808} \\ &= 5.30 \end{aligned}$$

The standard error of the estimate is a type of standard deviation, the standard deviation of the distribution of obtained Y scores about the predicted Y score, and it is used to develop a confidence interval around $\hat{Y}$.

INTERPRETATION OF THE STANDARD ERROR OF THE ESTIMATE

Recall from our discussion of the normal curve and standard deviation in Chapter 5 that

68% of the scores fall within the limits $\overline{X} \pm 1s$
95% of the scores fall within the limits $\overline{X} \pm 1.96s$
99% of the scores fall within the limits $\overline{X} \pm 2.58s$

Since the standard error of the estimate is a kind of standard deviation, we can make a similar interpretation regarding $\hat{Y}$ in that the obtained scores (Y) are normally distributed around $\hat{Y}$. Thus:

68% of the obtained values of Y fall within the limits $\hat{Y} \pm s_{yx}$
95% of the obtained values of Y fall within the limits $\hat{Y} \pm 1.96s_{yx}$
99% of the obtained values of Y fall within the limits $\hat{Y} \pm 2.58s_{yx}$

For example, suppose we have 10 students and we ask each his or her height and predict his or her weight on the basis of the regression equation.

$$\hat{Y} = -80.7 + 3.48X$$

We then go back and determine the actual weight of each of these 10 students and compare their actual weight to their predicted weight. The difference between their actual weight (Y) and their predicted weight ($\hat{Y}$) is error and describes how scores vary around the regression line that follows a normal distribution. Using the standard error of the estimate, we can develop 95% confidence interval for $\hat{Y}$. For small sample sizes, the t distribution for the appropriate number of degrees of freedom is used to calculate the confidence interval rather than the normal distribution.

$$CI_{95} = \hat{Y} \pm t_{0.05}(s_{yx})$$

For the predicted weight of 145.5, the standard error of estimate is 5.30. For seven degrees of freedom ($n - 2$), $\alpha = 0.05$, the critical t is 2.365. The 95% confidence interval is calculated:

$$\begin{aligned} CI_{95} &= \hat{Y} \pm t_{05}(s_{yx}) \\ &= 145.5 \pm 2.365\ (5.30) \\ &= 145.5 \pm 12.53 \\ &[132.97,\ 158.03] \end{aligned}$$

The interpretation is that for a height of 65 inches, 95% of the obtained weights will fall within a range of 132.97 to 158.03 pounds.

Caution must be exercised when predicting Y beyond the range of the actual observations upon which the data analysis is based. In our current problem where we are predicting weight from height, the relationship described by the straight line may hold for a height of 70 inches, which is not too far beyond the range of observations in our data set. But as we move to greater heights such as 75 inches or to lesser heights such as 36 inches, the same linear relationship may no longer continue.

Also, note that in the regression equation for predicting weight from height, the constant is -80.7. A negative value indicates that the regression line **intercepts** the y-axis at a point below zero. The literal interpretation would be that for certain heights, the predicted weight was less than zero. But we know that this is not possible. The constant is a fixed value that ensures that the predicted value "comes out right."

An example that illustrates both of these points is the case of a newborn whose length at birth is 20 inches. According to our regression equation, the predicted weight is -11.1 pounds.

$$\begin{aligned} \hat{Y} &= -80.7 + 3.48X \\ &= -80.7 + 3.48(20) \\ &= -80.7 + 69.6 \\ &= -11.1 \end{aligned}$$

We know that this result is impossible. In this example, the newborn's length is considerably outside the range of observations upon which our regression equation was modeled,

and the negative constant brings the predicted weight to less than zero. This demonstrates the importance of the researcher's judgment when using statistics. Empirical data together with the judgment of the researcher are required in the decision-making process.

HYPOTHESIS TESTING

Before using the regression model for actual predictions, we must conduct a statistical test to determine that the predicted slope does not equal zero. In regression, the hypothesis test is for the regression coefficient for the slope of the line ($\hat{\beta}_1$), which is indicative of the correlation between X and Y. The slope is defined as

$$\text{Slope} = \frac{\text{change in } X}{\text{change in } Y}$$

If the slope equals zero, Y will be a constant that does not change with changes in X. The null hypothesis is that the true slope of $\hat{\beta}_1$ of the population regression line is equal to zero:

$$H_0: \beta = 0$$
$$H_A: \beta \neq 0$$

The t distribution where df $= n - 2$ is used to test the null hypothesis, where

$$t = \hat{\beta}_1/(s_{yx}/\sqrt{\sum x^2}$$

For our example of height and weight, $\hat{\beta}_1 = 3\ 48$, $\Sigma x^2 = 78.2$, and $s_{yx} = 5.30$. So,

$$\begin{aligned} t &= 3.48/(5.30/\sqrt{78.2}) \\ &= 3.48 / .60 \\ &= 5.84 \end{aligned}$$

For $\alpha = 0.05$, the tabled t for 7 degrees of freedom is 2.365. Since the calculated t is greater than the critical t, we reject the null hypothesis of no linear relationship between X and Y.

To construct a 95% confidence interval around the regression coefficient, we have

$$\begin{aligned} CI_{95} &= \hat{\beta}_1 \pm t_{0.05}(s_{yx}/\sqrt{\sum x^2}) \\ &= 3.48 \pm 2.365\ (530/\sqrt{78.2}) \\ &= 3.48 \pm 2.365(0.60) \\ &= 3.48 \pm 1.42 \\ &[2.06, 4.90] \end{aligned}$$

Thus, we are 95% confident that the population regression coefficient (β) falls between 2.06 and 4.90.

COEFFICIENT OF DETERMINATION

We previously encountered the coefficient of determination (r^2) in our discussion of correlation. The r^2 indicates the explanatory power of our linear model. The range of r^2 is 0 to +1. The r^2 indicates the amount of variation in the dependent variable that is explained by the independent variable. When $r^2 = +1.0$, the independent variable accounts for 100% of the variation in the dependent variable, and all of the observations fall on the regression line. When r^2 equals zero, the IV accounts for no variation in *Y* and is not helpful in predicting *Y*. The closer the r^2 is to 1, the better the fit of the regression line to the data points. When r^2 is close to zero, the two variables are said to be independent of each other—that is, they do not vary together in any systematic way. A high value for r^2 is necessary if our predictions are to be accurate. We know that for the height and weight problem, $r = 0.899$ and $r^2 = 0.808$. Thus, 80.8% of the variation in weight is explained by an individual's height.

F TEST

SPSS provides an analysis of variance (ANOVA) model for the regression equation that indicates the significance of the regression model. The components of the model are

$$\text{Total variation} = \text{regression} + \text{residual (Error)}$$

The variation that can be explained by the regression model is represented in the "regression" component of the model, and the unexplained variance (error) is represented in the "residual" component of the model. Thus, the model may be expressed as:

$$\text{Total variation} = \text{explained variation} + \text{unexplained variation}$$

These terms are explained below:

Variation in the ANOVA Model Source of Variation Explanation

Source of Variation	*Explanation*
Total variation $\Sigma(Y - \overline{Y})^2$	The sum of the squares of the differences between the observed value of *Y* and the mean value of *Y*.
Explained variation $\Sigma(Y_c - \overline{Y})^2$	The sum of the squares of the differences between the calculated value of *Y* and the mean value of *Y*.
Unexplained variation $\Sigma(Y - Y_c)^2$	The sum of the squares of the differences between the observed value of *Y* and the calculated value of *Y* for a given *X*.

We will now use SPSS to verify our hand calculations. With slight differences due to rounding, the SPSS output in Exhibit 8–8 matches our hand calculations—for example, our hand-calculated standard error of the estimate is 5.3, and that provided by SPSS is 5.67. This is because the SPSS-provided standard error was based on the adjusted r^2 of 0.781 rather than the r^2 of 0.808 used in our calculations. The ANOVA indicates that the regression mode is significant ($F = 29.492$, df = 1, 7, $p = 0.001$). We obtain the formula for the regression model from the coefficients table in Exhibit 8–8. The constant is −80.625, and the β_1 coefficient for the slope is 3.48; the calculated t for the slope is significant. Our hand-calculated t is slightly different from the t provided by SPSS because of the slight difference in the standard error. As before, our regression model is

$$\hat{Y} = -80.625 + 3.48X$$

Exhibit 8–8 SPSS Output for Regression Model, Height and Weight

Model Summary

Model	*R*	*R Square*	*Adjusted R Square*	*Std. Error of the Estimate*
1	.899[a]	.808	.781	5.66768

[a] Predictors: (Constant), Height

ANOVA[b]

Model		*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
1	Regression	947.364	1	947.364	29.492	.001[a]
	Residual	224.858	7	32.123		
	Total	1172.222	8			

[a] Predictors: (Constant), Height
[b] Dependent Variable: Weight

Coefficients[a]

		Unstandardized Coefficients		*Standardized Coefficients*		
Model		*B*	*Std. Error*	*Beta*	*t*	*Sig.*
1	(Constant)	−80.625	40.772		−1.977	.089
	Height	3.480	0.641	.899	5.431	.001

[a] Dependent Variable: Weight

The Excel output for the regression model appears in Exhibit 8–9. The Excel output is similar to that for SPSS.

Exhibit 8–9 Excel Output for Regression Model, Height and Weight

SUMMARY OUTPUT

Regression Statistics	
Multiple *R*	0.899
R Square	0.808
Adjusted *R* Square	0.781
Standard Error	5.668
Observations	9

ANOVA

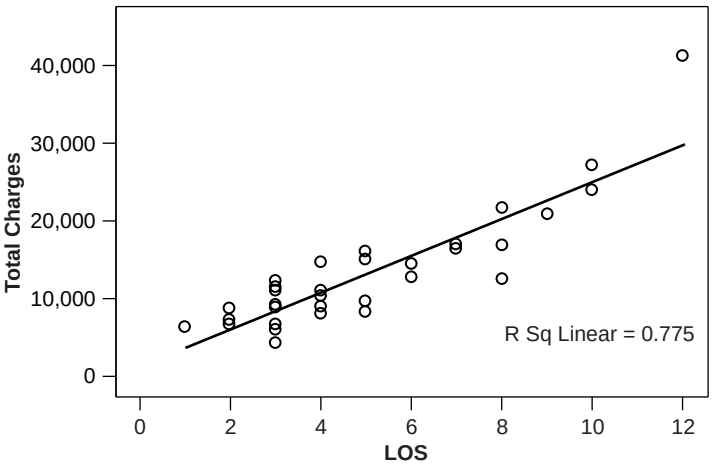
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	947.364	947.364	29.492	.001
Residual	7	224.858	32.123		
Total	8	1172.222			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	−80.625	40.772	−1.977	0.089	−177.035	15.785	−177.035	415.785
Height	3.480	0.641	5.431	0.001	1.965	4.995	1.965	4.995

REGRESSION MODEL FOR LENGTH OF STAY AND TOTAL CHARGES

Now let's return to the problem of LOS and total charges for DRG 087. We will now construct a new scatter diagram that includes the regression line (Figure 8–8).

Figure 8–8 Scatter Diagram, Length of Stay (LOS) and Total Charges, DRG 087



In this example, the means and standard deviations are provided for both LOS and total charges. The SPSS model summary displayed in Exhibit 8–10, indicates a fairly strong relationship between LOS and total charges that is statistically significant ($r = 0.880$, $p < 0.01$). In the model summary, the investigator can find r^2, the coefficient of determination, which in this example is 0.775, indicating that LOS accounts for approximately 77.5% of the variation in total charges. The standard error of the estimate is 3,552.182. The Excel regression statistics for LOS and total charges for DRG 087 are displayed in Exhibit 8–11.

In the "Coefficients" section, we find the regression coefficients, LOS or $\hat{\beta}_1 = 2{,}360.879$, and the constant, $\hat{\beta}_0 = 1{,}441.467$. SPSS also provides the 95% confidence interval and standard error for both coefficients. The calculated t values for both coefficients are also provided, along with their precise level of statistical significance, indicating that $\hat{\beta}_1$ (slope) is statistically significant. The ANOVA regression model indicates that the regression model is statistically significant ($F = 109.956$, df = 1, 32, $p < 0.01$). The regression model for predicting charges from LOS for DRG 087 is

$$\hat{Y} = 1{,}441.467 = 2{,}360.879X$$

We will now look at several applications for linear regression.

✓ To Obtain Linear Regression Using SPSS:

- From the menus choose:
  →Analyze
    →Regression
      →Linear
- In the "Linear Regression" dialog box, select a numeric dependent variable.
- Select one or more numeric independent variables.

Exhibit 8–10 SPSS Output for Linear Regression, Length of Stay (LOS) and Total Charges, DRG 087

Descriptive Statistics

	Total Charges	*LOS*
Mean	13454.1765	5.0882
Std. Deviation	7305.20369	2.72327
N	34	34

Model Summary

Model	*R*	*R Square*	*Adjusted R Square*	*Std. Error of the Estimate*
1	.880[a]	.775	.768	3522.18201

[a] Predictors: (Constant), LOS

ANOVA[b]

Model		*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
1	Regression	1364093516.247	1	1364093516.247	109.956	.000[a]
	Residual	396984516.695	32	12405766.147		
	Total	1761078032.941	33			

[a] Predictors: (Constant), LOS
[b] Dependent Variable: Total Charges

Coefficients[a]

		Unstandardized Coefficients		*Standardized Coefficients*			*95% Confidence Interval for B*	
Model		*B*	*Std. Error*	*Beta*	*t*	*Sig.*	*Lower Bound*	*Upper Bound*
1	(Constant)	1441.467	1295.091		1.113	.274	−1196.547	4079.482
	LOS	2360.879	225.146	.880	10.486	.000	1902.273	2819.486

[a] Dependent Variable: Total Charges

Example 1: Predicting Cancer Deaths from Age

As researchers for a statewide cancer registry, we have been asked to build a regression model for predicting colon cancer survival time based on the age of the patient at time of diagnosis. The raw data for cancer survival time appear in Table 8–3. In building the regression model, we have designated "Age at Diagnosis" as the independent variable (*X*), and "Survival Time in Months" as the dependent variable (*Y*).

Exhibit 8–11 Excel Output for Linear Regression, Length of Stay and Total Charges, DRG 087

SUMMARY OUTPUT

Regression Statistics

Multiple *R*	0.880
R Square	0.775
Adjusted *R* Square	0.768
Standard Error	3522.182
Observations	34

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1364093516.25	1364093516.25	109.96	0.000
Residual	32	396984516.69	12405766.15		
Total	33	1761078032.94			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95%*	*Upper 95%*
Intercept	1441.467	1295.09	1.11	0.27	−1196.54	4079.48	−1196.54	4079.48
Length of Stay	2360.879221	225.15	10.49	0.00	1902.27	2819.49	1902.27	2819.49

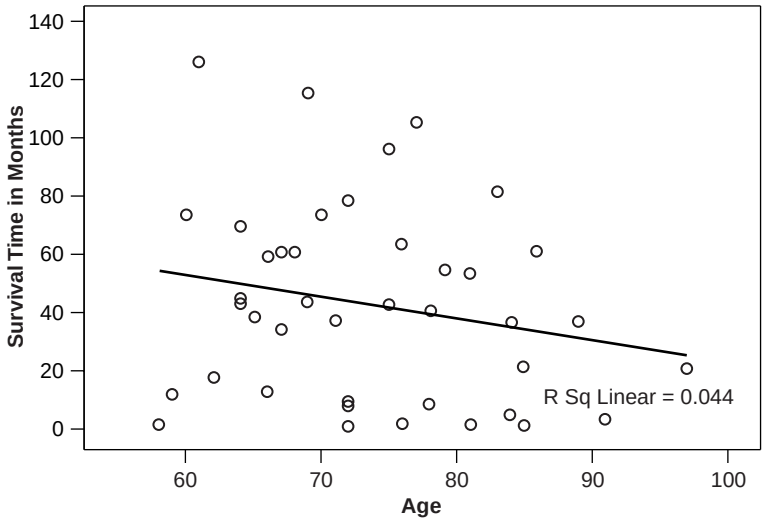
Table 8–3 Age at Diagnosis and Survival Time in Months, for Cases of Colon Cancer, Your Hospital, 20XX

Age at Diagnosis	*Survival Time in Months*	*Age at Diagnosis*	*Survival Time in Months*
61	126	89	36
78	8	75	96
69	115	84	36
62	17	64	69
77	105	72	78
81	53	67	60
81	1	60	73
83	81	70	73
72	1	76	63
85	21	86	60
58	1	66	59
64	43	69	43
68	60	85	1
79	54	76	1
75	42	64	44
78	40	65	38
67	34	71	37
97	20	72	9
72	8	91	3
84	4	59	11
66	12		

The first step is to prepare a scatter diagram, which appears in Figure 8–9. The plots are widely scattered, indicating that the relationship between age at diagnosis and survival time may not be linear and that the correlation between the two variables may be small. The regression line, which is somewhat flat, indicates a negative relationship between age and survival time.

Figure 8–9 Scatter Diagram—Age at Diagnosis of Colon Cancer and Survival Time in Months

Source: Data from *Self-Instructional Manual for Cancer Registries, Book 7: Statistics and Epidemiology for Cancer Registries,* p. 121, US Department of Health and Human Services, Public Health Service, National Institutes of Health, National Cancer Institute.



The null and alternative hypotheses are

$$H_0: \beta_1 = 0$$
$$H_A: \beta_1 \neq 0$$
$$\alpha = 0.05$$

Analysis of the SPSS output in Exhibit 8–12 does verify our suspicion that the relationship between the two variables is slight and that the relationship is negative and not statistically significant ($r = -0.210$, $p = 0.187$). The r^2 indicates that the age variable accounts for only 4% of the variance in survival time. The regression model (Exhibit 8–9) for predicting survival time from age at diagnosis is

$$Y = 97.063 - 0.743X$$

Exhibit 8–12 SPSS Output for Pearson *r* Correlation, Age at Diagnosis of Colon Cancer and Survival Time in Months

Correlations

		Age	*Survival Time in Months*
Age	Pearson Correlation	1	−.210
	Sig. (2-tailed)	.	.187
	N	41	41
Survival Time in Months	Pearson Correlation	−.210	1
	Sig. (2-tailed)	.187	.
	N	41	41

Source: Data from *Self-Instructional Manual for Cancer Registries, Book 7: Statistics Epidemiology for Cancer Registries*, p. 121, U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Cancer Institute.

As expected, the calculated t for the regression coefficient, β_1, is not significant ($t = -1.342$, $p = 0.187$). So we fail to reject the null hypothesis and conclude that the regression coefficient, β_1 equals zero. The ANOVA model (see Exhibit 8–13) also indicates that the regression model is not statistically significant. In this case, age at diagnosis is not an important indicator in predicting patient survival time. We therefore conclude that we cannot predict survival time from age at diagnosis. The β_1 coefficient indicates that there is not much change in Y associated with a change in X.

Exhibit 8–13 SPSS Output for Linear Regression, Age at Diagnosis of Colon Cancer and Survival Time in Months

Descriptive Statistics

	Mean	*Std. Deviation*	*N*
Survival Time in Months	42.3415	33.56681	41
Age	73.6098	9.48915	41

Model Summary

Model	*R*	*R Square*	*Adjusted R Square*	*Std. Error of the Estimate*
1	.210[a]	.044	.020	33.23527

[a] Predictors: (Constant), Age

ANOVA[b]

Model		*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
1	Regression	1990.474	1	1990.474	1.802	.187[a]
	Residual	43078.745	39	1104.583		
	Total	45069.220	40			

[a] Predictors: (Constant), Age
[b] Dependent Variable: Survival Time in Months

Coefficients[a]

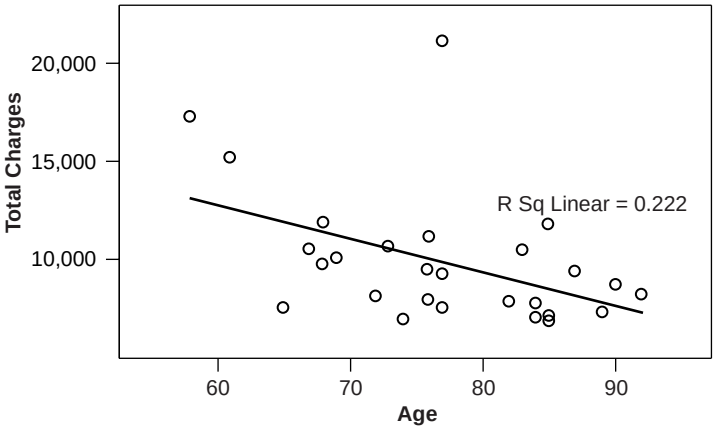
		Unstandardized Coefficients		*Standardized Coefficients*			*95% Confidence Interval for B*	
Model		*B*	*Std. Error*	*Beta*	*t*	*Sig.*	*Lower Bound*	*Upper Bound*
1	(Constant)	97.063	41.093		2.362	.023	13.944	180.182
	Age	−.743	.554	−.210	−1.342	.187	−1.864	.377

[a] Dependent Variable: Survival Time in Months

Example 2: Predicting Total Charges from Age

Earlier, we had an example where the total charges appeared to decrease with the age of the patient. We will now examine this phenomenon in more detail for DRG 336, Transurethral Prostatectomy with CC. The number of patients discharged from DRG 336 is 28. The descriptive statistics for total charges and age appear in Exhibit 8–1, and the scatter diagram with fitted regression line appears in Figure 8–10.

Figure 8–10 Scatter Diagram with Fitted Regression Line, Age and Total Charges



The Pearson *r* for age and total charges appears in Exhibit 8–14. The Pearson *r* is −0.471, $p = 0.011$. Even though the correlation appears to be moderate, age is accounting for only 22.2% of the variation in total charges. There must be something else at work that is causing this to occur. Another variable that affects total charges is LOS, as we have already demonstrated. The correlation matrix in Exhibit 8–14 provides the correlations between total charges and length of stay ($r = 0.838$) and total charges and age ($r = -0.471$).

Exhibit 8–14 SPSS Output for Pearson *r* Correlation Coefficients for Total Charges, Length of Stay (LOS), DRG 336

Correlations

		Total Charges	*AGE*	*LOS*
Total Charges	Pearson Correlation	1	−.471*	.838**
	Sig. (2-tailed)	.	.011	.000
	N	28	28	28
AGE	Pearson Correlation	−.471*	1	−.310
	Sig. (2-tailed)	.011	.	.108
	N	28	28	28
LOS	Pearson Correlation	.838**	−.310	1
	Sig. (2-tailed)	.000	.108	.
	N	28	28	28

*Correlation is significant at the 0.05 level (2-tailed).
**Correlation is significant at the 0.01 level (2-tailed).

For DRG 336, there are two third-party payers—Medicare and commercial. Since "Payer" is a nominal-level variable, we cannot correlate payer with total charges, age, or LOS using the Pearson *r*. However, if we compare the mean total charges, age, and LOS by third-party payer, we find the observed differences between the means to be statistically significant (Table 8–4). Substituting eta for the Pearson *r*, we find that the relationships between LOS and payer and between total charges and payer are moderate, with etas of 0.428 and 0.538, respectively. However, there is a strong relationship between payer and age, eta = 0.708. Thus, "Third-Party Payer" may be a confounding variable explaining why total charges decrease as age increases. In Table 8–4, we can see that the ALOS for Medicare patients, 1.82 days, is less than the ALOS for patients in the commercial payer category, 3.27 days.

Table 8–4 Mean Age, Length of Stay (LOS), and Total Charges by Third-Party Payer, DRG 336

	Medicare Mean	*Commercial Mean*	*F*	*p*	*eta*
Age	82.5	69.5	26.16	<.001	.708
LOS	1.82	3.27	5.82	.023	.428
Total Charges	$8,389.35	$11,989.64	10.61	.003	.538

Since these variables, including payer, correlate well with total charges, we can develop a simple regression model for each of these independent variables separately. (Linear regression allows us to use nominal-level variables in which there are two categories.) The regression models for LOS, age, and payer appear in Exhibits 8–15 through Exhibit 8–17. Each of the three models below is statistically significant:

Predictor	*Model*	r^2
Length of Stay	$Y = 5{,}845.578 + 1.654.162\text{I}$	0.702
Age	$Y = 23{,}111 + (-171.8781X)$	0.222
Payer	$Y = 7{,}489.282 + 900.071X$	0.290

Exhibit 8–15 SPSS Regression Statistics for Total Charges and Length of Stay (LOS), DRG 336

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.838[a]	.702	.691	1850.07943

[a] Predictors: (Constant), LOS

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	209811792.289	1	9209811792.289	61.298	.000[a]
	Residual	88992640.961	26	3422793.883		
	Total	298804433.250	27			

[a] Predictors: (Constant), LOS
[b] Dependent Variable: Total Charges

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	5845.578	614.679		9.510	.000	4582.086	7109.069
	Payer	1654.162	211.278	.838	7.829	.000	1219.874	2088.449

[a] Dependent Variable: Total Charges

The strongest model is the one produced by the independent variable, LOS. This is evidenced by the β_1 coefficient and the r^2, which indicates that 70.2% of the variation in total charges is accounted for by LOS. Does this mean that we can sum the r^2 for the other two variables to determine the total amount of variation in total charges? The answer is no. This is because each variable was analyzed separately—not in relation to how they act together. In addition, by looking at each variable separately, we increase the probability of making a type I error.

Exhibit 8–16 SPSS Regression Statistics for Total Charges and Age, DRG 336

Model Summary

Model	*R*	*R Square*	*Adjusted R Square*	*Std. Error of the Estimate*
1	.471[a]	.222	.192	2990.28523

[a] Predictors: (Constant), AGE

ANOVA[b]

Model		*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
1	Regression	66317483.834	1	66317483.834	7.417	.011[a]
	Residual	232486949.416	26	8941805.747		
	Total	298804433.250	27			

[a] Predictors: (Constant), AGE
[b] Dependent Variable: Total Charges

Coefficients[a]

		Unstandardized Coefficients		*Standardized Coefficients*			*95% Confidence Interval for B*	
Model		*B*	*Std. Error*	*Beta*	*t*	*Sig.*	*Lower Bound*	*Upper Bound*
1	(Constant)	23111.999	4919.310		4.698	.000	13000.213	33223.785
	Payer	−171.8778	63.113	−.471	2.723	.011	−301.608	−42.147

[a] Dependent Variable: Total Charges

Adjusted r^2 in SPSS:
The sample r^2 tends to optimistically estimate how well the models fit the population. The model usually does not fit the population as well as it fits the sample from which it is derived. Adjusted r^2 attempts to correct r^2 to more closely reflect the goodness of fit of the model in the population.

Source: Data from SPSS 12.0 for Windows, Copyright SPSS Inc. 2003, Chicago, Illinois, USA.

Exhibit 8–17 SPSS Regression Statistics for Total Charges and Payer, DRG 336

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.538[a]	.290	.262	2857.08646

[a] Predictors: (Constant), Payer

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	86567914.822	1	86567914.822	10.605	.003[a]
	Residual	212236518.428	26	8162943.016		
	Total	298804433.250	27			

[a] Predictors: (Constant), Payer
[b] Dependent Variable: Total Charges

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	7489.282	892.553		8.391	.000	5654.613	9323.951
	Payer	900.071	276.390	.538	3.257	.003	331.944	1468.198

[a] Dependent Variable: Total Charges

To determine the effect of the three independent variables together on the dependent variable, total charges, we can develop a **multiple regression** model. This discussion will serve only as a brief introduction to multiple regression. Basically, in multiple regression we are incorporating more than one independent variable into the model. This procedure usually provides a fuller explanation of the effects on the dependent variable. Also, the effect of a single variable is made more certain. The multiple regression model is an extension of the bivariate model and is symbolized as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots + \beta_k X_k + e$$

The interpretation of the constant in the above model is the same as that for simple regression—the average value of Y when each of the independent values equals zero. The interpretation for slope is slightly different. In the multiple-regression situation, the slope is interpreted as the average change in Y associated with a unit change in X when the other in-

dependent variables are held constant. Thus, we are able to separate out the effect of any independent variable (X_k) from any distorting influences of the other independent variables. This is sometimes referred to as partial slope or partial regression coefficient.

Now let us develop the regression model using the three independent variables together. SPSS provides several methods for entering the independent variables into the regression model: enter, where all variables appear in the model whether or not they are significant, and forward, backward, and stepwise. In the other methods, a variable is entered into the model on the basis of certain entry or removal criteria. We will use the stepwise multiple-regression procedure, which examines each variable in the model in a block (together) at each step for entry or removal of each independent variable. The stepwise method produces two regression models—one with LOS as the sole predictor variable and the other adding age as a second predictor variable. "Payer" does not enter the regression model.

Upon examination of the two models (Exhibit 8–18), we see that in model 1, the variable LOS alone accounts for 70.2% of the variance in total charges. If we add age to the model, the amount of variance accounted for by the two variables together improves to 75.1%. The model may help explain the variation in charges, but it still does not help us answer why charges decrease with age. This illustrates a problem that occurs with multiple regression: two of the independent variables are highly correlated with one another. In this example, payer and age are highly correlated (eta = 0.708). Since payer is highly correlated with age, the resultant statistics in the regression model become unstable. When two independent variables are highly correlated with each other, a phenomenon often termed **multicollinearity**, it is difficult to separate the effects of each independent variable on the dependent variable.

In this DRG, there are only two payers—Medicare and commercial PPO. The Medicare group has a smaller LOS (Table 8–5), and hence lower total charges than the commercial payer, even though the average age of Medicare patients is 82.53 versus 69.55 for the commercial payer group. Even though the payer variable does not appear in the multiple regression model, this could be an explanation for the decrease in total charges by either age or LOS because these two variables are so closely related to the payer variable. This may also suggest that the third-party payer has a strong influence on a patient's LOS.

Exhibit 8–18 SPSS Regression Statistics, Total Charges from Length of Stay, Age and Payer, DRG 336

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	.838[a]	.702	.691	1850.07943	.702	61.298	1	26	.000
2	.867[b]	.751	.732	1723.40382	.049	4.963	1	25	.035

[a] Predictors: (Constant), LOS
[b] Predictors: (Constant), LOS, AGE

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	209811792.289	1	209811792.289	61.298	.000[a]
	Residual	88992640.961	26	8162943.016		
	Total	298804433.250	27			
2	Regression	224551415.172	2	112275707.586	37.802	.000[b]
	Residual	74253018.078	25	2970120.723		
	Total	298804433.250	27			

[a] Predictors: (Constant), LOS
[b] Predictors: (Constant), LOS, AGE
[c] Dependent Variable: Total Charges

Coefficients[a]

Model		Unstandardized Coefficients: B	Unstandardized Coefficients: Std. Error	Standardized Coefficients: Beta	t	Sig.	95% Confidence Interval for B: Lower Bound	95% Confidence Interval for B: Upper Bound
1	(Constant)	5845.578	614.679		9.510	.000	4582.086	7109.069
	LOS	1654.162	211.278	.838	7.829	.000	1219.874	2088.449
2	(Constant)	12787.664	3168.430		4.036	.000	6262.161	19313.167
	LOS	1511.082	207.026	.765	7.299	.000	1084.704	1937.431
	AGE	−85.236	38.262	−.234	2.228	.035	−164.038	−6.434

[a] Dependent Variable: Total Charges

Excluded Variables[c]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics: Tolerance
1	AGE	−.234[a]	−.2.228	.035	−.407	.904
	Payer	.220[b]	1.958	.062	.365	.817
2	Payer	.101[c]	.671	.509	.136	.451

[a] Predictors in the Model: (Constant), LOS
[b] Predictors: in the Model: (Constant), LOS, AGE
[c] Dependent Variable: Total Charges

Table 8–5 SPSS Output of Means of Total Charges and Length of Stay (LOS) by Payer, DRG 336

Payer		*Total Charges*	*AGE*	*LOS*
Medicare	Mean	8389.3529	82.5294	1.8235
	N	17	17	17
	Std. Deviation	1376.61351	6.60604	.95101
Commercial PPO	Mean	11989.6364	69.5455	3.2727
	N	11	11	11
	Std. Deviation	4265.15513	6.48635	2.19504
Total	Mean	9803.7500	77.4286	2.3929
	N	28	28	28
	Std. Deviation	3326.68467	9.11827	1.68521

CONCLUSION

In this chapter, we explored correlation and linear regression. We can use the Pearson *r* correlation coefficient to determine whether there is a relationship between two variables, *X* and *Y*. For the Pearson *r*, the variables must be at least at the interval scale of measurement; furthermore, it is assumed that the relationship between the two variables is linear. The Pearson *r* correlation coefficient ranges from −1.00, indicating a perfect negative correlation, to +1.00, indicating a perfect positive correlation. A correlation coefficient of 0.00 indicates no linear relationship. Prior to calculating the Pearson *r*, a scatter diagram should be constructed to determine if the relationship between the two variables is linear.

If two variables have a high correlation with one another and the relationship is linear, it may be possible to predict one (*Y*) from our knowledge of the other (*X*). When predicting *Y* from *X*, we are constructing a regression model. The components of the regression model are $Y = a + BX$ where *a* is the intercept and *B* is the regression coefficient. The intercept is the average of *Y* when *X* is equal to zero; the regression coefficient represents the average change in *Y* that is associated with a unit change in *X*. In making predictions from regression models, we must be careful to limit the range of *X* to the observations that were used to develop the model.

In both the Pearson *r* correlation and linear regression, we can request the coefficient of determination, r^2. The r^2 can help us determine the power of the model; it tells us how much of the variation in the dependent variable can be explained by the independent variable.

ADDITIONAL RESOURCES

Lewis-Beck, M.S. 1986. *Applied regression: An introduction.* Beverly Hills, CA: Sage Publications.

Appendix 8–A

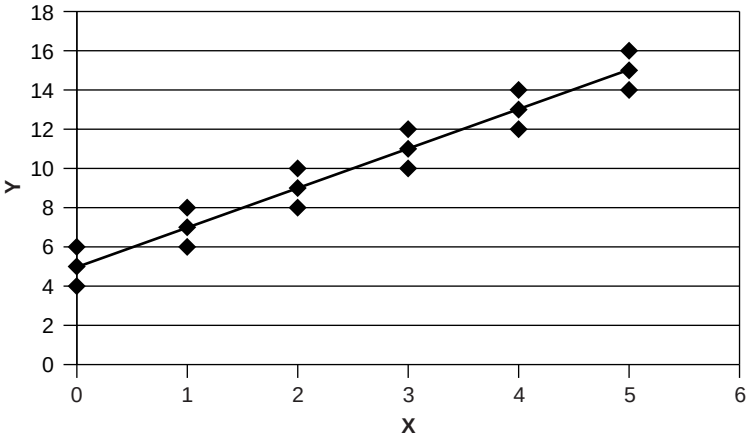
Exercises for Solving Problems

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
2. What is the Pearson *r*? At what level(s) of measurement must the variables be in order to use the Pearson *r*?
3. What is the range for the Pearson *r*? How is the Pearson *r* statistic interpreted? Explain the concepts of positive linear relationship and negative linear relationship.
4. Describe the relationship between the Pearson *r* and the coefficient of determination, r^2.
5. What is the interpretation of the regression line in a scatter diagram?
6. Explain simple linear regression and multiple regression. Give an example for each.

MULTIPLE CHOICE

1. To calculate the Pearson *r*, the two variables should be:
   a. normally distributed
   b. linearly related
   c. random
   d. all of the above
2. In the scatter diagram below, the Pearson *r* can be described as:
   a. equal to 0.0
   b. equal to +1.0
   c. negative
   d. positive



3. If the Pearson *r* is equal to +1.0, we can say that:
   a. *x* causes *y*
   b. *y* causes *x*
   c. there is a perfect positive relationship between *x* and *y*
   d. *x* and *y* are negatively correlated

4. Which of the following values for the Pearson *r* indicates the strongest relationship?
   a. +0.85
   b. +0.76
   c. 0.0
   d. −0.89

5. Which of the following values for the Pearson *r* indicates the weakest relationship?
   a. +0.85
   b. +0.76
   c. 0.0
   d. −0.89

6. If we cannot predict *y* from *x*, we would conclude that the Pearson *r* is:
   a. 0.0
   b. negative
   c. positive
   d. cannot determine from information provided

7. We have calculated a Pearson *r* for length of stay (*Y*) and age (*X*). Our result is $r = -1.16$.We therefore conclude that:
   a. as age increases, the length of stay decreases
   b. as age increases, the length of stay increases
   c. we cannot predict length of stay from age
   d. we have made an error in our calculations

8. We have calculated the Pearson r for length of stay (X) and total charges (Y). Our result is $r = +0.64$. We therefore conclude that:
   a. as length of stay increases, total charges decrease
   b. as length of stay increases, total charges increase
   c. we cannot predict total charges from length of stay
   d. we have made an error in our calculations

PROBLEMS

1. You are studying DRG 105, Cardiac Valve Procedures and Other Major Cardiothoracic Procedures without Cardiac Catheterization, for Critical Care Hospital. Using the data provided in Table 8–A–1, calculate the Pearson r for each of the following pairs:
   - Age and length of stay
   - Total charges and length of stay
   - Age and total charges

   a. State the null and alternative hypotheses and alpha level for each.
   b. Construct a scatter diagram with regression line for each.
   c. State your conclusions for each.

2. Construct a regression model for predicting total charges from length of stay for DRG 105.
   a. State the null and alternative hypotheses and alpha level.
   b. Prepare a scatter diagram with the regression line for the two variables.
   c. What are the r and r^2? What is the importance of the r and r^2 results?
   d. What is the regression equation?
   e. What are your conclusions?

Table 8–A–1 Case Summaries for DRG 105, Cardiac Valve Procedures and Other Major Cardiothoracic Procedures without Cardiac Catheterization

	Gender	*Age*	*LOS*	*Charges*	*Payor*
1	Female	47	20	$91,683	Medicaid
2	Female	75	43	$93,708	Medicare
3	Female	84	7	$21,446	Medicare
4	Female	50	13	$37,797	Medicare
5	Male	77	14	$54,364	Medicare
6	Male	57	4	$17,626	Medicare
7	Male	73	4	$12,832	Medicare
8	Female	56	1	$36,153	Medicaid
9	Male	69	1	$14,907	Medicaid
10	Female	81	23	$104,148	Medicare
11	Male	21	5	$21,423	Medicaid
12	Female	37	5	$24,971	Medicaid
13	Female	69	4	$17,022	Medicare
14	Female	89	17	$50,652	Medicare
15	Male	28	35	$186,496	Medicaid
16	Male	47	6	$24,441	Medicaid
17	Male	87	11	$35,349	Medicare
18	Female	85	5	$22,155	Medicare
19	Male	56	5	$24,455	Managed Care
20	Male	45	11	$36,401	Medicaid
21	Male	82	6	$25,783	Medicare
22	Female	65	10	$37,055	Managed Care
23	Male	67	4	$19,236	Medicare
24	Male	59	23	$60,132	Other
25	Female	67	7	$35,777	Medicare
26	Male	53	4	$19,972	Managed Care
27	Male	71	7	$25,409	Medicare
28	Female	79	6	$281,140	Medicare
29	Male	63	1	$41,283	Medicaid
30	Male	53	19	$71,439	Medicaid
31	Female	75	9	$33,735	Medicare
32	Female	68	9	$37,830	Gov Mngd Care
33	Male	37	4	$22,311	Medicaid
Total *N*	33	33	33	33	33

CHAPTER 9

Chi-Square

KEY TERMS

Contingency tables
χ^2 test of independence
  Standardized residuals
  Yates correction for continuity
  Phi coefficient
  Contingency coefficient
  Cramer's V
  Fisher's exact test
χ^2 goodness-of-fit test
McNemar test

LEARNING OBJECTIVES

Upon completion of this chapter, you should be able to:

1. Define key terms
2. Differentiate between parametric and nonparametric statistical procedures.
3. Outline the assumptions for nonparametric procedures.
4. Conduct the χ^2 test of independence for given situations.
5. Explain the concept of standardized residuals.
6. Analyze results of the χ^2 test of independence using standardized residuals.
7. Conduct the McNemar test for given situations.
8. Use microcomputer statistical software to solve nonparametric problems.

The statistical procedures discussed thus far are called *parametric statistics*. Parametric statistical procedures require certain assumptions about the underlying population, most particularly that the underlying population is normally distributed, that measures are at the interval or ratio level of measurement, and that the samples are randomly drawn and independent. In

contrast, nonparametric procedures have less restrictive assumptions. There are no assumptions that the underlying population distribution is normal, and the distributions may take on any shape, so they are not limited to the bell shape of the normal distribution. Thus, nonparametric statistical procedures are often referred to as *distribution-free statistics*.

Nonparametric procedures may be used to analyze data about populations that consist of nominal, ordinal, interval, and/or ratio data. Many nonparametric tests analyze the ranks or orders of the data set rather than the numerical values of the observations. Nonparametric procedures also are used when sample sizes are small (≤ 30) or when there are extreme values in the data set so that the median rather than the mean is more representative of the distribution. Also, when the shape of the distributions is either unknown or non-normal, nonparametric procedures may be more powerful, reducing the chance of committing a type II error.

Nonparametric methods may be used for testing hypotheses regarding

- relationships between variables
- relationships between variables in paired samples
- relationships between variables in two independent samples
- relationships between variables in three or more independent samples

There are, however, disadvantages associated with the use of nonparametric tests, especially when it is possible to use a corresponding parametric procedure. Since nonparametric procedures involve testing orders or ranks rather than interval/ratio data that are continuous, some information is lost. Consequently, we have a conflict because nonparametric procedures may be considered less powerful than their parametric counterparts, also increasing the probability of committing a type II error. It is therefore imperative that the researcher select the appropriate test in view of the level of measurement used in the data collection process and the sample size. Table 9–1 provides a listing of tests that may be used in relation to the variables' levels of measurement.

Table 9–1 Parametric/Nonparametric Procedures by Level of Measurement

	Variable 1		
Variable 2	*Nominal*	*Ordinal*	*Interval/Ratio*
Nominal	χ^2 test Fisher exact test McNemar χ^2 (paired) Cramer's V Kappa coefficient		
Ordinal	Sign test Wilcoxon signed ranks test (paired) Mann-Whitney Wilcoxon test Kruskal-Wallis test	Spearman rho	
Interval/ratio	Student's paired *t* test Student's *t* test for two independent samples One-way ANOVA	Spearman rho	Pearson *r* Simple regression

The power of nonparametric tests can be improved by increasing the sample size. Less powerful tests are less likely to detect a small difference between groups when one exists. If it is important to the researcher to detect small differences, a parametric procedure should be used if possible. Parametric and nonparametric procedures are compared in Table 9–2.

In this chapter, we will discuss the various forms of the chi-square test. In Chapter 10, we review other nonparametric procedures.

Table 9–2 Comparison of Parametric and Nonparametric Procedures

Parametric Procedures	*Nonparametric Procedures*
For Two Independent Samples or Paired Samples	
Student's *t* for two independent samples	Fisher's exact test
ANOVA	Mann-Whitney Wilcoxon *U* test
	χ^2 test for two independent samples
For Three or More Independent Samples	
ANOVA	Kruskal-Wallis analysis of ranks
	χ^2 test for *k* independent samples
For Relationships Between Variables	
Pearson *r* (bivariate)	Spearman rho (bivariate)
	χ^2 test
	Cramer's V
	Phi coefficient
	Fisher's exact test

CHI-SQUARE (χ^2) TESTS

One of the more commonly used nonparametric tests is the chi-square (χ^2) test. The χ^2 distribution is a positive distribution based on the number of degrees of freedom. There is more than one type of χ^2 test that can be used to analyze frequency data; however, all are based on comparing actual observations with expected frequencies. With frequency data, we are reporting the percentages or frequencies of an independent variable on the characteristic of interest. **Contingency tables** are used to display frequency data. A two-by-two (2×2) contingency table is the simplest form.

In a 2×2 table, the distribution of one variable is conditionally dependent, or contingent, upon the other. The table is made up of cells, which are specific locations in the matrix created by the two variables under study. Each cell represents the joint frequencies for the categories on each variable. The categories that make up each variable are dichotomous and must be mutually exclusive. The sums of the rows and columns are placed in the margins and are thus called marginal frequencies. The total for the row and column marginals is the cell in the lower right-hand corner (grand total); this sum is equal to *N*. The basic shell for a 2×2 contingency table is presented in Exhibit 9–1.

Exhibit 9–1 Shell for a 2 × 2 Contingency Table

		Variable 1		
		Category 1	*Category 2*	*Total*
Variable 2	Category 1	*a*	*b*	*a* + *b*
	Category 2	*c*	*d*	*c* + *d*
		a + *c*	*b* + *d*	*a* + *b* + *c* + *d*

Source: Adapted from *Principles of Epidemiology: An Introduction to Applied Epidemiology and Biostatistics*, p. 210, 1992, U.S. Department of Health and Human Services, Public Health Service.

Contingency tables may be larger than 2 × 2. If there are more than two categories for a given variable, the table is referred to as an R × C table (R = rows, C = columns). We will now use hypothetical data collected to construct a two-by-two contingency table. A survey was conducted in which data were collected on two variables: professional credential and geographic location. "Credential" had two categories, Registered Health Information Technician (RHIT) and Registered Health Information Administrator (RHIA), and geographic location was classified into two categories: urban/suburban and rural. The question of interest was whether there is a relationship between professional credential and geographic practice location—urban/suburban or rural. The results of the classification appear in Table 9–3. When we read a 2 × 2 contingency table or R × C table, the percentages that appear in the cells represent the percentages of the column totals.

We will now conduct the χ^2 test of independence on these data.

Table 9–3 Contingency Table of Professional Credential and Geographic Location

	Credential		
Location	*RHIT*	*RHIA*	*Total*
Urban/Suburban	30 (35.3%)	76 (67.3%)	106 (53.5%)
Rural	55 (64.7%)	37 (32.7%)	92 (46.5%)
Total	85	113	198

THE χ^2 TEST OF INDEPENDENCE

In the **χ^2 test of independence**, we can determine whether a relationship exists between two variables in a 2 × 2 contingency table or R × C table. It is one of the most widely used statistics in health care. The question that we are trying to answer is whether the categories of the row variable are distributed differentially across the column variables—that is, how one

variable relates to another. If the variables are independent, a change in one variable does not have any effect on the other. We will use the data in Table 9–3 to calculate χ^2.

The null hypothesis is that the two variables, credential and geographic location, are independent of each other—in other words, that there is no relationship between the two variables. If the null is true, we would expect that geographic location would appear in the same proportions in both the ART population and the RRA population. The alternative hypothesis is nondirectional and states that there is an association between the two variables of interest.

H_0: There is no association between credential and geographic location.

H_A: There is an association between credential and geographic location.

$\alpha = 0.05$

Exhibit 9–2 summarizes the important aspects of the χ^2 test of independence.

Exhibit 9–2 χ^2 Test of Independence—Points to Remember

1. The χ^2 statistic is used when both the independent variable and the dependent variable are at the nominal level of measurement and the categories of each variable are mutually exclusive.
2. Data contained in the contingency table or R × C table are frequencies, not scores.
3. Observations must be independent of each other (except for McNemar's χ^2).
4. The total number of observations (*N*) should be greater than 20, and the expected frequency per cell should be equal to or greater than 5. For large tables, 20% of the expected frequencies may be less than 5, but should not be less than 2.

The χ^2 test involves comparing observed frequencies with expected frequencies. In a contingency table, the expected frequencies are the proportion of each category that we would expect to find in each cell by chance over the long run. In our example, there are 198 total observations. Of these 198, 106, or 53.5%, practice in urban/suburban areas, and 92, or 46.5%, practice in rural areas. Therefore, in the long run, we would expect these frequencies to occur in each group. In our contingency table, there are 85 RHITs and 113 RHIAs. Therefore, we would expect that 53.5% in each group would be practicing in urban/suburban locations—45.5 RHITs and 60.5 RHIAs. The expected frequencies are generated on this assumption. An easy way to generate the expected frequencies is given by the following formula:

$$\text{Expected frequency} = \frac{\text{row marginal} \times \text{column marginal}}{\text{grand total } (N)}$$

The next step is to subtract the expected cell frequency (E) from the observed cell frequency (O). This gives the amount of deviation in each cell. The sum of the deviates for both rows and columns is equal to zero. The deviates in each cell are then squared, $(O - E)^2$, and divided by the expected value for each cell, $(O - E)^2/E$. This is similar to the numerator of the variance, $(X - \overline{X})^2$, where $\overline{X}$ is the expected value. But where the denominator for the variance is divided by the degrees of freedom, $n - 1$, the denominator for χ^2 is the expected (E) frequency. Therefore, the basic statistical method for measuring variation in a data set, the total sum of squares (TSS), is rewritten for χ^2 as the sum of $(O - E)^2$.

The formula for calculating the χ^2 statistic is:

$$\chi^2_{\text{calc}} = \sum[(O - E)^2]/E$$

which is defined as the sum of the squared deviations divided by the expected frequency for each cell. To calculate the expected frequency for the first cell, RHIT × urban/suburban, we multiply the column marginal by the row marginal (Table 9–3) and divide by the grand total: $(85 \times 106)/198 = 45.5$. We proceed in the same manner for each cell. The expected frequency for each cell is subtracted from the observed frequency, squared, and divided by the expected frequency. The results for each cell are then summed. For the data in Table 9–4, χ^2 is calculated as:

$$\begin{aligned}\chi^2_{\text{calc}} &= \sum[(O - E)^2]/E \\ &= 5.28 + 6.08 + 3.97 + 4.58 \\ &= 19.91\end{aligned}$$

Table 9–4 χ^2 Test of Independence for Credential and Geographic Location

	Credential					
	RHIT			*RHIA*		
Location	*O*	*E*	$(O - E)^2$	*O*	*E*	$(O - E)^2$
Urban/Suburban	30	45.5	5.28	76	60.5	3.97
Rural	55	39.5	6.08	37	52.5	4.58
Total	85	85		113	113	

In our example, the calculated χ^2 is fairly large, indicating that the observed frequencies do differ markedly from the theoretical or expected frequencies. This is verified by comparing the calculated χ^2 with the critical value of χ^2 in Appendix B, Table B–6. In a 2×2 contingency table, degrees of freedom are equal to 1 (df = 1). In larger tables, the number of degrees of freedom is determined by

$$\text{df} = (\text{R} - 1) \times (\text{C} - 1).$$

Conceptual models for degrees of freedom in contingency tables and R × C tables are displayed in Exhibit 9–3.

Exhibit 9–3 Degrees of Freedom

In a 2 × 2 contingency table, the best estimate of the expected counts in a distribution is given in the row and column totals. Therefore, the row and column totals are fixed. Once an observed count is entered into a cell in a 2 × 2 table, no other cells are free to vary.

	Column 1	*Column 2*	*Total*
Row 1	Degree of freedom	Fixed	Fixed
	a	b	$a + b$
Row 2	Fixed	Fixed	Fixed
	c	d	$c + d$
Total	Fixed	Fixed	Fixed
	$a + c$	$b + d$	$a + b + c + d$ (N)

We can extend this idea to R × C tables. As above, the row and column totals are fixed. Now assume that the last column and the bottom row are never free to vary because they must consist of numbers to make the totals come out right. This is illustrated in the 4 × 3 table below (4 rows, 3 columns):

	Column 1	*Column 2*	*Column 3*	*Total*
Row 1	Degree of freedom	Degree of freedom	Fixed	Fixed row total
Row 2	Degree of freedom	Degree of freedom	Fixed	Fixed row total
Row 3	Degree of freedom	Degree of freedom	Fixed	Fixed row total
Row 4	Fixed	Fixed	Fixed	Fixed row total
Total	Fixed column total	Fixed column total	Fixed column total	Fixed grand total

From Table B–6 in Appendix B, the χ^2 critical for $\alpha = 0.05$ and for one degree of freedom is 3.841. Since the calculated χ^2 is greater than the critical χ^2, we reject the null and conclude that the variables, credential and geographic location, are related or that they vary together in some systematic way.

There are several points to remember when calculating χ^2 from a contingency table or the larger R × C table. First, the expected counts for each cell should be 5 or more. For larger tables, 20% of the expected counts could be less than 5 but should not be less than 2. When expected cell counts are less than 5, consider collapsing the number of categories.

The χ^2 test of independence is an example of statistical modeling. Statistical modeling is a process by which a model is developed to predict the relationship of one or more dependent variables with an independent variable. We have already had experience with statistical modeling with linear regression. On the basis of our results in the previous example, we would expect that for any sample drawn from a population of RHITs and RHIAs, RHIAs would be more likely to practice in urban areas than RHITs.

EXAMINATION OF RESIDUALS

A large χ^2 often results in statistical significance, indicating an association between the two variables under study. However, this does not always tell us which levels of the variable are contributing the most to the χ^2 statistic. By examining the χ^2 residuals, we can determine which cells are contributing the most to the calculated χ^2.

A residual is defined as the difference between the actual frequencies and the expected frequencies in each cell ($O - E$). SPSS provides unstandardized, standardized, and adjusted residuals for each cell. We most often use the **standardized residuals**, which are obtained by

$$\text{Standardized residual} = \frac{O - E}{\sqrt{E}}.$$

If the value of the obtained standardized residual is greater than $+2$ or less than -2 in any cell, we can conclude that the cell in question is an important contributor to the significance of χ^2. In the SPSS output in Exhibit 9–4, the RHIT cells for geographic location are contributing the most to the calculated χ^2, indicating that the observed frequencies deviate most from the expected frequencies.

YATES CORRECTION FOR CONTINUITY

Because the χ^2 test is based on comparing the calculated values of χ^2, which form a discontinuous distribution with the theoretical values of χ^2, which in turn forms a continuous distribution, many recommend that the Yates continuity correction be used when the table numbers are small and df = 1. The only difference in the χ^2 formula is that the 0.5 is subtracted from the absolute value of $(O - E)$ in each cell.

$$\text{Yates } \chi^2 = \sum [(|O - E| - 0.5)^2]/E$$

Obviously, this will reduce the size of the calculated χ^2, thus reducing the chance of finding statistical significance and increasing the probability of making a type II error—failing to reject the null hypothesis when it is false. Recalculating our credential and geographic location data using the **Yates correction for continuity** reduces chi-square slightly, $\chi^2 = 18.66$, but it still remains statistically significant.

PHI COEFFICIENT

Chi-square tells us if there is an association between two variables, but it does not tell us the degree of association, as does the Pearson r correlation coefficient. Also, we cannot compare values of χ^2 across samples because the calculated χ^2 is a function of sample size. The **phi coefficient** corrects for sample size and measures the degree of association between the two variables under study. The formula for phi is

$$\phi = \sqrt{\chi^2/n}.$$

Exhibit 9–4 SPSS Output for Chi-Square Procedure, Credential and Geographic Location

CREDENTIAL * GEOGRAPHIC LOCATION Crosstabulation

			GEORAPHIC LOCATION		
			URBAN/ SUBURBAN	*RURAL*	*Total*
CREDENTIAL	RHIT	Count	30	55	85
		Expected Count	45.5	39.5	85.0
		Std. Residual	−2.3	2.5	
	RHIA	Count	76	37	113
		Expected Count	60.5	52.5	113.0
		Std. Residual	2.0	−2.1	
Total		Count	106	92	198
		Expected Count	106.0	92.0	198.0

Chi-Square Tests

	Value	*df*	*Asymp. Sig. (2-sided)*	*Exact Sig. (2-sided)*	*Exact Sig. (1-sided)*
Pearson Chi-Square	19.923[b]	1	.000		
Continuity Correction[a]	18.659	1	.000		
Likelihood Ratio	20.213	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	19.8222	1	.000		
N of Valid Cases	198				

[a] Computed only for 2 × 2 table
[b] 0 cells (.0%) have expected count less than 5. The minimum expected count is 39.

Symmetric Measures

		Value	*Approx. Sig.*
Nominal by Nominal	Phi	−.317	.000
	Cramer's V	.317	.000
	Contingency Coefficient	.302	.000
N of Valid Cases		198	

[a] Not assuming the null hypothesis.
[b] Using the asymptotic standard error assuming the null hypothesis.

The phi coefficient is interpreted in the same way as the Pearson *r*, but the range of phi is 0 to1. As a general rule of thumb, a value less than 0.30 may be interpreted as a trivial association. The phi coefficient for the data in Table 9–4 is calculated as

$$\phi = \sqrt{\chi^2/n}$$
$$\phi = \sqrt{19.91/198}$$
$$= 0.317.$$

Therefore, the relationship between credential and geographic location is not very strong. We can see that phi provides us with more knowledge about our sample profile than χ^2 alone.

CONTINGENCY COEFFICIENT

The **contingency coefficient** (*C*) is an alternative to the phi coefficient when one dimension of the contingency table is greater than two ($2 \times k$). The contingency coefficient is calculated as

$$C = \sqrt{\chi^2/\chi^2 + N}$$

The value of the contingency coefficient ranges from 0 to 1.0. However, its maximum value depends on the number of rows and columns in the table.

CRAMER'S V

The phi coefficient is a useful statistic when working with large samples where statistical significance can be easily achieved. Phi adjusts for sample size and can be considered the correlation coefficient for data in 2×2 tables. **Cramer's V** is used when the table is larger than 2×2.

$$\text{Cramer's V} = \sqrt{\chi^2/N(m-1)}$$

where *m* is the smaller of the number of the rows or columns in the R × C table. The range for Cramer's V is 0 to 1.0, and the significance level will be the same as that for the χ^2.

SPSS will perform the procedures described: chi-square, Yates correction for continuity, phi coefficient, and Cramer's V. When categorical variables are used, they must be coded for correct entry onto the data sheet. For the variables in Table 9–3, credential was coded as RHIT = 1 and RHIA = 2. Geographic location was coded as urban/suburban = 1 and rural = 2.

✓ To Obtain a Chi-Square Test Using SPSS:

- From the "Analyze" menu, choose:
  →Descriptive statistics
  →crosstabs
- Select one or more row variables and one or more column variables.
  →Click "Statistics" for tests and measures of association (chi-square, phi coefficient, Cramer's V, etc.).
  →Click "Cells" for observed and expected values, percentages, and residuals.

The results of the SPSS calculations appear in Exhibit 9–4. The Pearson χ^2 provided by SPSS is 19.923, and the continuity correction is 18.659. (Rounding accounts for the differences between our hand-calculated results and the SPSS results.) SPSS provides the same results for the phi coefficient and Cramer's V. Note that SPSS reports the phi coefficient as a negative value—an idiosyncrasy of the program, as a square root cannot be negative. The standardized residuals indicate that the RHIT cells are contributing the most to the calculated χ^2 statistic.

The likelihood ratio, **Fisher's exact test**, and "Linear × Linear Association" appear by default. The likelihood ratio is a goodness-of-fit statistic similar to the χ^2 goodness of fit, which we will discuss later in this chapter. For large sample sizes, they are identical. The advantage of the likelihood ratio is that it can be subdivided into interpretable parts that sum to the total. The significance level of the likelihood ratio is of more interest than the actual value of the statistic itself.

The χ^2 test for linear association tells us whether the relationship between the two variables under study is a linear one. That is, the observations on both variables increase in the same direction. Although SPSS reports the χ^2 test for linearity by default, it is not appropriate for nominal-level data.

FISHER'S EXACT TEST

There may be times that we have nominal level data on two variables but the sample size is too small, usually defined as less than or equal to 20, for the χ^2 test of independence. In these cases, **Fisher's exact test** is a useful procedure. The purpose of Fisher's exact test is to examine whether two populations differ from each other in the proportion of subjects who fall into one of two classifications of the variables. In the most common form of Fisher's exact test, there are usually two levels for each variable and the collected data are classified into a 2 × 2 table, as in Table 9–5. However, Fisher's exact test may be extended to larger tables. The null and alternative hypotheses are

H_0: There is no association between the two variables of interest.

H_A: There is an association between the two variables of interest.

Table 9–5 Contingency Table of Smoking by Sex

	Sex		
Smoking	*Male*	*Female*	*Total*
Yes	7	2	9
No	3	8	11
Total	10	10	20

We calculate Fisher's exact test by

$$p = \frac{(a + b)!(c + d)!(a + c)!(b + d)!}{N!a!b!c!d!},$$

where p is the probability of obtaining the observed frequencies that appear in the contingency variable. As you might expect, calculating the p for this test with a hand-held calculator is quite tedious. When using SPSS, Fisher's exact test is the default of the χ^2 test of independence when sample sizes are less than 20 or when the expected frequencies for each cell are less than 5. The assumptions for Fisher's exact test are displayed in Exhibit 9–5.

Exhibit 9–5 Assumptions for Fisher's Exact Test

1. Both variables of interest are dichotomous.
2. Assigned categories are mutually exclusive.
3. Data contained within the table are frequencies, not scores.

As an example, let's assume that we are studying the relatedness of smoking and sex; we will limit the sample size to 20. The null and alternative hypotheses are

H_0: There is no association between smoking and sex.

H_A: There is an association between smoking and sex.

$\alpha = 0.05$

The results of our sampling appear in Table 9–5. Given these data, p would be calculated as:

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{N!a!b!c!d!},$$

$$= \frac{(9)!(11)!(10)!(10)!}{(20)!(7)!(2)!(3)!(8)!}$$

$$= 0.032$$

In this case, the resulting p is the actual probability of obtaining this sampling result. The calculation of $p = 0.032$ is a one-tailed test result. Since our null is nondirectional (i.e., two-tailed), we double the obtained p value of 0.032, which results in a two-tailed p value of 0.064. Since $p > 0$ .05, we fail to reject the null and conclude that the two variables, smoking and sex, are independent.

To calculate Fisher's exact test using SPSS, we must first code the dichotomous variables to be used in the analysis. The sex variable will be coded as male = 1 and female = 2. The smoking variable will be coded as smoking = 1 and no smoking = 2. Fisher's exact test is provided by default. The SPSS results for Fisher's exact test appear in Exhibit 9–6. SPSS provides both the one- and two-tailed p values. The two-tailed p value is 0.070. SPSS adjusts the p value because the row margin totals are not equal; that is, $(a + b) \neq (c + d)$.

Exhibit 9–6 SPSS Output for Fisher's Exact Test

SMOKING * SEX Crosstabulation

			SEX		
			MALE	*FEMALE*	*Total*
SMOKING	SMOKING	Count	7	2	9
		Expected Count	4.5	4.5	9.0
		Std. Residual	1.2	−1.2	
	NO SMOKING	Count	3	8	11
		Expected Count	5.5	5.5	11.0
		Std. Residual	−1.1	1.1	
Total		Count	10	10	20
		Expected Count	10.0	10.0	20.0

Chi-Square Tests

	Value	*df*	*Asymp. Sig. (2-sided)*	*Exact Sig. (2-sided)*	*Exact Sig. (1-sided)*
Pearson Chi-Square	5.051[b]	1	.025		
Continuity Correction[a]	3.232	1	.072		
Likelihood Ratio	5.300	1	.021		
Fisher's Exact Test				.070	.035
Linear-by-Linear Association	4.798	1	.028		
N of Valid Cases	20				

[a] Computed only for a 2x2 table
[b] 2 cells (50.0%) have expected count less than 5. The minimum expected count is 4.50.

χ^2 GOODNESS OF FIT

In the χ^2 **goodness-of-fit test**, we are also comparing actual frequencies with expected or theoretical frequencies in a distribution. However, the expected counts instead of being based on the collected data, as in Table 9–3, are based on our knowledge of the population under study. For example, if 75% of the physicians practicing in our hospital are men and 25% are women, we will expect to draw a sample composed of 75% men and 25% women. We know that the drawn sample will not precisely match the proportions in the population, but we want to be assured that the sample proportions are not significantly different from the underlying population proportions. In this procedure, the goal is to determine how well the observed counts in our sample fit the expected counts based on the model—that is, how well the observed count matches a theoretical frequency distribution.

As an example, a researcher received 355 responses to a survey that was mailed to directors of health information management (HIM) centers in acute care facilities across the United States. A χ^2 goodness-of-fit test was conducted to ensure that the respondent profile matched that of the distribution of practitioners in the general HIM population. A variable used to compare the respondent profile to the general population was the hospital bed size in which professionals practiced. The obtained frequency distribution of HIM respondents is compared to the respondent profile in Table 9–6.

Table 9–6 Hypothetical HIM Population Distribution and Respondent Profile by Hospital Size

Hospital Size	*HIM Population % 1990*	*Respondent %*
< 100 beds	20.1	21.7
101–300 beds	38.9	41.4
301–500 beds	21.9	22.3
501+ beds	19.0	14.6

The null hypothesis, based on the actual HIM frequency distribution, is

H_0: $p_1 = 0.201$, $p_2 = 0.389$, $p_3 = 0.219$, $p_4 = 0.190$

H_A: At least one of the hypothesized proportions is different from the expected.

The expected frequencies are generated on the basis of the actual frequency distribution of the characteristic of interest in the general population. In the example, 20.1% of the HIM professionals practice in hospitals with fewer than 100 beds; therefore, the expected frequency count for the actual respondents for this category is 71.4 ($355 \times 0.201 = 71.4$). Table 9–7 displays the observed and expected frequencies. If the calculated χ^2 is small, there is close agreement between the observed and expected frequencies. As the discrepancy between the observed and expected frequencies increases, the calculated χ^2 increases, and the more likely we are to reject the null hypothesis and conclude that the population proportions are significantly different from the expected population proportions.

Table 9–7 Observed *(O)* and Expected *(E)* Frequencies of Respondents by Hospital Bed Size (Hypothetical Data)

Hospital Bed Size	*O*	*E*	*(O − E)*	*$(O - E)^2$*	*$(O - E)^2/E$*
< 100 beds	77	71.4	5.6	31.36	0.44
101–300 beds	147	138.2	8.8	77.4	0.56
301–500 beds	79	77.8	1.2	1.44	0.02
501+ beds	52	67.5	−15.5	240.25	3.56
Total	355				4.58

In this example, we are comparing observed frequencies with the expected frequencies across four categories (rows) on the variable hospital bed size. In the one-variable case, degrees of freedom are based on the number of categories (k) and are equal to the number of categories minus 1 (i.e., $k - 1$). In this example, $k - 1 = 3$, and for 3 df, the critical value of χ^2, $\alpha = 0.05$, is 7.82. Since the calculated χ^2 does not equal or exceed the critical value of χ^2, we fail to reject the null hypothesis and conclude that the respondent profile "fits" the population profile. Even though we can generalize our results to our population of interest, it is important to note that the category "501+ beds" contributes the most to the χ^2 statistic. It is this category that deviates the most from the expected frequencies.

The SPSS calculations for the data in Table 9–7 appear in Exhibit 9–7. SPSS reports the χ^2 as 4.575, $p = 0.206$. SPSS requires that the expected frequencies for each category be entered into the dialog box. The expected frequencies may be specified as proportions, percentages, or actual values. The assumptions for the χ^2 goodness-of-fit test are summarized in Exhibit 9-8.

Exhibit 9–7 SPSS Output for Chi-Square Goodness of Fit

Hospital Size

	Observed N	*Expected N*	*Residual*
6–100 Beds	77	71.4	5.6
101–300 Beds	147	138.2	8.8
301–500 Beds	79	77.8	1.2
501+ Beds	52	67.5	−15.5
Total	355		

Test Statistics

	Hospital Size
Chi-Square[a]	4.576
df	3
Asymp. Sig.	0.206

[a] 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 67.5.

Exhibit 9–8 Assumptions for χ^2 Goodness of Fit

1. Select a criterion upon which to compare the selected sample with the underlying population, such as sex, level of education, or certification.
2. Select the null and alternative hypotheses and the significance level for rejection of the null.
3. The criterion variable must have two or more categories, and the categories must be mutually exclusive and exhaustive.
4. The data contained within the cells are frequencies, not scores.
5. Determine the expected or theoretical frequencies under the assumption that the null hypothesis is true. The expected frequencies are obtained by the researcher's knowledge of the target population. Frequencies must be specified in advance.
6. If categories are dichotomous, the expected frequencies for each cell should be at least 5. If there are more than two categories, no more than 20% of the cells should have frequencies less than 5. If this requirement cannot be met, consider collapsing the categories.
7. Compare the observed frequencies to the expected frequencies for each cell.
8. If the aggregate discrepancy (χ^2calc) between the observed and expected frequencies is too great to attribute to chance at the selected significance level, reject the null.

✓ To Obtain a Chi-Square Test Using SPSS:

- From the "Analyze" menu, choose:
  →Nonparametric
  →chi-square
- Select one or more test variables. Each variable produces a separate test.
- Click "Options" for descriptive statistics, etc.

χ^2 TEST FOR PAIRED DATA—McNEMAR TEST

When we previously discussed identifying differences between paired data, we reviewed the paired *t* test. In the paired *t* test, the dependent variable is continuous and the independent variable is categorical. The **McNemar test** is used when we have paired data and the variable under study is nominal. The McNemar test is a modified χ^2 with one degree of freedom. If the resultant χ^2 is significant, the conclusion is that there was a change from the pretest condition to the posttest condition, or that there is an association between the treatment and the effect. Exhibit 9–9 outlines the assumptions for the McNemar test.

Exhibit 9–9 Assumptions for McNemar Test

1. Observations are dichotomous.
2. Dichotomous measures are paired observations of the same subjects or matched pairs.
3. Dichotomous categories are mutually exclusive.
4. Data that are contained in the cells of the table are the number of pairs.

As an example, we are interested in comparing student attitudes before enrollment in a professional HIM baccalaureate degree program and after graduation. Fifty students were asked to complete an attitude survey prior to formal program enrollment and one week post-graduation. Responses were recorded as either positive or negative—a dichotomous response. The subjects (students) are serving as their own controls, as it is their pre-enrollment and post-graduation scores that are being compared.

For the McNemar test, we first prepare a 2×2 contingency table. The pre-enrollment attitudes appear on the left, and post-graduation attitudes appear across the top (Table 9–8).

Table 9–8 Pre-Enrollment and Post-Graduation Attitudes of HIM Students

	Post-Graduation		
Pre-Enrollment	*Positive*	*Negative*	*Total*
Positive	38 *a*	5 *b*	43 *a* + *b*
Negative	2 *c*	5 *d*	7 *c* + *d*
Total	40 *a* + *c*	10 *b* + *d*	50 *a* + *b* + *c* + *d*

Each cell in the table represents one of the four following combinations:

a = positive before and after—no change
b = change from positive to negative
c = change from negative to positive
d = negative before and after—no change

Only the b and c cells are used in the analysis because they represent change—if it occurred. Cells a and d do not contribute anything to the analysis, since the cell frequencies remain the same before enrollment in the baccalaureate program and after graduation. The null and alternative hypotheses are

H_0: pre-enrollment attitude = post-graduation attitude.
H_A: pre-enrollment attitude ≠ post-graduation attitude.

The formula for the McNemar χ^2 is

$$\begin{aligned}\text{McNemar } \chi^2 &= (|b - c| - 1)^2/(b + c)\\ &= (|5 - 2| - 1)^2/(5 + 2)\\ &= 0.57.\end{aligned}$$

The critical value for χ^2, $\alpha = 0.05$, with one degree of freedom, is 3.841. Since the calculated χ^2 does not equal or exceed the critical value, the McNemar test result is not statistically significant. So even though there were observed changes in attitudes, in that three students changed their attitudes from positive to negative, the changes are not statistically significant. But even though these results are not statistically significant, it would be worthwhile to investigate the reasons for change in student attitudes.

The SPSS calculations for the McNemar test appear in Exhibit 9–10. The two paired variables for analysis are the pre-enrollment attitudes and the post-graduation attitudes. Positive attitudes were coded as "1," and negative attitudes were coded as "2." SPSS does not display the actual McNemar statistic, and it uses the binomial distribution (see Chapter 10) to calculate the level of significance ($p = 0.453$).

Exhibit 9–10 SPSS Output for McNemar Test

Pre-Enrollment * Post-Enrollment Crosstabulation

		Post-Enrollment		
		Positive	*Negative*	*Total*
Pre-Enrollment	positive	38	5	43
	negative	2	5	7
Total		40	10	50

Chi-Square Tests

	Value	*Exact Sig. (2-sided)*
McNemar Test		0.453[a]
N of Valid Cases	50	

[a] Binomial distribution used.

✓ To Obtain McNemar's Test Using SPSS:

- From the "Analyze" menu, choose:
  →descriptive statistics
  →crosstabs
- Select one or more row variables and one or more column variables.
  →Click "statistics for McNemar test"

CONCLUSION

In this chapter, we were introduced to the various forms of the chi-square test, which is a nonparametric procedure. Nonparametric procedures are less restrictive than their parametric counterparts. Nonparametric procedures do not make assumptions regarding the underlying

population distribution; nor do they require large sample sizes. Nonparametric methods may also be used for analyzing frequencies for nominal- and ordinal-level data.

We can use the chi-square test to measure the relationship between two variables. This is called the chi-square test of independence. Other frequency measures of association based on chi-square include Cramer's V, the phi coefficient, and Fisher's exact test. These tests use frequencies rather than the values of observations in the calculations.

If we are interested in making comparisons between two independent samples, the forms of chi-square to be used are Fisher's exact test and the chi-square test for two independent samples. If we are interested in making comparisons between paired samples, we can use McNemar's chi-square.

ADDITIONAL RESOURCES

Besag, F.P., and P.L. Besag. 1985. *Statistics for the helping professions.* Beverly Hills, CA: Sage Publications.

Fleiss, J.L. 1981. *Statistical methods for rates and proportions.* New York: John Wiley & Sons.

Gibbon, J.D. 1993. *Non-parametric statistics: An introduction.* Newbury Park, CA: Sage Publications.

Jekel, J.F., et al. 1996. *Epidemiology, biostatistics and preventive medicine.* Philadelphia: W.B. Saunders.

Pett, M.A. 1997. *Nonparametric statistics for health care research: Statistics for small samples and unusual distributions.* Thousand Oaks, CA: Sage Publications.

Statsoft Electronic Textbook. 1998. www.statsoft.com/textbook

Appendix 9–A

Exercises for Solving Problems

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
2. Describe the major differences between parametric and nonparametric procedures.
3. Describe the circumstances under which it would be appropriate to use the chi-square test.
4. What questions would a researcher be attempting to answer when using the chi-square goodness-of-fit test?

MULTIPLE CHOICE

1. We use the chi-square test to test for differences between:
   a. means
   b. variances
   c. frequencies
   d. proportions
   e. c and d
   f. all of the above
2. You want to compare the number of male and female patients discharged by service at Critical Care Hospital. There are 10 services. For the chi-square test, the number of degrees of freedom is:
   a. 1
   b. 2
   c. 9
   d. 10

3. In the chi-square test, if the result is not significant, the:
   a. observed frequencies are similar to the expected frequencies
   b. observed frequencies do not match the expected frequencies
   c. expected frequencies are greater than the observed frequencies
   d. observed frequencies are greater than the expected frequencies

For questions 4 through 9, refer to the following R × C table:

	Phys A	*Phys B*	*Phys C*	*Total*
Male				60
Female	28	42		140
Total	40	60	100	200

4. The observed number of male patients discharged by physician C is:
   a. 10
   b. 20
   c. 30
   d. not enough information provided

5. The expected number of male patients discharged by physician B is:
   a. 10
   b. 18
   c. 20
   d. not enough information provided

6. The observed number of males discharged by physician A is:
   a. 10
   b. 20
   c. 30
   d. not enough information provided

7. The expected number of females discharged by physician A is:
   a. 20
   b. 28
   c. 30
   d. not enough information provided

8. The number of degrees of freedom for this problem is equal to:
   a. 2
   b. 4
   c. 6
   d. 199

9. For the above problem, χ^2_{calc} is equal to 16.87; χ^2_{crit} is equal to 5.99. We therefore:
   a. reject the null hypothesis
   b. fail to reject the null hypothesis
   c. conclude that these physicians have the same number of discharges
   d. conclude that these physicians prefer to treat men over women

10. The phi coefficient is used to measure the:
   a. correlation of two ordered variables
   b. correlation of interval-level variables
   c. strength of the association between variables in a 2×2 table
   d. all of the above

PROBLEMS

1. You are assisting Dr. Hartman in studying the number of deaths due to acute myocardial infarctions (AMIs). Dr. Hartman is particularly interested in knowing if more men than women died from AMIs. To answer this question, you review discharges by sex for DRG 123, Circulatory Disorders with AMI, Expired. Use the nonparametric procedure for chi-square to determine if there is an association between sex and deaths due to AMI at Critical Care Hospital. A frequency distribution of discharges by sex and age from DRG 123 appears in Table 9–A–1.
   a. State the null and alternative hypotheses.
   b. State the alpha level.
   c. What is the result of the chi-square test?
   d. State your conclusions.

Table 9–A–1 Frequency Distribution of Discharges by Age and Gender, DRG 123 (SPSS Output)

Age * Gender Crosstabulation

Count

		Gender		
		FEMALE	*MALE*	*Total*
Age	49	0	1	1
	50	0	1	1
	61	0	1	1
	66	0	1	1
	75	1	0	1
	76	0	1	1
	77	0	1	1
	88	1	0	1
	88	0	1	1
Total		2	7	9

2. Using the same information in Table 9–A–1, use the chi-square test to determine if there is an association between age and sex for discharges from DRG 123. Calculate the phi coefficient.
   a. State the null and alternative hypotheses.
   b. State the alpha level.
   c. What is the result of the chi-square test?
   d. What does the phi coefficient indicate?
   e. State your conclusions.

CHAPTER 10

Nonparametric Methods

KEY TERMS
Spearman rho
Sign test
Wilcoxon signed ranks test
Mann-Whitney Wilcoxon test
Kruskal-Wallis test

LEARNING OBJECTIVES
At the conclusion of this chapter, you should be able to:

1. Define key terms.
2. Conduct the following tests for given situations: Spearman rho, Sign test, Wilcoxon signed ranks test, Mann-Whitney Wilcoxon test, and the Kruskal-Wallis test.
3. Distinguish between the Pearson *r* correlation coefficient and the Spearman rho correlation coefficient.
4. Use microcomputer statistical software to solve nonparametric problems.

In Chapter 9, we discussed the various forms of chi square and their respective applications. In this chapter, we will discuss some of the other commonly used nonparametric procedures. We will first discuss the Spearman rho rank order correlation coefficient. This test is the nonparametric counterpart of the Pearson *r* correlation coefficient.

THE SPEARMAN RHO RANK ORDER CORRELATION COEFFICIENT

An alternative to the Pearson *r* correlation coefficient that we previously discussed is the Spearman rho rank order correlation coefficient. The Spearman rho is used when at least one of the two variables under study falls on the ordinal scale of measurement. The

correlation coefficient obtained from the **Spearman rho** procedure is the result of the rankings of the observations, not the actual values of the observations.

To calculate the Spearman rho, we rank the observations on each variable from lowest to highest. Tied observations are assigned the average of the ranks. If two observations are tied for second position, they actually occupy positions 2 and 3, and both are assigned the average rank of (2 + 3)/2, or 2.5. For example, consider the following data set:

Observed score	2	5	5	5	7
Rank	1	2	3	4	5

The total number of possible ranks is five, but three subjects have the same observation of "5," and it is incorrect to assign them different ranks. To obtain the average rank, we sum the ranks 2, 3, and 4 and divide by 3 to obtain the average rank of 3. The assigned ranks become:

Observed score	2	5	5	5	7
Rank	1	3	3	3	5

Once the ranks have been assigned, the differences between the ranks on the X and Y variables are obtained, summed, and squared (Table 10–1). The resulting values are substituted into the following formula:

$$r_{\text{rho}} = 1 - [(6 \Sigma D^2)/n(n^2 - 1)]$$

The range for the Spearman rho is the same as that for the Pearson r, -1.0 to $+1.0$, and has the same interpretation.

As an example, let's consider a hypothetical case of eight subjects who smoke. The R_1 column (Table 10–1) indicates the rank for each patient in terms of number of cigarettes

Table 10–1 Patient Ranks by Smoking and Severity of Illness

	Number of Cigarettes Smoked	*Severity of Illness*	*Difference in Ranks*	
Patient	R_1	R_2	$D(R_1 - R_2)$	D^2
1	1	2	−1	1
2	2	4	−2	4
3	3	3	0	0
4	4	1	3	9
5	5	7	−2	4
6	6	5	1	1
7	7	8	−1	1
8	8	6	2	4
Total			0	24

smoked, from lowest (rank = 1) to highest (rank = 8). The R_2 column indicates the rank for each patient in terms of severity of illness, from least severe (rank = 1) to most severe (rank = 8). To complete the table, R_2 is subtracted from R_1, and the difference is squared. The null and alternative hypotheses are

H_0: There is no relationship between number of cigarettes smoked and severity of illness.

H_A: There is a relationship between number of cigarettes smoked and severity of illness.

$\alpha = 0.05$

Note that the *D* column sums to zero. This column does not need to be summed, but it does serve as a check on our calculations. The difference between the ranks column should always sum to zero. Substituting our obtained values into the formula:

$$\begin{aligned} r_{\text{rho}} &= 1 - [(6 \Sigma D^2)/n(n^2 - 1)] \\ &= 1 - [6(24)]/[8(64 - 1)] \\ &= 0.71 \end{aligned}$$

To test the significance of rho, we use the Pearson *r* table (Appendix B, Table B–5), where df = $n - 2$. However, when the sample size is less than 10, we must use the *t* distribution where df = $n - 2$ (Appendix B, Table B–2). The formula for calculating *t* is

$$t = \text{rho}\sqrt{n - 2}/\sqrt{1 - \text{rho}^2}$$

substituting in the formula

$$\begin{aligned} t &= 0.71\sqrt{6}/\sqrt{1 - 0.71^2} \\ &= 1.74/0.7 \\ &= 2.49. \end{aligned}$$

For six degrees of freedom, $\alpha = 0.05$, $t_{\text{crit}} = 2.447$; since the calculated *t*, 2.49, is greater than the critical *t*, it falls in the region of rejection. We therefore reject the null hypothesis and conclude that there is a statistically significant positive relationship between the number of cigarettes smoked and severity of illness.

The Spearman rho can also be calculated when one of the variables falls on the interval scale of measurement, but it is first necessary to convert the observations to ranks. As an example, we will calculate the Spearman rho for a data set in which variable *X* falls on the ordinal scale of measurement and variable *Y* falls on the interval scale, as in Table 10–2. Variable *Y* must be converted to ranks. The null and alternative hypotheses are

Table 10–2 Spearman Rho for Ordinal and Interval Level Data

	X	*Y*	*Difference in Ranks*		
Patient	R_1		R_2	$S(R_1 - R_2)$	D^2
1	8.5	135	6.5	2	4
2	8.5	120	9	−0.5	0.25
3	6	140	5	1	1
4	6	130	8	−2	4
5	6	135	6.5	−0.5	0.25
6	4	145	4	0	0
7	3	150	2.5	0.5	0.25
8	1.5	150	2.5	−1	1
9	1.5	160	1	0.5	0.25
Total				0	11

H_0: There is no relationship between variables *X* and *Y*.

H_A: There is a relationship between variables *X* and *Y*.

$\alpha = 0.05$

The calculations for the Spearman rho are

$$\begin{aligned} r_{\text{rho}} &= 1 - [(6 \Sigma D^2)/n(n^2 - 1)] \\ &= 1 - [6(11)]/[9(81 - 1)] \\ &= 0.91 \end{aligned}$$

Since we have a sample size that is less than 10, we evaluate the significance of rho by using *t*:

$$t = \text{rho}\sqrt{n - 2}/\sqrt{1 - \text{rho}^2}$$

substituting in the formula

$$\begin{aligned} t &= 0.91\sqrt{7}/\sqrt{1 - 0.91^2} \\ &= 2.41/0.41 \\ &= 5.88. \end{aligned}$$

The critical value of *t*, for seven degrees of freedom, $\alpha = 0.05$, is 2.365. Since the calculated *t* is greater than the critical *t*, it falls in the region of rejection. We reject the null hypothesis and conclude that there is a strong positive relationship between variables *X* and *Y*.

✓ To Obtain the Spearman Rho Using SPSS:

- From the "Analyze" menu, choose:
  →Correlate
  →Bivariate
- Select two numeric variables.
  →Select options.
  →Click "Spearman rho."

The SPSS output for the Spearman rho appears in Exhibit 10–1. SPSS does not require that interval-level data be recoded for the Spearman rho.

Exhibit 10–1 SPSS Output for Spearman Rho

			X	*Y*
Spearman's rho	X	Correlation Coefficient	1.000	−.905*
		Sig. (2-tailed)	.	.001
		N	9	9
	Y	Correlation Coefficient	−.905*	1.000
		Sig. (2-tailed)	0.001	.
		N	9	9

* Correlation is significant at the 0.01 level (2-tailed).

LOCATION TESTS FOR SINGLE AND PAIRED SAMPLES

The **sign test** and the **Wilcoxon signed ranks test** are used for analyzing data from single data sets or data collected in pairs. The results of these tests are inferences concerned with the median of a population (M) and the median (M_D) of the population differences for paired samples. These are examples of location tests. In a location test, we are concerned with the value of a measure of central tendency, or central location, and its associated confidence intervals. The parametric counterpart of these tests is the paired t test.

Sign Test

The sign test is used to test hypotheses about the location of a population distribution. The test is often used when evaluating data in the form of matched pairs—that is, "before" and "after" data such as pretests and posttests that are in the form of a single sample. In this case, the test is for a median difference of zero between the matched pairs rather than a mean difference of zero, as in the paired sample t test. The sign test does not require that the underlying population be normally distributed. In the sign test, the null hypothesis is stated as:

$$H_0: M = M_0$$

As you recall, the median is the value that divides the area under the curve in half. In a normal distribution, the mean and median are equal. If the null hypothesis is true, M_0 is the central value—one-half of the observations in the sample should be larger than M_0, and one-half should be smaller. To evaluate the null, the sign statistic, S, is defined as

S = number of plus signs among the differences $X_1 - M_0$, $X_2 - M_0 \ldots X_n - M_0$.

The null distribution of S is the binomial distribution, with n and $p = 0.5$.

$$p(S = x) = \binom{n}{x}(0.5)^x(0.5)^{n-x}$$

The binomial distribution describes the possible number of times that an event will occur in a given number of trials. When using the binomial distribution, the variables of interest are at the nominal level of measurement and are dichotomous. We use the binomial distribution when we are interested in the frequency of occurrence of an event—for example, how many patients survived/did not survive a particular cancer treatment.

In using the binomial distribution, there must only be two possible mutually exclusive outcomes, such as survived/did not survive, yes/no, or success/failure. Each observation must be independent, and the outcome of one trial must not influence another. The two possible outcomes are designated as

p = probability that a successful event (x) will occur in a single trial.

$1 - p$ = probability that a successful event (x) will not occur in a single trial.

To determine the value of p or $1 - p$, it is necessary to know the number of ways in which success or failure can occur in a specified number of trials (n). This is obtained from the binomial coefficient $\binom{n}{x}$, where n equals the number of trials and x is the number of times a successful outcome will occur.

As an example, let's consider the probability of drawing a delinquent medical record from the incomplete file. We know that 10% of the records are delinquent at any one time. If we randomly select five records from the incomplete file, what is the probability that one will be delinquent? The binomial coefficient is $\binom{5}{1}$; that is, the number of possible combinations of an event where five records are randomly drawn with the probability of one delinquent record. This is illustrated as

D N N N N
N D N N N
N N D N N
N N N D N
N N N N D

(where D = delinquent record and N = nondelinquent record)

The formula for determining the probability of a designated number of successes $p(x)$ in n trials equals the number of possible combinations of the event multiplied by the probability of success, 10%, and failure.

$$p(x) = \binom{n}{x}(p)^x(1-p)^{n-x}$$

The probability of drawing one delinquent record in five trials is:

$$\begin{aligned} p(x) &= \binom{5}{1}(0.10)^1(1-0.10)^{5-1} \\ &= 5(0.10)(0.90)^4 \\ &= 5(0.10)(0.6561) \\ &= 0.32805, \text{ or } 0.33 \end{aligned}$$

Thus, the probability that one delinquent record will be drawn in five trials is 0.33, or 33%.

For sample sizes greater than 20, we can use the normal approximation to the binomial distribution with the test statistic z:

$$z = \frac{S - 0.5n \pm 0.5}{0.5\sqrt{n}}$$

where ± 0.5 is a continuity correction to improve the normal approximation to the binomial distribution. If the calculated sign statistic, S, is less than $0.5n$, $+0.5$ is used in the above formula; if S is greater than $0.5n$, -0.5 is used. When calculating S, any difference between X_i and X_0 that is equal to zero is called zero. As long as zero differences are few, ignore them, and reduce the size of n accordingly.

Under the previously stated null hypothesis, there are two one-sided alternatives and one two-sided:

$H_{A+}: M > M_0$

$H_{A-}: M < M_0$

$H_A: M \neq M_0$

In the positive-sided alternative, we expect that the number of observations greater than M_0 will be large and that the calculated S will be large. Conversely, in the negative-sided alternative, we expect that the number of observations less than M_0 will be large and that the calculated S will be small.

We will now look at a hypothetical example for calculating the S statistic. In a health information management (HIM) coding class, the students have been reluctant to use encoding software to code ICD-9-CM diagnoses and procedures. The instructor believes that if the students used the computer program just once they would have a more positive attitude

toward using it. To test this hypothesis, the instructor administers a pre- and postuse attitude assessment on the use of the computerized software. The assessment consisted of 20 questions, all stated in positive terms regarding computerized encoders. The null and alternative hypotheses are

H_0: $M = M_0$.

H_A: $M > M_0$.

To test the hypothesis in the single-sample sign test, the average preuse assessment score for each item is subtracted from the average postuse assessment score for each item. The differences are recorded as either plus (greater posttest score) or minus (greater pretest score). The average difference for each item appears in Table 10–3.

Table 10–3 Pretest and Posttest Coding Results, Sign Test

Average Pretest Posttest			*Average Pretest Posttest*		
Item	*Difference*	*Sign*	*Item*	*Difference*	*Sign*
1	+0.3	+	11	−0.3	−
2	+0.1	+	12	+0.5	+
3	−0.4	−	13	+0.1	+
4	+0.2	+	14	+0.2	+
5	+0.5	+	15	−0.5	−
6	+0.3	+	16	+0.4	+
7	−0.2	−	17	+0.1	+
8	+0.6	+	18	+0.3	+
9	+0.4	+	19	+0.2	+
10	−0.1	−	20	+0.1	+

Our calculations result in $S = 15$; that is, 15 plus signs. Conclusions for nonparametric tests are often reported in terms of the p value, and we use the binomial table, where $p = 0.50$, to determine the value of p when $n \leq 20$ (Appendix B, Table B–7). For a given n, the entry in the column labeled "p" is the left-tail cumulative probability for the corresponding number in the column labeled "Left S," and this same p value is the right-tail probability for the corresponding number in the column labeled "Right S." A partial table for the binomial distribution for $n = 20$ is presented in Exhibit 10–2.

For our example, $n = 20$, the Pr ($S \geq 15$) = 0.0207. Thus, we reject the null and conclude that students had a more positive attitude regarding computerized encoders after actually using the computer programs. Using the same data, we can also calculate the corresponding large sample approximation (the continuity correction, -0.5, is used because $S > 1/2\ n$):

Exhibit 10–2 Partial Binomial Distribution, $p = 0.50$

n	*Left S*	*p*	*Right S*
20	0	0.0000	20
	1	0.0000	19
	2	0.0002	18
	3	0.0013	17
	4	0.0059	16
	5	0.0207	15
	6	0.0577	14
	7	0.1316	13
	8	0.2517	12
	9	0.4119	11
	10	0.5881	10

Source: Reprinted from National Bureau of Standards.

$$z = \frac{S - 0.5n \pm 0.5}{0.5\sqrt{n}}$$
$$= (15 - 10 - 0.5)/0.5\sqrt{20}$$
$$= 4.5/2.24$$
$$= 2.01$$

The corresponding p value for $z = 2.01$ is 0.0222 (Appendix B, Table B–1). The p value can also be calculated using the binomial test available on SPSS. From the "Analyze" menu, select "Nonparametric" and then select "Binomial." The p value must be set at 0.50 to correspond to the sign test. The results of the SPSS output appear in Exhibit 10–3. Note that the p value for the SPSS binomial test is 0.041—this is for a two-tailed test.

Exhibit 10–3 SPSS Output for the Sign Test

Binomial Test

	PRE-, POSTTEST DIFFERENCES		
	Group 1	*Group 2*	*Total*
Category	Positive	Negative	
N	15	5	20
Observed Prop.	.75	.25	1.00
Test Prop.	.50		
Exact Sig. (2-tailed)	.041		

Dividing the p value in half, we obtain $p = 0.0205$, which is very close to the p value for the binomial distribution that appears in Exhibit 10–2.

The Wilcoxon Signed Ranks Test

In the sign test, the only information we used to calculate S was the signs of the differences between the pre- and posttest scores. Information regarding the size of the differences between the pre- and posttest scores was not considered. As we shall see, the Wilcoxon signed ranks test uses more information in its calculations and thus is a more powerful test. The Wilcoxon signed ranks test assumes that the observations are symmetric about the median M. If the underlying population is extremely nonsymmetric, the Wilcoxon signed ranks test should not be used. The assumptions for the Wilcoxon signed ranks test are summarized in Exhibit 10–4.

Exhibit 10–4 Assumptions for the Wilcoxon Signed Ranks Test

1. Data are paired observations from a single random sample either constructed as matched pairs or using subjects as controls.
2. Data are continuous and at least at the ordinal level of measurement.
3. There is symmetry of the difference scores about the true median for the population.

In the sign test discussed previously, we looked at the average differences for each item. We will now consider the differences in pre- and posttest scores for 15 students in the class. Here we will view the pre- and posttest scores as matched pairs—similar to the t test for matched pairs. What we want to know is whether there is a significant difference in the pre- and posttest medians. In the Wilcoxon signed ranks test, we calculate T where $T+$ is the sum of the positive ranks and $T-$ is the sum of the negative ranks. To calculate T, we take the absolute values of the differences between the pre- and posttest scores, keeping track of the original positive and negative signs, and assign ranks from lowest to highest. In the example, the rankings will range from 1 to 15. If there are observations that are tied, we assign the average or the midpoint of the ranks they would have if they were not tied. The absolute values of the differences between pre- and posttest scores and their corresponding ranks appear in Table 10–4. The null and alternative hypotheses are the same as for the sign test:

$$H_0: M = M_0$$

$$H_{A+}: M > M_0$$

Under the null hypothesis, we would expect that the sum of the positive rankings, $T+$, would be similar to the sum of $T-$. In our example, the sum of $T+$ equals 91, and the sum of $T-$ equals 29—a wide difference in values. To determine the significance of T, we use

Table 10–4 The Sign Test of Pretest and Posttest Coding Results

Coder	*Pretest Score*	*Posttest Score*	*Difference D*	*Sign*	$\lvert D \rvert$	*Rank*
1	13	20	+7	+	7	11.5
2	15	16	+1	+	1	1
3	15	11	−4	−	4	5.5
4	12	16	+4	+	4	5.5
5	10	15	+5	+	5	7
6	8	11	+3	+	3	3.5
7	12	10	−2	−	2	2
8	9	15	+6	+	6	9
9	9	18	+9	+	9	14
10	5	12	+7	+	7	11.5
11	10	20	+10	+	10	15
12	7	1	−6	−	6	9
13	5	13	+8	+	8	13
14	16	10	−6	−	6	9
15	10	7	−3	−	3	3.5
				$S = 10$		
		$\Sigma T_+ = 91$	$\Sigma T_- = 29$			

the Wilcoxon signed ranks distribution of T. For $T+ = 91$, $n = 15$, we obtain a p value of 0.042. We therefore reject the null and conclude that $M > 0$; thus, students' use of computerized encoders did have a positive impact on attitude.

Tables for the critical values of T are often not large enough for sample sizes greater than 15. When the sample size is greater than 15, the normal approximation must be used to calculate T, where

$$z = \frac{T - n(n + 1)/4}{\sqrt{n(n + 1)(2n + 1)/24}},$$

where T is the sum of the positive or negative ranks, depending on the proposed alternative hypotheses, and n is the number of positive and negative ranks, excluding ties.

We can use SPSS to calculate the sign test and the Wilcoxon signed ranks test (Exhibit 10–5). For comparison purposes, the Student's t for matched pairs is also calculated (Exhibit 10–6). The p value for the Wilcoxon signed ranks test is given for the normal approximation of T. Since the p value is for a two-tailed test, we divide the given p value of 0.078 by 2, and the significance level becomes 0.039. Also, the p value for the Student's t test is for the two-sided alternative. For the one-sided alternative, where $t = 1.818$ and df $= 14$, the p value is 0.0455 (0.091/2).

Exhibit 10–5 SPSS Output for Wilcoxon Signed Ranks Test

Ranks

POSTTEST–PRETEST

	Negative Ranks	*Positive Ranks*	*Ties*	*Total*
N	5[a]	10[b]	0[c]	15
Mean Rank	5.80	9.10		
Sum of Ranks	29.00	91.00		

[a] POSTTEST < PRETEST
[b] POSTTEST > PRETEST
[c] POSTTEST = PRETEST

Test Statistics[a]

	POSTTEST–PRETEST
Z	−01.763[b]
Asymp. Sig. (2-tailed)	.078

[a] Based on negative ranks.
[b] Wilcoxon Signed Ranks Test.

Exhibit 10–6 SPSS Output for Student's *t* Test for Matched Pairs

Paired Samples Statistics

		Mean	*N*	*Std. Deviation*	*Std. Error Mean*
Pair 1	PRETEST	10.4000	15	3.43927	.88802
	POSTTEST	13.0000	15	5.04268	1.30201

			Pair 1 PRETEST–POSTTEST
Paired Differences	Mean		−2.60000
	Std. Deviation		5.53947
	Std. Error Mean		1.43028
	95% Confidence Interval of the Difference	Lower	−5.66766
		Upper	.46766
t			−1.818
df			14
Sig. (2-tailed)			.091

✓ To Obtain Wilcoxon Signed Ranks Test Using SPSS:

- From the menus, choose:
  →Statistics
    →Nonparametric
      →2 related samples
- Select one or more pairs of variables.

MANN-WHITNEY WILCOXON TEST

When two independent samples violate the assumptions associated with the independent samples *t* test, the **Mann-Whitney Wilcoxon test**, also known as the Mann-Whitney *U* test (*U*) and the Wilcoxon rank sum test (*T* or *W*), may be used in its place. In this test, we are interested in comparing the medians of two independent samples, *X* and *Y*. The Mann-Whitney Wilcoxon test is based on the ranks of observations, with the two independent samples treated as one. Because the test takes into account the rankings of measurements in each sample, it uses more information than the previously discussed sign test.

For the Mann-Whitney Wilcoxon test, the null hypothesis is

$$H_0: M_X = M_Y$$

and the alternative hypotheses are

$$H_{T+}: M_X > M_Y$$
$$H_{T-}: M_X < M_Y$$
$$H_A: M_X \neq M_Y$$

The Mann-Whitney Wilcoxon statistic provides a measurement of the difference between the ranked observations of the two samples and provides evidence of the difference of the medians between the two populations. If the null was true, we would expect the average of the two summed ranks to be about the same. The total possible sum or all the ranks is determined by $N(N + 1)/2$. In the example in Table 10–5, the sum of all the ranks is equal to 300; therefore, we would expect the ranks for each group to average 150 (300/2). If the average rank for hospital sample *X* is greater than the average rank for hospital sample *Y*, most of the physician satisfaction scores for hospital sample *X* will be greater than the physician satisfaction scores from sample *Y*, and vice versa. The procedure for calculating the Mann-Whitney Wilcoxon statistic is outlined in Exhibit 10–7.

Let's consider an example. The director of HIM distributes a satisfaction survey to 12 physicians in the facility (hospital X) who are consistent users of HIM services. The same survey is distributed to 12 physicians at another hospital (hospital Y) within the corporate group. The lowest score that may be obtained on the instrument is 1, and the highest score is 20. The director wants to determine if there is a difference in the median levels of satisfaction between the two hospitals. The null and alternative hypotheses are:

$$H_0: M_X = M_Y$$
$$H_A: M_X \neq M_Y$$

Table 10–5 Calculation of Mann-Whitney Wilcoxon Statistic for Physician Satisfaction Level in Hospital X and Hospital Y

Hospital X			*Hospital Y*		
Doctor	*Score*	*Rank*$_X$	*Doctor*	*Score*	*Rank*$_Y$
1	7	7	13	15	17.5
2	6	5.5	14	14	15
3	4	2.5	15	15	17.5
4	11	10.5	16	16	21
5	16	21	17	16	21
6	4	2.5	18	16	21
7	10	9	19	14	15
8	6	5.5	20	12	12
9	9	8	21	14	15
10	3	1	22	17	24
11	13	13	23	16	21
12	5	4	24	11	10.5
	$\Sigma R_X = 89.5$			$\Sigma R_Y = 210.5$	
	Mean $R_X = 7.5$			Mean $R_Y = 7.5$	
	$T = 89.5$				

Exhibit 10–7 Calculation of Mann-Whitney Wilcoxon Statistic

1. Designate one sample as *X* and the other as *Y*. If the sample sizes are unequal, designate the sample with the fewest observations as sample *X*.
2. Rank all the observations in order from lowest to highest, without regard to whether the observation is from sample *X* or sample *Y*. If ranks are tied, assign the average of the ranks they would have if they were not tied.
3. The observations from the first sample are identified, and their ranks are summed. The result is the *T* statistic.
4. The calculated value of *T* is compared to the critical value of *T*. The critical value is related to the number of observations in sample *X* (n_1) and sample *Y* (n_2).

The results appear in Table 10–5. To assign the ranks, we combine the 24 scores into a single ordered array; we keep track of which sample produced which score by underlining the scores from hospital X:

3 4 4 5 6 6 7 9 10 11 11 12 13 14 14 14 15 15 16 16 16 16 16 17

From eyeballing the raw data, it appears that the physicians in hospital Y are more satisfied than the physicians in hospital X—the scores from hospital X are generally lower than the scores from hospital Y. In addition, the sum of the ranks for the samples X and Y are markedly different from the expected sum, 150. The Mann-Whitney Wilcoxon statistic, *T*, is the sum of the ranks assigned to the *X* observations, or, as in this example, the sum of the ranks for physicians practicing in hospital X.

A large value of T_X supports the alternative hypothesis (H_{T+}: $M_X > M_Y$) and requires a right-tailed p value. A small value of T_X supports H_{T-}: $M_X < M_Y$ and requires a left-tailed p value. For the alternative H_T: $M_X \neq M_Y$, double the smallest p value.

To determine the critical values of T, refer to Appendix B, Table B–8. For $n_1 = 12$ and $n_2 = 12$, $\alpha = 0.05$, the critical values of X range from 120 to 180. The null is rejected if ΣR_X equals or falls below the lower number (120) or if it equals or exceeds the higher number (180). Since $\Sigma R_X < R_{crit}(120)$, we reject the null and conclude that the median level of physician satisfaction with HIM services is significantly different between the two hospitals.

An alternative to calculating T is the U statistic. U is calculated as

$$U_1 = n_1 n_2 + [n_1(n_1 + 1)/2] - \Sigma R_1$$
$$U_2 = n_1 n_2 + [n_2(n_2 + 1)/2] - \Sigma R_2$$

where n_1 is the number of observations in sample 1, n_2 is the number of observations in sample 2, R_1 is the sum of ranks for sample 1, and R_2 is the sum of ranks for sample 2. The null is rejected if the calculated value of U, the smaller of U_1 or U_2, is smaller than the critical value of U at the predetermined alpha level.

In our example, the Mann-Whitney U statistic is calculated as:

$$\begin{aligned} U_1 &= n_1 n_2 + [n_1(n_1 + 1)/2] - \Sigma R_1 \\ &= (12)(12) + [12(12 + 1)/2] - 89.5 \\ &= 144 + 78 - 89.5 \\ &= 132.5 \end{aligned}$$

$$\begin{aligned} U_2 &= n_1 n_2 + [n_2(n_2 + 1)/2] - \Sigma R_2 \\ &= (12)(12) + [12(12 + 1)/2] - 210.5 \\ &= 144 + 78 - 210.5 \\ &= 11.5 \end{aligned}$$

The calculated U statistic is 11.5, $p < 0.05$.

The assumptions for the Mann-Whitney Wilcoxon test are summarized in Exhibit 10–8.

When sample sizes are too large to use the tables, an alternate method for determining statistical significance between the two groups is to calculate z, where

$$z = \frac{\Sigma R_1 - 0.5[n_1(n_1 + n_2 + 1)]}{\sqrt{n_1 n_2(n_1 + n_2 + 1/12}}.$$

Exhibit 10–8 Assumptions for the Mann-Whitney Wilcoxon Test

1. The independent variable is dichotomous, and the scale of measurement for the dependent variable is at least ordinal.
2. Data are collected from a randomly selected sample of independent observations from two independent groups.
3. The categories of the independent variable are mutually exclusive.
4. The population distributions of the two independent samples share a similar unspecified shape but with a possible difference in measures of central tendency.

Using the data in Table 8–14, we have

$$z = \frac{89.5 - 0.5[12(12 + 12 + 1)]}{\sqrt{(12 \times 12)(12 + 12 + 1)/12}}$$
$$= 89.5 - (150\sqrt{300})$$
$$= -3.5.$$

For $\alpha = 0.05$, the critical value of z is -1.96. Since our calculated value, -3.5, falls in the region of rejection, we reject the null and conclude that the level of physician satisfaction between hospitals X and Y is significantly different.

The SPSS output for the Mann-Whitney U test appears in Exhibit 10–9. The Wilcoxon W, which is provided by default, is the same as our calculated T in Table 10–5. SPSS provides both the Wilcoxon W, which is the sum of the ranks of the smaller of U_1 and U_2, and the calculated z statistic, -3.514.

Exhibit 10–9 SPSS Output for Mann-Whitney Wilcoxon Test

Ranks

	Hospital	*N*	*Mean Rank*	*Sum of Ranks*
Rank	X	12	7.46	89.50
	Y	12	17.54	210.50
	Total	24		

Test Statistics[a]

	Rank
Mann-Whitney U	11.500
Wilcoxon W	89.500
Z	−3.514
Asymp. Sig. (2-tailed)	.000
Exact Sig. [2*(1-tailed Sig.)]	.000[b]

[a] Grouping Variable: Hospital.
[b] Not corrected for ties.

✓ To Obtain the Mann-Whitney U Using SPSS:

- From the menu, choose "Analyze"
  →Nonparametric
  →2 independent samples
- Select one or more numeric variables.
  →Click "Mann-Whitney *U*"

KRUSKAL-WALLIS TEST

The nonparametric procedure that is comparable to analysis of variance (ANOVA) is the Kruskal-Wallis procedure. The **Kruskal-Wallis test** is used when the populations under study violate the assumptions of normality. The Kruskal-Wallis test does require that the samples be independent and that there be three or more groups ($k \geq 3$). The procedure for calculating the Kruskal-Wallis statistic, which is similar to that for the Mann-Whitney Wilcoxon statistic, is described in Exhibit 10–10.

Exhibit 10–10 Calculation of Kruskal-Wallis Statistic

1. Combine the k samples into one single ordered array.
2. Rank-order each observation from lowest to highest, keeping track of the samples. When ranks are tied, assign the average of the ranks that would have been assigned.
3. Sample sizes are noted by $n_1, n_2, \ldots, n_k$. The sample sizes do not have to be equal.
4. Calculate the sum and average of the ranks for each sample $R_1, R_2, \ldots, R_k$, and $R_1, R_2, \ldots, R_k$.
5. Calculate the Kruskal-Wallis statistic Q. Let N be the size of the combined samples ($N = n_1 + n_2 + n_3, \ldots, n_i$). Calculate the Q statistic according to

$$Q = 12/N(N+1)\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(N+1)$$

   where R_1 is the sum of the ranks in sample 1, etc.; n_1 is the number of cases in subgroup 1, etc.; and N is the number of cases in all samples.
6. The resulting Q statistic is compared to the critical value of χ^2 for $k - 1$ degrees of freedom.

Like the Mann-Whitney Wilcoxon test, the Kruskal-Wallis test is based on the ranks of the data. The null hypothesis is that all medians are equal:

$$H_0: M_1 = M_2 = M_3$$

The alternative hypothesis is that not all of the medians are the same:

$$H_A: M_1 \neq M_2 \neq M_3$$

Under the null hypothesis, we expect the average of the ranks for each group to be about equal. If the sum of all the ranks is equal to $N(N + 1)/2$, we would expect that the sum of the ranks for each group would be the average of the sum of all the ranks $\frac{N(N + 1)/2}{k}$. In the example in Table 10–6, there are 36 possible ranks, and the total sum of the ranks is (36(36+1)/2), or 666. Therefore, we expect the sum of the ranks for each group to be 222. The Kruskal-Wallis test statistic Q, also referred to as the H statistic, is a function of the weighted sum of squares of the deviations of the actual rank sums for each group from the expected rank sums for each group:

$$Q = 12/N(N + 1)\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(N + 1)$$

Table 10–6 Calculation of the Kruskal-Wallis *Q* Statistic for Physician Satisfaction Levels in Hospitals A, B, and C

Hospital A			*Hospital B*			*Hospital C*		
Phys	*Score*	*Rank 1*	*Phys*	*Score*	*Rank 2*	*Phys*	*Score*	*Rank 3*
1	7	7	13	15	24	25	10	10.5
2	6	5.5	14	14	19.5	26	20	36
3	4	2.5	15	15	24	27	18	35
4	11	12.5	16	16	28.5	28	13	15.5
5	16	28.5	17	16	28.5	29	14	19.5
6	4	2.5	18	16	28.5	30	8	8
7	10	10.5	19	14	19.5	31	17	33
8	6	5.5	20	12	14	32	17	33
9	9	9	21	14	19.5	33	14	19.5
10	3	1	22	17	33	34	14	19.5
11	13	15.5	23	16	28.5	35	15	24
12	5	4	24	11	12.5	36	16	28.5
	$\Sigma R_A = 104$			$\Sigma R_B = 280$			$\Sigma R_C = 282$	
	$\bar{R}_A = 8.67$			$\bar{R}_B = 23.33$			$\bar{R}_C = 23.5$	

The value of the Q is a special case of the χ^2 distribution with $k - 1$ degrees of freedom.

Referring to the procedure outlined in Exhibit 10–10, and extending the physician satisfaction problem to three samples instead of two, we can calculate the Kruskal-Wallis Q statistic (Table 10–6):

$$\begin{aligned} Q &= 12/N(N + 1)\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(N + 1) \\ &= 12/36(36 + 1)[(104^2/12) + (280^2/12) + (282^2/12)] - [3(36 + 1)] \\ &= [(12/1{,}332) \times (901.33 + 6{,}533.33 + 6{,}627)] - 111 \\ &= 126.55 - 111 \\ &= 15.85 \end{aligned}$$

In the Kruskal-Wallis procedure, the degrees of freedom are equal to $k - 1$, where k equals the number of groups. In our example, the number of groups is three; therefore, the df is equal to 2 ($k - 1 = 3 - 1 = 2$). With three groups and four or more cases per group, the χ^2 distribution can be used to evaluate the significance of Q_{calc}. For 2 degrees of freedom, $\alpha = 0.05$, $\chi^2_{\text{crit}} = 5.991$. Since the calculated Q exceeds the critical value, we reject the null and conclude that the physician satisfaction level varies by hospital.

The result of the Kruskal-Wallis procedure indicates that there is a significant difference in the group medians. But which groups are different from each other? Just as with the ANOVA procedure, we can determine where the differences lie by conducting multiple pairwise comparisons. The number of possible comparisons that can be made is determined by $k(k - 1)/2$; in our example we have three groups, so the total number of comparisons that can be made is three ($3[3 - 1]/2 = 3$). We would then compare hospitals A and B, A and C, and B and C. We can determine the value of the difference between the average ranks that must exist in order for the difference to be statistically significant by

$$\left|\overline{R}_i - \overline{R}_j\right| > Z_c[\sqrt{[N(N + 1)/12][1/n_i + 1/n_j}]$$

where z comes from the standard normal distribution, but not in the usual way. With c pairwise comparisons and an overall level of α, z_c is the critical value that corresponds to a right-tailed p value of $\alpha/2_c$. The value of $c = k(k - 1)/2$. Now, what does this mean? Remember that when we use the t test or z, we are comparing two independent samples, and for a two-tailed test, when $\alpha = 0.05$, we cut 0.025 off each tail to determine significance. In this situation, we are going to make three comparisons, and if we use the same approach, we may be too conservative in our decision-making, which could result in a type I error.

A more liberal method for making comparisons in this situation is called the Bonferroni test for inequality. In this method, we set a more liberal alpha—that is, 0.10—to better detect difference between groups. We then have

$$\alpha/k = 0.10/3 = 0.033.$$

We then divide $\alpha/2$, as usual for our nondirectional test, which results in $0.033/2 = 0.017$. The critical value of z, $\alpha = 0.017$, is 2.128 (from Appendix B, Table B–1). The absolute difference between the average of the ranks that must be achieved is:

$$\begin{aligned}
&\left|\overline{R}_i - \overline{R}_j\right| > Z_c[\sqrt{[N(N + 1)/12][1/n_i + 1/n_j}] \\
&= 2.128\ \sqrt{[36(36 + 1)/12][1/12 + 11/12]} \\
&= 2.128\ \sqrt{111(0.083 + 0.083)} \\
&= 2.128(4.3) \\
&= 9.15
\end{aligned}$$

Comparing the absolute values of the average ranks, we have:

$$|\overline{R}_A - \overline{R}_B| = |8.67 - 23.33| = 14.66$$
$$|\overline{R}_A - \overline{R}_C| = |8.67 - 23.50| = 14.83$$
$$|\overline{R}_C - \overline{R}_B| = |23.50 - 23.33| = 0.17$$

Thus we determine that the median for hospital A is significantly different from the medians for hospitals B and C.

Carrying out these calculations by hand can become quite tedious. We can use SPSS to carry out the calculations for us. From the "Statistics" menu, we select "nonparametric statistics." To conduct the Kruskal-Wallis test, we select the option for "k independent samples." The grouping variable is "hospital," and the test variable is "score." Note that SPSS (Exhibit 10–11) gives the average rank for each group rather than the sum of the ranks for each group. However, the resulting Kruskal-Wallis χ^2 statistic is 15.851. SPSS does not provide post hoc procedures for the Kruskal-Wallis statistic.

Exhibit 10–11 SPSS Output for Kruskal-Wallis Test

Descriptive Statistics

	N	*Mean*	*Std. Deviation*	*Minimum*	*Maximum*
SCORE	36	12.3889	4.49938	3.00	20.00
Hospital	36	2.0000	.82808	1.00	3.00

Ranks

	Hospital	*N*	*Mean Rank*
SCORE	A	12	8.67
	B	12	23.33
	C	12	23.50
	Total	36	

Test Statistics[a,b]

	SCORE
Chi-Square	15.851
df	2
Asymp. Sig.	.000

[a] Kruskal-Wallis Test.
[b] Grouping Variable: Hospital.

✓ To Obtain the Kruskall-Wallis Statistic Using SPSS:

- From the menus, choose "Analyze"
  →Nonparametric
  →*k* independent samples* Select one or more numeric variables
- Select a grouping variable
  →Click "define range" to specify minimum and maximum integer values for the grouping variable

CONCLUSION

In this chapter, we have explored nonparametric statistical procedures other than chi square. Nonparametric procedures are less restrictive than their parametric counterparts. Nonparametric procedures do not make assumptions regarding the underlying population distribution nor do they require large sample sizes. Nonparametric methods may also be used for the nominal and ordinal scales of measurement.

For assessing relationships between variables, we can use the Spearman rho correlation coefficient, which is analogous to the Pearson *r* product moment correlation coefficient. If we are interested in making comparisons between two independent samples, we can use the Mann-Whitney Wilcoxon test as a substitute for the Student's *t* and ANOVA parametric procedures. For paired samples, nonparametric procedures include the sign test and the Wilcoxon signed ranks test. The corresponding parametric procedure is Student's paired *t* test. If we are interested in comparing three or more independent samples, the Kruskal-Wallis test is the nonparametric alternative to the ANOVA procedure.

ADDITIONAL RESOURCES

Campbell, M.J. (1997). *Statistics at Square One.* Ninth Edition. Southhampton, UK: University of Southhampton. http://bmj.bmjjournals.com/collections/statsbk.

Gibbon, J.D. 1993. *Nonparametric statistics: An introduction.* Newbury Park, CA: Sage Publications.

Rice Virtual Lab in Statistics. Hyperstat Online Textbook. http://davidmlane.com/hypestat.

Jekel, J.F., et al. 1996. *Epidemiology, biostatistics and preventive medicine.* Philadelphia: W.B. Saunders.

Pett, M.A. 1997. *Nonparametric statistics for health care research: Statistics for small samples and unusual distribution.* Thousand Oaks, CA: Sage Publications.

Statsoft Electronic Textbook. 1998. www.statsoft.com/textbook/.

Appendix 10–A

Exercises for Solving Problems

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
2. Compare the Pearson *r* with the Spearman rho correlation coefficient.
3. Under what conditions is it appropriate to use the sign test?
4. Under what conditions is it appropriate to use the Wilcoxon signed ranks test?
5. Under what conditions is it appropriate to use the Mann-Whitney Wilcoxon test?
6. Under what conditions is it appropriate to use the Kruskal-Wallis test?

MULTIPLE CHOICE

1. The Spearman rho rank order correlation coefficient
   a. is used for ordered data.
   b. has the same interpretation as the Pearson *r*.
   c. may be used when one sample is ordered and the second sample is at the ratio level of measurement.
   d. a and b
   e. all of the above
2. The sign test is an alternative approach to test a hypothesis about:
   a. a single mean
   b. the difference between matched pairs
   c. the difference between two independent samples
   d. the difference between three or more independent samples

3. The sign test is not as likely to detect statistical significance as its parametric counterpart. This means that:
   a. the probability of making a type II error is increased
   b. the sign test is less powerful than its parametric counterpart
   c. the sign test is not as sensitive as its parametric counterpart
   d. all of the above

4. Nonparametric procedures:
   a. are appropriate for ordered data
   b. have less power than their parametric counterparts
   c. have fewer assumptions regarding the underlying population distribution than their parametric counterparts
   d. b and c
   e. all of the above

5. The parametric counterpart of the Kruskal-Wallis procedure is the:
   a. paired *t* test
   b. independent sample *t* test
   c. *t* test for two independent samples
   d. analysis of variance (ANOVA)

6. The sum of a set of ranks may be found by:
   a. $n(n - 1) / 2$
   b. $n(n + 1) / 2$
   c. $(n - 1) / 2$
   d. $(n - 1)2$

PROBLEMS

1. You are analyzing length of stay by physician for DRG 124, Circulatory Disorders, Except Acute Myocardial Infarction with Cardiac Catheterization and Complex Diagnosis. You are focusing on physicians 2050, 2210, and 8290. The lengths of stay for the patients of these three physicians appear in Table 10–A–1. Since the sample size for each physician is small, you decide to conduct the Kruskal-Wallis test to compare the mean lengths of stay.
   a. State the null and alternative hypotheses.
   b. State the alpha level.
   c. What is the result of the Kruskal-Wallis test?
   d. State your conclusions.

Table 10–A–1 LOS Physician Crosstabulation (SPSS Output)

Count		*Physician*			
		2050	*2210*	*8290*	*Total*
LOS	1	1	2	2	5
	2	7	4	2	13
	3	1	2	1	4
	4	4	2	2	8
	5	0	1	0	1
	6	0	1	2	3
	8	0	1	0	1
Total		13	13	9	35

2. You want to determine if more men than women are discharged from DRG 127, Heart Failure and Shock. You believe that since more men than men suffer from heart disease, that more men should be discharged from this DRG. The frequency distribution of discharges by sex appears in Table 10–A–2. You decide to use the sign test (binomial test) to determine if there is a difference in the proportion of discharges by sex for DRG 127.
   a. State the null and alternative hypotheses.
   b. State the alpha level.
   c. What is the result of the sign test?
   d. State your conclusions.

Table 10–A–2 Frequency Distribution of Discharges from DRG 127 by Gender

		Frequency	*Percent*	*Valid Percent*	*Cumulative Percent*
Valid	FEMALE	20	32.8	32.8	32.8
	MALE	41	67.2	67.2	100.0
	Total	61	100.0	100.0	

3. Review Exhibits 10–A–1 and 10–A–2 for discharges from DRG 127, Heart Failure and Shock. Use the Mann-Whitney *U* test to determine if there is a difference in age by sex and length of stay by sex for discharges from DRG 127.
   a. State the null and alternative hypotheses.
   b. State the alpha level.
   c. What are the results of the Mann-Whitney *U* tests?
   d. State your conclusions.
   e. Use the ANOVA procedure to run the same analyses. Compare the ANOVA results with the Mann-Whitney *U* test results.

Exhibit 10–A–1 Case Summaries for Female Discharges from DRG 127 (SPSS Output)

			PDX	*Age*	*LOS*
Gender	Female	1	404.91	83	4
		2	404.93	88	4
		3	402.91	88	6
		4	428.0	82	4
		5	428.40	80	10
		6	428.0	80	1
		7	428.0	78	4
		8	428.0	76	8
		9	428.0	72	3
		10	428.0	66	36
		11	428.0	64	6
		12	428.0	64	6
		13	428.0	63	3
		14	428.0	61	15
		15	428.0	60	3
		16	428.0	59	2
		17	428.0	57	27
		18	428.0	57	11
		19	428.0	56	8
		20	402.01	54	13
	Total	N	20	20	20

Exhibit 10–A–2 Case Summaries for Male Discharges from DRG 127 (SPSS Output)

			PDX	*Age*	*LOS*
Gender	Male	1	428.0	87	10
		2	428.0	86	3
		3	428.0	85	5
		4	428.0	84	4
		5	428.0	83	2
		6	428.0	80	3
		7	428.0	80	2
		8	428.0	79	5
		9	428.0	77	3
		10	428.0	77	1
		11	428.0	77	3
		12	428.0	76	14
		13	428.0	75	17
		14	428.0	75	2
		15	428.0	73	3
		16	428.0	73	11

(Continued)

Exhibit 10–A–2 Case Summaries for Male Discharges from DRG 127 (SPSS Output) (Continued)

			PDX	*Age*	*LOS*
Gender	Male	17	428.0	70	6
		18	428.0	69	5
		19	428.0	66	7
		20	428.0	65	3
		21	428.0	65	5
		22	428.0	64	3
		23	428.43	64	1
		24	404.91	63	3
		25	428.0	63	3
		26	428.0	60	5
		27	428.0	54	1
		28	428.0	53	10
		29	428.0	52	2
		30	428.0	51	3
		31	428.0	51	1
		32	428.0	51	2
		33	428.0	49	1
		34	428.0	47	8
		35	404.03	47	3
		36	428.0	47	10
		37	428.0	42	8
		38	428.0	39	11
		39	428.0	39	3
		40	404.91	37	3
		41	428.0	37	16
Total	N		41	41	41

APPENDIX A

Glossary

Age-adjusted death rate—The crude death rate is adjusted when population proportions by age are different; this adjustment eliminates the effects of different age distributions in different populations; the crude death rate is adjusted when the death rates of two populations are to be compared.

Age-specific death rate (ASDR)—The total number of deaths for a given age group for a certain time frame divided by the estimated population for the same age group and for the same time frame; the ASDR is usually expressed as the number of deaths per 10^n depending on the size of the population.

Alpha error—See type I error.

Alpha level—The point at which the null hypothesis will be rejected. The alpha level, which is set prior to conducting the research, is usually set at 0.05 for small sample sizes and at 0.01 for larger sample sizes.

Alternative hypothesis—See hypothesis testing.

Analysis of variance (ANOVA)—A statistical method for comparing the differences between two or more means for statistical significance; the independent variable is usually nominal, and the dependent variable is at the ratio or interval level of measurement.

Apparent limits—See class interval.

Asymptotic curve—A property of the normal distribution in which the tails of the curve approach the *x*-axis but never touch it.

Attributable risk—A measure of the impact of a disease on a population; measures additional risk of illness as a result of exposure to the risk factor.

Bar chart— A graphic technique for displaying discrete or nominal-level data; one or more variables may be displayed in a bar chart; a bar chart that displays two or more variables is called a grouped bar chart.

Beta error—See type II error.

Between-group variance (SSB)—See sum of squares between.

Box head—In a table, the box head contains the column headings.

Case fatality rate—The total number of deaths due to a specific illness during a given time period divided by the total number of cases during the same time period.

Cause-specific death rate—The total number of deaths due to a specific cause during a given time interval divided by the estimated mid-interval population.

Cell—In a table, a cell is where the row and column variables intersect.

Central limit theorem—In a repeated number of random samples of size *N* drawn from a population, the distribution of the sample means approaches the normal distribution as *N* becomes large. This occurs even when the population distribution is not normal.

Chi-square goodness-of-fit test—See χ^2 goodness-of-fit test.

Chi-square test of independence—See χ^2 test of independence.

Class interval—A method used to classify interval/ratio level data into categories for analysis; the limits of the class intervals are referred to as either "apparent" or "real." The apparent limits are the upper and lower boundaries of the class interval. For example, the upper and lower limits of two successive intervals may be 18–22 and 23–27. The real limits depict the continuous nature of the frequency distribution, indicating that there are no gaps between successive intervals. For example, the real limits for previously stated intervals are 17.5–22.5 and 22.5–27.5. The upper limit of one class interval is the lower limit for the next class interval.

Cluster sampling—A sampling technique in which the sampling units are groups rather than individuals.

Coefficient alpha—See Cronbach's alpha.

Coefficient of determination—The square of the Pearson *r*. States how much of the variation in the dependent variable *Y* is explained by the independent variable *X*.

Confidence interval—Calculated from the standard error of the mean, it is an estimate of the true limits within which the true population mean lies; the range of values that may reasonably contain the true population mean.

Confounding—The relationship between two variables is so close that the effects of either variable cannot be separated from the other.

Confounding factor—See confounding variable.

Confounding variable—A factor or variable that contributes differentially across the categories or levels of another variable; confusion of two independent variables so that the effect of one variable cannot be differentiated from the effect of the other.

Construct validity—A type of validity that is the link between a theory and the property being measured; a measurement instrument with construct validity is representative of the property of interest.

Content validity—The adequacy of the sample or number of items used to represent the content area being measured; content validity is a matter of judgment and is evaluated by a panel of experts.

Contingency coefficient—A measure of association between two nominal level variables; it is an alternative to the phi coefficient when one variable has more than two categories; the range of the contingency coefficient is 0 to 1.

Contingency table—A table that displays the relationship between two variables, whether the distribution of one variable is dependent on or related to the distribution of a second variable.

Continuous variable—A measure taken on the interval or ratio level of measurement.

Cramer's V—A statistic used to adjust the χ^2 statistic for sample size. It is a measure of correlation coefficient with a range of 0.0 to 1.0. It is used in place of the phi coefficient when the contingency table is greater than 2×2.

Criterion-related validity—A type of validity in which a measuring instrument correlates with a criterion known to accurately measure the property of interest.

Critical region—In hypothesis testing, the portion of the test statistic distribution that is equal to or beyond the critical value of the test statistic; the region of the *z, t, F,* or χ^2 distribution that results in rejection of the null hypothesis.

Cronbach's alpha—Also known as coefficient alpha. A measure of internal consistency. Cronbach's alpha has a range of 0.0 to 1.0; the minimum acceptable criterion for internal consistency is 0.70.

Crude death rate—The total number of deaths in a given population for a given time period divided by the mid-interval population for the same time period; the crude death rate is usually expressed as the number of deaths per 1,000 population.

Death-to-case ratio—The total number of deaths due to a specific disease during a given time period divided by the number of new cases of the disease reported during the same time period.

Degrees of freedom—The number of observations in a data set that are free to vary after the mean of the distribution has been determined.

Dichotomous variable—A variable that falls on the nominal scale of measurement but is limited to only two categories.

Direct standardization—When mortality rates of two populations are compared, each population is assigned the same standard population proportion for each age group; the standard population proportion is multiplied by the age-specific death rate for each age group in each population; the sum results in the age-adjusted death rate.

Discrete variable—A dichotomous or nominal variable whose values are placed into categories.

Effect—A change in one variable that may be associated with a change in the second variable.

***F* test—**The ratio of the between-group variance (SSB) to the within-group variance (SSW) in the ANOVA procedure. If the *F* ratio is statistically significant, the observed differences between the group means of the independent variables under study will be significantly different from each other.

Fisher's exact test—A statistical test that is used as a substitute for the χ^2 test of independence when the frequencies in a contingency table are less than or equal to 20.

Footnote—See note.

Frequency distribution—A table or graph that displays the number of times (frequency) a particular observation occurs.

Frequency polygon—A line graph of a frequency distribution of interval or ratio level data.

Grouped bar chart—Used to illustrate data from a two- or three-variable table when an outcome variable has only two categories; the bars within a group are adjoining.

Histogram—A graphic technique used to display the frequency distribution of interval or ratio level data; the frequency distribution can be displayed as either numbers or percentages in a series of bars.

Hypothesis testing—A statement regarding the research question to be tested. There are two forms of the hypothesis: the null hypothesis and the alternative hypothesis. The null hypothesis states that there is no difference between the population means or proportions that are being compared, or that there is no association between the two variables that are being compared. The alternative hypothesis states that there is a significant difference in the population means or proportions that are being compared, or that there is an association between the two variables that are being compared. The alternative hypothesis usually states what the researcher believes to be true regarding the problem under study.

Incidence rate—The number of new cases of a particular disease in a population for a given time period divided by the average population for the same time period.

Individuality—For nominal-level data, the circumstance in which the number of observations in a given category is limited to one.

Infant mortality rate—The number of deaths of persons under one year of age during a given time period divided by the number of live births reported during the same time period.

Intercept—Represented by "*a*" in the regression model. It is the point at which the regression line crosses the *y*-axis. The intercept is the average value of *Y* when *X* is equal to zero.

Internal consistency—The extent to which the items on a measuring instrument are consistent with one another; Cronbach's alpha is used to evaluate internal consistency.

Interrater agreement—See interrater reliability.

Interrater reliability—The percentage of agreement between raters using the same measuring instrument; the kappa coefficient is used to measure interrater agreement.

Interval scale—Similar to the ratio scale of measurement, but there is no true zero; intervals between successive intervals are equal and continuous.

Kaplan Meier survival analysis—A statistical method used to analyze the survival time of individuals with a specific disease; it is most often used in analyzing the survival time of cancer patients.

Kappa coefficient—A measure of agreement, beyond what would occur by chance, between two raters on the same measuring instrument. The Kappa coefficient ranges from 0.0 to 1.0.

Kruskal-Wallis test—The nonparametric counterpart to the ANOVA procedure; it is used to compare two or more groups of ordinal data.

Kurtosis—The vertical stretching of a frequency distribution.

Level of significance (or significance level)—A cutoff value for evaluating the *p* value that results from a statistical test; indicates the level of risk that we are willing to take for rejecting the null hypothesis when it is true.

Line graph—A graphic technique that consists of a line connecting a series of points on an arithmetic scale; a line graph is often used to display time trends and survival curves; a line graph does not represent a frequency distribution.

Linear regression—A statistical test used to measure the strength of the linear relationship between two variables; the variables under study are at the interval or ratio level of measurement.

Line of best fit—See regression line.

Mann-Whitney *U* test—also known as the Mann-Whitney Wilcoxon test. The nonparametric counterpart to the Student's *t* test for two independent samples; used to compare two groups of ordinal-level data.

Mann-Whitney Wilcoxon Test—See Mann-Whitney *U* Test.

Maternal mortality rate—The number of deaths assigned to pregnancy-related causes during a given time period divided by the total number of live births during the same period.

McNemar test—The nonparametric test used for paired data; analogous to the paired *t* test.

Mean—A measure of central tendency; arithmetic average of the observations in a frequency distribution; the sum of the values of the observations in a frequency distribution divided by the total number of observations.

Mean square—Also known as the variance. The sum of squares divided by the appropriate number of degrees of freedom.

Measurement—The process of measuring an attribute or property of a person, object, or event according to a particular set of rules; the set of rules is used to assign numbers to the attribute or property being measured.

Measures of central tendency—A single value that summarizes a frequency distribution or illustrates the most typical value in a frequency distribution; measures of central tendency are the mean, median, and the mode. The measure selected to represent the data set depends on the characteristics and shape of the distribution.

Measures of variation—Describes how much spread there is in a frequency distribution; measures of variation include the range, standard deviation, and the variance.

Median—A measure of central tendency; midpoint of a frequency distribution when the observations have been arranged in order from lowest to highest; point at which 50% of the observations fall above and 50% of the observations fall below.

Mode—A measure of central tendency; the most frequently occurring observation in a frequency distribution.

Mortality rate—See crude death rate.

Multicollinearity—In multiple regression, when two independent variables are highly correlated with one another, and it is difficult to separate the effects of each independent variable on the dependent variable.

Multiple regression—A statistical procedure used to explain the effects of two or more independent variables on a dependent variable; it is an extension of the linear regression model.

Neonatal death rate—The number of deaths of newborns under 28 days of age during a given time period divided by the number of live births during the same time period.

Neonatal mortality rate—See neonatal death rate.

Nominal scale—A level of measurement in which the frequencies of observations on variables are placed into categories.

Noncritical region—In hypothesis testing, the area of the test statistic distribution that is between the critical values of the test statistic; if the calculated value of the test statistic falls within this region, we fail to reject the null hypothesis.

Nonparametric statistical tests—Tests used to test for statistical significance when the underlying population distribution does not meet the requirements of the normal distribution or when sample sizes are small.

Normal distribution—A continuous frequency distribution characterized by a bell-shaped curve; a normal distribution is symmetrical with 50% of the values falling above the mean and 50% of the values falling below the mean.

Note—Explanation that appears below a table, chart, or graph to explain symbols or abbreviations that may have been used in the table, chart, or graph.

Null hypothesis—See hypothesis testing.

Odds ratio—A relative measure of occurrence of an illness; the odds of exposure in a diseased group divided by the odds of exposure in a non-diseased group.

One-sample *t* test—See *t* tests.

100% component bar chart—A variant of the stacked bar chart; all bars in the chart are of the same height; each bar displays the variable categories as percents of the total number; each bar is like its own pie chart.

One-tailed test (directional test)—In hypothesis testing, the researcher is trying to determine if the sample mean is greater than or less than the population parameter.

Ordinal scale—Scale of measurement in which measures are placed into ordered categories or measures are ranked in some predetermined order such as lowest score on a test to highest score on a test; the width between categories or ranks may not be equal.

Outliers—Extreme values in a frequency distribution.

***p* value**—The probability that the observed difference could have been obtained by chance alone, given random variation and a single test of the null hypothesis.

Paired *t* test—See *t* test.

Parametric statistical tests—Statistical tests of significance used when the underlying population distributions are assumed to be normal.

Pearson *r* correlation coefficient—A measure of the strength of the linear association between two variables that fall on the interval or ratio level of measurement. The Pearson *r* correlation coefficient has a range of -1.0 to $+1.0$; a negative correlation indicates that the variables change in opposite direction to one another, a positive correlation that the variables change in the same direction and a correlation of 0.0 that there is no relationship between the two variables.

Percentile Rank—The proportion of scores in a distribution that a specific score is greater than or equal to. For example, if a hospital ranks in the 98th percentile in terms of severity of illness it means that the hospital has a patient population sicker than or equal to 98% of the hospitals ranked.

Phi coefficient—A measure that indicates the degree of association between two nominal level variables; the range of the phi coefficient is 0 to 1; it has the same interpretation as the Pearson *r* product moment correlation coefficient.

Pie chart—A graphic technique in which the proportions of a nominal variable are displayed as "pieces" of the pie.

Point estimate—The numerical value calculated from a sample that is assumed to best represent the population parameter.

Point prevalence rate—The number of current cases, both new and old, of a specified disease at a given point in time compared to the estimated population at the same point in time; the point prevalence rate is usually expressed as the number of cases per 10^n, depending on the size of the population.

Polarization—The maximum spread or variability in a frequency distribution.

Population—All members of a group that is under study; the group to which the sample results are generalizable; members of the group share some measurable characteristic.

Population parameter—A measure that results from the compilation of data from a population.

Post hoc procedures—Statistical follow-up tests following the ANOVA procedure when three or more means are being compared; the post hoc test indicates whether all means compared are significantly different from each other or if only several are significantly different from each other.

Postneonatal mortality rate—The number of deaths of persons aged 28 days up to and not including one year during a given time period divided by the number of live births for the same time period.

Predictive value—The number of cases correctly identified by a measure out of the total number of cases with the property of interest.

Prevalence rate—The number of cases of a particular disease in a population for a given time period divided by the population for the same time period.

Proportion—A particular type of ratio; in a proportion, the numerator is always included in the denominator.

Proportionate mortality ratio (PMR)—The total number of deaths due to a specific cause during a given time period divided by the number of deaths due to all causes. Proportionate mortality is not a rate because the denominator is the number of deaths during the time period, not the population size during the time period.

Race-specific death rate—The number of deaths in a specific ethnic group for a given timeframe compared to the estimated population total for the same ethnic group for the same time frame; the race-specific death rate is usually expressed as the number of deaths per 10^n, depending on the size of the population.

Range—A measure of variability; the difference between the smallest and largest values in a frequency distribution.

Rate—A measure used to compare an event over time; comparison of the number of times an event did happen (numerator) to the number of times an event could have happened (denominator); rates are expressed as the event of interest per 100, 1,000, 10,000, or 100,000 cases.

Ratio—A comparison of categories of dichotomous variables either to each other (e.g., male discharges to female discharges), or of one category to the whole (e.g., male discharges to total discharges).

Ratio scale—The highest level of measurement; intervals between successive intervals are equal and continuous; measurement scale with a true zero.

Real limits—See class interval.

Regression line—The slope in the regression model. It is the straight line that best fits all of the data points in the regression problem. If the correlation is perfect, the regression line will go through all of the data points simultaneously.

Relative risk—See risk ratio.

Reliability—A characteristic of a measuring instrument that results in consistent measures over repeated trials; measurement results are approximately the same on repeated trials.

Risk ratio—Also called relative risk. A ratio that compares the risk of disease between two groups.

Sample—Items drawn from a population for a study; ideally, members of the sample are drawn randomly and independently from the population of interest; a subset of the population under study.

Sampling method—The process of selecting individuals from a larger group or population in such a way that the resultant sample is representative of the underlying population.

Scales of measurement—Nominal, ordinal, interval, and ratio scales of measurement. The level at which data are collected determines the types of statistics that may be used to describe the population under study.

Sample statistic—The measures that result from analysis of data compiled from samples.

Scatter diagram—A graphic representation of the *X* and *Y* variables. The *X* variable is plotted on the horizontal axis and the *Y* variable is plotted on the vertical axis. It is used to

assess the linearity of the relationship between variables *X* and *Y* when conducting either the Pearson *r* or simple linear regression. If the two variables appear to approximate a straight line, the variables are linearly related.

Scheffé test—A post hoc test used in the analysis of variance procedure when there are three or more sample means being compared. It is a test used to determine which of the means are significantly different from each other.

Sensitivity—An aspect of data accuracy (validity); a measure is sensitive if it identifies the property of interest when that property is truly present.

Sex-specific death rate—The number of male or female deaths for a given time frame compared to the estimated male or female population total for the same time frame; the sex specific death rate is usually expressed as the number of deaths per 10^n, depending on the size of the population.

Sign test—The nonparametric counterpart to the paired-sample *t* test; compares whether one group did better than another group.

Simple random sampling—A sampling technique in which each member of a population has an equal chance of being included in the sample.

Skewness—The horizontal stretching of a frequency distribution.

Slope—In the regression model, the slope (*b*) represents the average change in *Y* that is associated with *X*. The greater the slope, the greater the change in *Y* that is associated with a change in *X*, and the greater the relationship between *X* and *Y*.

Source—A statement following a table, chart, or graph that indicates the resource used to generate the table, chart, or graph.

Spearman rho—The nonparametric counterpart of the Pearson *r* used when one of the variables is ordered; the interpretation of the Spearman rho is the same as that for the Pearson *r*.

Specificity—An aspect of data accuracy (validity); a measure that is specific excludes cases when the property of interest is truly absent.

SSB—See sum of squares between.

SSW—See sum of squares within.

Stability—A type of reliability in which the same or similar results are obtained on repeated measures by administering the same instrument to the same group on two different occasions to obtain a reliability coefficient that ranges from 0.0 (no reliability) to 1.0 (perfect reliability).

Stacked bar chart—A type of bar chart in which the categories of a nominal level variable are stacked like building blocks on top of one another to form a single bar; the bar represents the total number of cases that occurred in the category and the segments represent the frequencies within the category.

Standard deviation—A measure of variability; square root of the variance; describes deviation from the mean in terms of the original unit of measurement (e.g., height, age, blood pressure).

Standard error of the estimate—In linear regression, a measure of the scatter or spread of the observed values of *Y* around the corresponding values of *Y* estimated from the regression equation.

Standard error of the mean—A measure of how close the sample mean is to the population mean; it is influenced by the sample size and standard deviation.

Standard mortality rate—See standard mortality ratio.

Standard mortality ratio (SMR)—Compares the actual number of deaths in a group or population under study compared to the expected number of deaths based on standard population death rates applied to the study group or population; this measure is always multiplied by 100.

Standard normal deviate—In the standard normal distribution, the distance between the observed value and the mean, μ; it is also referred to as the *z* score or *z* value.

Standard normal distribution—A normal distribution with a mean equal to zero and a standard deviation equal to one.

Standardized residuals—In the χ^2 test of independence, a residual is the difference between the observed cell frequencies and the expected cell frequencies. To obtain the standardized residual, the difference between the observed and expected cell frequencies for each cell is divided by the square root of the expected cell frequency. An obtained standardized residual of greater than $+2$ or less than -2 is an indication that the cell in question is an important contributor to the calculated χ^2.

Statistical power analysis—Statistical power analysis assists the researcher in determining the appropriate sample size for conducting a statistical test while controlling for both type I and type II error.

Statistical significance—In hypothesis testing, when a statistical test results in a *p* value that is less than or equal to the preset alpha level; this is usually set at 0.05 for small samples and 0.01 for large samples. The interpretation of the *p* value is that the result obtained would occur by chance no more than 5 out of 100 times when $p = 0.05$, or 1 out of 100 times when $p = 0.01$.

Stratified random sampling—A sampling technique in which each stratum within a population is proportionately represented in the sample.

Stub—The row captions in a table.

Sum of squares between (SSB)—In the analysis of variance procedure, the SSB is the variation of the sample means around the grand mean.

Sum of squares within (SSW)—In the analysis of variance procedure, the SSW is the variation of the observations in each sample around their respective sample means.

Symmetrical—See symmetry.

Symmetry—A property of the normal distribution in which 50% of the observations fall above the mean and 50% of the observations fall below the mean.

Systematic sampling—A sampling technique in which every *k*th member of a population is selected for inclusion in the sample.

***t* statistic**—A statistic that follows the *t* distribution; see *t* test.

***t* test**—A parametric statistical test that compares the difference between the means of two groups; the *t* test may be used when comparing a single sample mean to a known population parameter (independent sample *t* test), when comparing the means of two independent samples (*t* test for two independent samples), or when comparing the means of matched pairs (paired *t* test).

***t* test for comparing two independent sample means**—See *t* test.

Table—A set of data arranged in rows and columns.

Table shell—Prepared prior to collection of data to show how the data will be organized and displayed after data collection; table shells are complete except for the actual data; table shells show titles, headings, and categories.

Test-retest reliability—See stability.

Timeliness—The collection, analysis, and reporting of data/information within a time frame useful for decision making.

Total sum of squares (TSS)—In the analysis of variance procedure, the TSS is the variation of all the observations (combined into one sample) around the grand mean.

Trimmed mean—The calculation of the mean or a frequency distribution after the elimination (trimming) of outliers from the distribution.

TSS—See total sum of squares.

Tukey HSD—A post hoc test used in the analysis of variance procedure when there are three or more sample means being compared. It is a test used to determine which of the sample means are significantly different from each other.

Two-tailed test (nondirectional test)—In hypothesis testing, the researcher is interested in determining whether the sample mean and the population parameter are significantly different from each other; the direction of the inequality is not an issue.

Type I error—Also called alpha error; the rejection of the null hypothesis when it is true.

Type II error—Also called beta error; failure to reject the null hypothesis when it is false.

Uniformity—Even distribution of observations in the categories of a nominal variable.

Unimodal—Property of the normal distribution in which there is only one mode.

Validity—Accuracy in measurement; a valid measuring instrument accurately measures what it is intended to measure.

Variable—A characteristic or property that may take on different values.

Variance—A measure of variability; the average of the squared deviations from the mean; measures variability in original units of measurement squared.

Weighted mean—A mean that takes into account differences in sample size; the weighted mean is equal to the sum of the means times the number of observations in each sample divided by the total number of observations in the samples combined.

Wilcoxon signed ranks test—The nonparametric counterpart of the paired *t* test; used when sample sizes are small or when the underlying population distribution is not normally distributed.

Winsorized mean—A type of mean (average) which has been adjusted for extreme values in the frequency distribution; the most extreme values (highest and lowest) in the distribution are changed to the next less extreme values.

Within-group variance (SSW)—Also known as the error term in the ANOVA procedure; a measure of variation within each group that is being compared; it is a measure of the sample observations around the sample mean.

χ^2 goodness-of-fit test—A nonparametric test used to determine whether the observed frequencies in a distribution are significantly different from the expected or theoretical frequencies in a distribution, based on the researcher's knowledge of a population understudy. The null hypothesis states that there is no difference between the observed and expected frequencies; the alternative hypothesis states that there is a difference between the observed and expected frequencies. The χ^2 goodness-of-fit test is often used to determine whether proportions in a randomly drawn sample are significantly different from the underlying or theoretical population proportions.

χ^2 test of independence—A nonparametric test used to determine whether a relationship exists between two variables in a 2×2 contingency table or an $R \times C$ table. Measures for the variables are nominal or dichotomous. The test uses frequencies of the variables, not the actual observations. The null hypothesis states that there is no relationship between the two variables; thus, they are independent of each other. The alternative hypothesis states that there is a relationship between the two variables.

Yates correction for continuity—An adjustment made to the chi-square procedures when the counts in a contingency table are small or when any expected cell count is less than five.

***z* score**—In the standard normal distribution, the number of standard deviation units that the observed value is away from the mean, μ.

***z* statistic**—The critical value of *z*; the calculated value of *z* is compared to the *z* statistic in order to determine statistical significance.

***z* test for comparing two independent population means—**See *z* tests.

***z* test for comparing two population proportions**—See *z* tests.

***z* tests**—A parametric test of statistical significance; test is used to make comparisons for two independent population means or two independent population proportions or for comparing a sample mean to a population mean when population parameters are known. It is assumed that the population distributions are normal. In a one-tailed or directional *z* test, the researcher is trying to determine if the sample mean is significantly less or significantly greater than the population parameter μ. In a two-tailed or nondirectional *z* test, the researcher is trying to determine if the sample mean is significantly different (either greater than or less) from the population parameter μ.

***z*-value—**See *z* score.

APPENDIX B

Statistical Tables

Table B–1 Areas under the Normal Curve

z	*Cum p*	*Tail p*	*z*	*Cum p*	*Tail p*	*z*	*Cum p*	*Tail p*
0.00	0.5000	0.5000	0.32	0.6255	0.3745	0.64	0.7389	0.2611
0.01	0.5040	0.4960	0.33	0.6293	0.3707	0.65	0.7422	0.2578
0.02	0.5080	0.4920	0.34	0.6331	0.3669	0.66	0.7454	0.2546
0.03	0.5120	0.4880	0.35	0.6368	0.3632	0.67	0.7486	0.2514
0.04	0.5160	0.4840	0.36	0.6406	0.3594	0.68	0.7517	0.2483
0.05	0.5199	0.4801	0.37	0.6443	0.3557	0.69	0.7549	0.2451
0.06	0.5239	0.4761	0.38	0.6480	0.3520	0.70	0.7580	0.2420
0.07	0.5279	0.4721	0.39	0.6517	0.3483	0.71	0.7611	0.2389
0.08	0.5319	0.4681	0.40	0.6554	0.3446	0.72	0.7642	0.2358
0.09	0.5359	0.4641	0.41	0.6591	0.3409	0.73	0.7673	0.2327
0.10	0.5398	0.4602	0.42	0.6628	0.3372	0.74	0.7704	0.2296
0.11	0.5438	0.4562	0.43	0.6664	0.3336	0.75	0.7734	0.2266
0.12	0.5478	0.4522	0.44	0.6700	0.3300	0.76	0.7764	0.2236
0.13	0.5517	0.4483	0.45	0.6736	0.3264	0.77	0.7794	0.2206
0.14	0.5557	0.4443	0.46	0.6772	0.3228	0.78	0.7823	0.2177
0.15	0.5596	0.4404	0.47	0.6808	0.3192	0.79	0.7852	0.2148
0.16	0.5636	0.4364	0.48	0.6844	0.3156	0.80	0.7881	0.2119
0.17	0.5675	0.4325	0.49	0.6879	0.3121	0.81	0.7910	0.2090
0.18	0.5714	0.4286	0.50	0.6915	0.3085	0.82	0.7939	0.2061
0.19	0.5753	0.4247	0.51	0.6950	0.3050	0.83	0.7967	0.2033
0.20	0.5793	0.4207	0.52	0.6985	0.3015	0.84	0.7995	0.2005
0.21	0.5832	0.4168	0.53	0.7019	0.2981	0.85	0.8023	0.1977
0.22	0.5871	0.4129	0.54	0.7054	0.2946	0.86	0.8051	0.1949
0.23	0.5910	0.4090	0.55	0.7088	0.2912	0.87	0.8078	0.1922
0.24	0.5948	0.4052	0.56	0.7123	0.2877	0.88	0.8106	0.1894
0.25	0.5987	0.4013	0.57	0.7157	0.2843	0.89	0.8133	0.1867
0.26	0.6026	0.3974	0.58	0.7190	0.2810	0.90	0.8159	0.1841
0.27	0.6064	0.3936	0.59	0.7224	0.2776	0.91	0.8186	0.1814
0.28	0.6103	0.3897	0.60	0.7257	0.2743	0.92	0.8212	0.1788
0.29	0.6141	0.3859	0.61	0.7291	0.2709	0.93	0.8238	0.1762
0.30	0.6179	0.3821	0.62	0.7324	0.2676	0.94	0.8264	0.1736
0.31	0.6217	0.3783	0.63	0.7357	0.2643	0.95	0.8289	0.1711

continues

Table B–1 continued

z	*Cum p*	*Tail p*	*z*	*Cum p*	*Tail p*	*z*	*Cum p*	*Tail p*
0.96	0.8315	0.1685	1.44	0.9251	0.0749	1.92	0.9726	0.0274
0.97	0.8340	0.1660	1.45	0.9265	0.0735	1.93	0.9732	0.0268
0.98	0.8365	0.1635	1.46	0.9279	0.0721	1.94	0.9738	0.0262
0.99	0.8389	0.1611	1.47	0.9292	0.0708	1.95	0.9744	0.0256
1.00	0.8413	0.1587	1.48	0.9306	0.0694	**1.96**	**0.9750**	**0.0250**
1.01	0.8438	0.1562	1.49	0.9319	0.0681	1.97	0.9756	0.0244
1.02	0.8461	0.1539	1.50	0.9332	0.0668	1.98	0.9761	0.0239
1.03	0.8485	0.1515	1.51	0.9345	0.0655	1.99	0.9767	0.0233
1.04	0.8508	0.1492	1.52	0.9357	0.0643	2.00	0.9772	0.0228
1.05	0.8531	0.1469	1.53	0.9370	0.0630	2.01	0.9778	0.0222
1.06	0.8554	0.1446	1.54	0.9382	0.0618	2.02	0.9783	0.0217
1.07	0.8577	0.1423	1.55	0.9394	0.0606	2.03	0.9788	0.0212
1.08	0.8599	0.1401	1.56	0.9406	0.0594	2.04	0.9793	0.0207
1.09	0.8621	0.1379	1.57	0.9418	0.0582	2.05	0.9798	0.0202
1.10	0.8643	0.1357	1.58	0.9429	0.0571	2.06	0.9803	0.0197
1.11	0.8665	0.1335	1.59	0.9441	0.0559	2.07	09808	0.0192
1.12	0.8686	0.1314	1.60	0.9452	0.0548	2.08	0.9812	0.0188
1.13	0.8708	0.1292	1.61	0.9463	0.0537	2.09	0.9817	0.0183
1.14	0.8729	0.1271	1.62	0.9474	0.0526	2.10	0.9821	0.0179
1.15	0.8749	0.1251	1.63	0.9484	0.0516	2.11	0.9826	0.0174
1.16	0.8770	0.1230	**1.64**	**0.9495**	**0.0505**	2.12	0.9830	0.0170
1.17	0.8790	0.1210	1.65	0.9505	0.0495	2.13	0.9834	0.0166
1.18	0.8810	0.1190	1.66	0.9515	0.0485	2.14	0.9838	0.0162
1.19	0.8830	0.1170	1.67	0.9525	0.0475	2.15	0.9842	0.0158
1.20	0.8849	0.1151	1.68	0.9535	0.0465	2.16	0.9846	0.0154
1.21	0.8869	0.1131	1.69	0.9545	0.0455	2.17	0.9850	0.0150
1.22	0.8888	0.1112	1.70	0.9554	0.0446	2.18	0.9854	0.0146
1.23	0.8907	0.1093	1.71	0.9564	0.0436	2.19	0.9857	0.0143
1.24	0.8925	0.1075	1.72	0.9573	0.0427	2.20	0.9861	0.0139
1.25	0.8944	0.1056	1.73	0.9582	0.0418	2.21	0.9864	0.0136
1.26	0.8962	0.1038	1.74	0.9591	0.0409	2.22	0.9868	0.0132
1.27	0.8980	0.1020	1.75	0.9599	0.0401	2.23	0.9871	0.0129
1.28	0.8997	0.1003	1.76	0.9608	0.0392	2.24	0.9875	0.0125
1.29	0.9015	0.0985	1.77	0.9616	0.0384	2.25	0.9878	0.0122
1.30	0.9032	0.0968	1.78	0.9625	0.0375	2.26	0.9881	0.0119
1.31	0.9049	0.0951	1.79	0.9633	0.0367	2.27	0.9884	0.0116
1.32	0.9066	0.0934	1.80	0.9641	0.0359	2.28	0.9887	0.0113
1.33	0.9082	0.0918	1.81	0.9649	0.0351	2.29	0.9890	0.0110
1.34	0.9099	0.0901	1.82	0.9656	0.0344	2.30	0.9893	0.0107
1.35	0.9115	0.0885	1.83	0.9664	0.0336	2.31	0.9896	0.0104
1.36	0.9131	0.0869	1.84	0.9671	0.0329	2.32	0.9898	0.0102
1.37	0.9147	0.0853	1.85	0.9678	0.0322	**2.33**	**0.9901**	**0.0099**
1.38	0.9162	0.0838	1.86	0.9686	0.0314	2.34	0.9904	0.0096
1.39	0.9177	0.0823	1.87	0.9693	0.0307	2.35	0.9906	0.0094
1.40	0.9192	0.0808	1.88	0.9699	0.0301	2.36	0.9909	0.0091
1.41	0.9207	0.0793	1.89	0.9706	0.0294	2.37	0.9911	0.0089
1.42	0.9222	0.0778	1.90	0.9713	0.0287	2.38	0.9913	0.0087
1.43	0.9236	0.0764	1.91	0.9719	0.0281	2.39	0.9916	0.0084

continues

Table B–1 continued

z	*Cum p*	*Tail p*	*z*	*Cum p*	*Tail p*	*z*	*Cum p*	*Tail p*
2.40	0.9918	0.0082	2.80	0.9974	0.0026	3.20	0.9993	0.0007
2.41	0.9920	0.0080	2.81	0.9975	0.0025	3.21	0.9933	0.0007
2.42	0.9922	0.0078	2.82	0.9976	0.0024	3.22	0.9994	0.0006
2.43	0.9925	0.0075	2.83	0.9977	0.0023	3.23	0.9994	0.0006
2.44	0.9927	0.0073	2.84	0.9977	0.0023	3.24	0.9994	0.0006
2.45	0.9929	0.0071	2.85	0.9978	0.0022	3.25	0.9994	0.0006
2.46	0.9931	0.0069	2.86	0.9979	0.0021	3.26	0.9994	0.0006
2.47	0.9932	0.0068	2.87	0.9979	0.0021	3.27	0.9995	0.0005
2.48	0.9934	0.0066	2.88	0.9980	0.0020	3.28	0.9995	0.0005
2.49	0.9936	0.0064	2.89	0.9981	0.0019	3.29	0.9995	0.0005
2.50	0.9938	0.0062	2.90	0.9981	0.0019	3.30	0.9995	0.0005
2.51	0.9940	0.0060	2.91	0.9982	0.0018	3.31	0.9995	0.0005
2.52	0.9941	0.0059	2.92	0.9982	0.0018	3.32	0.9995	0.0005
2.53	0.9943	0.0057	2.93	0.9983	0.0017	3.33	0.9996	0.0004
2.54	0.9945	0.0055	2.94	0.9984	0.0016	3.34	0.9996	0.0004
2.55	0.9946	0.0054	2.95	0.9984	0.0016	3.35	0.9996	0.0004
2.56	0.9948	0.0052	2.96	0.9985	0.0015	3.36	0.9996	0.0004
2.57	**0.9949**	**0.0051**	2.97	0.9985	0.0015	3.37	0.9996	0.0004
2.58	0.9951	0.0049	2.98	0.9986	0.0014	3.38	0.9996	0.0004
2.59	0.9952	0.0048	2.99	0.9986	0.0014	3.39	0.9997	0.0003
2.60	0.9953	0.0047	3.00	0.9987	0.0013	3.40	0.9997	0.0003
2.61	0.9955	0.0045	3.01	0.9987	0.0013	3.41	0.9997	0.0003
2.62	0.9956	0.0044	3.02	0.9987	0.0013	3.42	0.9997	0.0003
2.63	0.9957	0.0043	3.03	0.9988	0.0012	3.43	0.9997	0.0003
2.64	0.9959	0.0041	3.04	0.9988	0.0012	3.44	0.9997	0.0003
2.65	0.9960	0.0040	3.05	0.9989	0.0011	3.45	0.9997	0.0003
2.66	0.9961	0.0039	3.06	0.9989	0.0011	3.46	0.9997	0.0003
2.67	0.9962	0.0038	3.07	0.9989	0.0011	3.47	0.9997	0.0003
2.68	0.9963	0.0037	3.08	0.9990	0.0010	3.48	0.9997	0.0003
2.69	0.9964	0.0036	3.09	0.9900	0.0010	3.49	0.9998	0.0002
2.70	0.9965	0.0035	3.10	0.9900	0.0010	3.50	0.9998	0.0002
2.71	0.9966	0.0034	3.11	0.9991	0.0009			
2.72	0.9967	0.0033	3.12	0.9991	0.0009	3.60	0.9998	0.0002
2.73	0.9968	0.0032	3.13	0.9991	0.0009			
2.74	0.9969	0.0031	3.14	0.9992	0.0008	3.70	0.9999	0.0001
2.75	0.9970	0.0030	3.15	0.9992	0.0008			
2.76	0.9971	0.0029	3.16	0.9992	0.0008	3.80	0.9999	0.0001
2.77	0.9972	0.0028	3.17	0.9992	0.0008	3.90	1.000	0.0000
2.78	0.9973	0.0027	3.18	0.9993	0.0007			
2.79	0.9974	0.0026	3.19	0.9993	0.0007			

Source: Copyright © Dr. Victor Bissonnette.

Table B–2 Critical Values of the *t* Distribution

	Two-Tailed Testing/(One-Tailed Testing)					
df	*0.2* *0.1*	*0.1* *0.05*	*0.05* *0.025*	*0.02* *0.01*	*0.01* *0.005*	*0.001* *0.0005*
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	0.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.768
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
50	1.299	1.676	2.009	2.403	2.678	3.496
60	1.296	1.671	2.000	2.390	2.660	3.460
80	1.292	1.664	1.990	2.374	2.639	3.416
100	1.290	1.660	1.984	2.364	2.626	3.390
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.327	2.576	3.291

Source:

Table B–3 Critical Values of the *F Distribution,* $\alpha = .05$

	df between										
df within	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*12*	*24*	∞
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.68	4.53	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.79	2.61	2.41
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.69	2.51	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.20	2.01	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.18	1.98	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.16	1.96	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.12	1.91	1.66
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	1.92	1.70	1.39
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	1.88	1.65	1.33
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.85	1.63	1.28
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.83	1.61	1.26
∞	3.84	3.00	2.61	2.37	2.22	2.10	2.01	1.94	1.75	1.52	1.00

Source:

Table B–4 Critical Values of the Studentized Range Statistic

		No. of Groups								
df within	*a*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
5	0.05	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99
	0.01	5.70	6.98	7.80	8.42	8.91	9.32	9.67	9.97	10.24
6	0.05	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49
	0.01	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10
7	0.05	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16
	0.01	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37
8	0.05	3.26	4.04	4.53	4.89	5.17	4.50	5.60	5.77	5.92
	0.01	4.75	5.64	6.20	6.62	6.96	7.24	7.47	7.68	7.86
9	0.05	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74
	0.01	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.33	7.49
10	0.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60
	0.01	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21
11	0.05	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49
	0.01	4.39	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99
12	0.05	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39
	0.01	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.67	6.81
13	0.05	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32
	0.01	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67
14	0.05	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25
	0.01	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54
15	0.05	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20
	0.01	4.17	4.84	5.25	5.56	5.80	5.99	6.16	6.31	6.44
16	0.05	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15
	0.01	4.13	4.79	5.19	5.49	5.72	5.92	6.08	6.22	6.35
17	0.05	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11
	0.01	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27
18	0.05	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07
	0.01	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20
19	0.05	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04
	0.01	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14
20	0.05	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01
	0.01	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09
24	0.05	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92
	0.01	3.96	4.55	4.91	5.17	5.37	5.54	5.69	5.81	5.92
30	0.05	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82
	0.01	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76
40	0.05	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73
	0.01	3.82	4.37	4.70	4.93	5.11	5.26	5.39	5.50	5.60
60	0.05	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65
	0.01	3.76	4.28	4.59	4.82	4.99	5.13	5.25	5.36	5.45
120	0.05	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56
	0.01	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30
∞	0.05	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47
	0.01	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16

Source: Data from E.S. Pearson and H.O. Hartley, *Biometricka Tables for Statisticians,* © 1966, Cambridge University Press; and Harter, Tables of Range and Studentized Range, *Annals of Mathematical Studies,* Vol. 31, pp. 1122–1147.

Table B–5 Critical Values of *r*

	Two-Tailed Testing /(One-Tailed Testing)					
df	*0.2 (0.1)*	*0.1 (0.05)*	*0.05 (0.025)*	*0.02 (0.01)*	*0.01 0.005*	*0.001 0.0005*
3	0.687	0.805	0.878	0.934	0.959	0.991
4	0.608	0.729	0.811	0.882	0.917	0.974
5	0.551	0.669	0.754	0.833	0.875	0.951
6	0.507	0.621	0.707	0.789	0.834	0.925
7	0.472	0.582	0.666	0.750	0.798	0.898
8	0.443	0.549	0.632	0.715	0.765	0.872
9	0.419	0.521	0.602	0.685	0.735	0.847
10	0.398	0.497	0.576	0.658	0.708	0.823
11	0.380	0.476	0.553	0.634	0.684	0.801
12	0.365	0.458	0.532	0.612	0.661	0.780
13	0.351	0.441	0.514	0.592	0.641	0.760
14	0.338	0.426	0.497	0.574	0.623	0.742
15	0.327	0.412	0.482	0.558	0.606	0.725
16	0.317	0.400	0.468	0.543	0.590	0.708
17	0.308	0.389	0.456	0.529	0.575	0.693
18	0.299	0.378	0.444	0.516	0.561	0.679
19	0.291	0.369	0.433	0.503	0.549	0.665
20	0.284	0.360	0.423	0.492	0.537	0.652
21	0.277	0.352	0.413	0.482	0.526	0.640
22	0.271	0.344	0.404	0.472	0.515	0.629
23	0.265	0.337	0.396	0.462	0.505	0.618
24	0.260	0.330	0.388	0.453	0.496	0.607
25	0.255	0.323	0.381	0.445	0.487	0.597
26	0.250	0.317	0.374	0.437	0.479	0.588
27	0.245	0.311	0.367	0.430	0.471	0.579
28	0.241	0.306	0.361	0.423	0.463	0.570
29	0.237	0.301	0.355	0.416	0.456	0.562
30	0.233	0.296	0.349	0.409	0.449	0.554
40	0.202	0.257	0.304	0.358	0.393	0.490
50	0.181	0.231	0.273	0.322	0.354	0.443
60	0.165	0.211	0.250	0.295	0.325	0.408
80	0.143	0.183	0.217	0.257	0.283	0.357
100	0.128	1.164	0.195	0.230	0.254	0.321
120	0.117	0.150	0.178	0.210	0.232	0.294
140	0.108	0.139	0.165	0.195	0.216	0.273
160	0.101	0.130	0.154	0.183	0.202	0.256
180	0.095	0.122	0.146	0.172	0.190	0.242
200	0.091	0.116	0.138	0.164	0.181	0.230
300	0.074	0.095	0.113	0.134	0.148	0.188
400	0.064	0.082	0.098	0.116	0.128	0.164
500	0.057	0.073	0.088	0.104	0.115	0.146

Source: Copyright © Dr. Victor Bissonnette.

Table B–6 Critical Values of χ^2 Distribution

	Two-Tailed Testing /(One-Tailed Testing)				
df	*0.10* *0.05*	*0.05* *0.025*	*0.02* *0.01*	*0.01* *0.005*	*0.001* *0.0005*
1	2.706	3.841	5.412	6.635	10.827
2	4.605	5.991	7.824	9.210	13.815
3	6.251	7.185	9.837	11.345	16.268
4	7.779	9.488	11.668	13.277	18.465
5	9.236	11.070	13.388	15.086	20.517
6	10.645	12.592	15.033	16.812	22.457
7	12.017	14.067	16.622	18.475	24.322
8	13.362	15.507	18.168	20.090	26.125
9	14.684	16.919	19.679	21.666	27.877
10	15.987	18.307	21.161	23.209	29.588
11	17.275	19.675	22.618	24.725	31.264
12	18.549	21.026	24.054	26.217	32.909
13	19.812	22.362	25.742	27.688	34.528
14	21.064	23.685	26.873	29.141	36.123
15	22.307	24.996	28.259	30.578	37.697
16	23.542	26.296	29.633	32.000	39.252
17	24.769	27.587	30.955	33.409	40.790
18	25.989	28.869	32.346	34.805	42.312
19	27.204	30.144	33.687	36.191	43.820
20	28.412	31.410	35.020	37.566	45.315
21	29.615	32.671	36.343	38.932	46.797
22	30.813	33.924	37.659	40.289	48.268
23	32.007	35.172	38.968	41.638	49.728
24	33.196	36.415	40.270	42.980	51.178
25	34.382	37.652	41.566	44.314	52.620
26	35.563	38.885	42.856	45.642	54.052
27	36.741	40.133	44.140	46.963	55.476
28	37.916	41.337	45.419	48.278	56.893
29	39.087	42.557	46.693	49.588	58.302
30	40.256	43.733	47.962	50.892	59.703

Source:

Table B–7 Binomial Distribution with $p = 0.50$

N	*Left S*	*p*	*Right S*
1	0	0.5000	1
2	0	0.2500	2
	1	0.7500	1
3	0	0.1250	3
	1	0.5000	2
4	0	0.0625	4
	1	0.3125	3
	2	0.6875	2
5	0	0.0312	5
	1	0.1875	4
	2	0.5000	3
6	0	0.0156	6
	1	0.1094	5
	2	0.3438	4
	3	0.6562	3
7	0	0.0078	7
	1	0.0625	6
	2	0.2266	5
	3	0.5000	4
8	0	0.0039	8
	1	0.0352	7
	2	0.1445	6
	3	0.3633	5
	4	0.6367	4
9	0	0.0020	9
	1	0.0195	8
	2	0.0898	7
	3	0.2539	6
	4	0.5000	5
10	0	0.0010	10
	1	0.0107	9
	2	0.0547	8
	3	0.1719	7
	4	0.3770	6
	5	0.6230	5
11	0	0.0005	11
	1	0.0059	10
	2	0.0327	9
	3	0.1133	8
	4	0.2744	7
	5	0.5000	6
12	0	0.0002	12
	1	0.0032	11
	2	0.0193	10
	3	0.0730	9
	4	0.1938	8
	5	0.3872	7
	6	0.6128	6
13	0	0.0001	13
	1	0.0017	12
	2	0.0112	11
	3	0.0461	10
	4	0.1334	9
	5	0.2905	8
	6	0.5000	7
14	0	0.0000	14
	1	0.0009	13
	2	0.0065	12
	3	0.0287	11
	4	0.0898	10
	5	0.2120	9
	6	0.3953	8
	7	0.6047	7
15	0	0.0000	15
	1	0.0005	14
	2	0.0037	13
	3	0.0176	12
	4	0.0592	11
	5	0.1509	10
	6	0.3036	9
	7	0.5000	8
16	0	0.0000	16
	1	0.0003	15
	2	0.0021	14
	3	0.0106	13
	4	0.0384	12
	5	0.1051	11
	6	0.2272	10
	7	0.4018	9
	8	0.5982	8
17	0	0.0000	17
	1	0.0001	16
	2	0.0012	15
	3	0.0064	14
	4	0.0245	13
	5	0.0717	12
	6	0.1662	11
	7	0.3145	10
	8	0.5000	9

continues

Table B–7 continued

N	*Left S*	*p*	*Right S*
18	0	0.0000	18
	1	0.0001	17
	2	0.0007	16
	3	0.0038	15
	4	0.0154	14
	5	0.0481	13
	6	0.1189	12
	7	0.2403	11
	8	0.4073	10
	9	0.5927	9
19	0	0.0000	19
	1	0.0000	18
	2	0.0004	17
	3	0.0022	16
	4	0.0096	15
	5	0.0318	14
	6	0.0835	13
	7	0.1796	12
	8	0.3238	11
	9	0.5000	10
20	0	0.0000	20
	1	0.0000	19
	2	0.0002	18
	3	0.0013	17
	4	0.0059	16
	5	0.0207	15
	6	0.0577	14
	7	0.1316	13
	8	0.2517	12
	9	0.4119	11
	10	0.5881	10

Note: Entries labled *P* in the table are the cumulative probability from each extreme to the value of *S*, for a given *n* when $p = 0.5$. Left tail probabilities are given for $S \leq 0.5n$, and right tail for $S \geq 0.5n$.

Source: Reprinted from National Bureau of Standards.

Table B–8 Critical Values for ΣR_X for the Mann-Whitney Wilcoxon Rank Sum Test

	$n_1 = 3$			
n_2	*0.005*	*0.01*	*0.025*	*0.05*
3				6–15
4				6–18
5			6–21	7–20
6			6–23	8–22
7		6–27	7–26	8–25
8		6–30	8–28	9–27
9	6–33	7–32	8–31	10–29
10	6–36	7–35	9–33	10–32
11	6–39	7–38	9–36	11–34
12	7–41	8–40	10–38	11–37
13	7–44	8–43	10–41	12–39
14	7–47	8–46	11–43	13–41
15	8–49	9–48	11–46	13–44

	$n_1 = 4$			
n_2	*0.005*	*0.01*	*0.025*	*0.05*
4			10–26	11–25
5		10–30	11–29	12–28
6	10–34	11–33	12–32	13–31
7	10–38	11–37	13–35	14–34
8	11–41	12–40	14–38	15–37
9	11–45	13–43	14–42	16–40
10	12–48	13–47	15–45	17–43
11	12–52	14–50	16–48	18–46
12	13–55	15–53	17–51	19–49
13	13–59	15–57	18–54	20–52
14	14–62	16–60	19–57	21–55
15	15–65	17–63	20–60	22–58

	$n_1 = 5$			
n_2	*0.005*	*0.01*	*0.025*	*0.05*
5	15–40	16–39	17–38	19–36
6	16–44	17–43	18–42	20–40
7	16–49	18–47	20–45	21–44
8	17–53	19–51	21–49	23–47
9	18–57	20–55	22–53	24–51
10	19–61	21–59	23–57	26–54
11	20–65	22–63	24–61	27–58
12	21–69	23–67	24–64	28–62
13	22–73	24–71	27–68	30–65
14	22–78	25–75	28–72	31–69
15	23–82	56–79	29–76	33–72

	$n_1 = 6$			
n_2	*0.005*	*0.01*	*0.025*	*0.05*
6	23–55	24–54	26–52	28–50
7	24–60	25–59	27–57	29–55
8	25–65	27–63	29–61	31–59
9	26–70	28–68	31–65	33–63
10	27–75	29–73	32–70	35–67
11	28–80	30–78	34–74	37–71
12	30–84	32–82	35–79	38–76
13	31–89	33–87	37–83	40–80
14	32–94	34–92	38–88	42–84
15	33–99	36–96	40–92	44–88

	$n_1 = 7$			
n_2	*0.005*	*0.01*	*0.025*	*0.05*
7	32–73	34–71	36–69	39–66
8	34–78	35–77	38–74	41–71
9	35–84	37–82	40–79	43–76
10	37–89	39–87	42–84	45–81
11	38–95	40–93	44–89	47–86
12	40–100	42–98	46–94	49–91
13	41–106	44–103	48–99	52–95
14	43–111	45–109	50–104	54–100
15	44–117	47–114	52–109	56–105

	$n_1 = 8$			
n_2	*0.005*	*0.01*	*0.025*	*0.05*
8	43–93	45–91	49–87	51–85
9	45–99	47–97	51–93	54–90
10	47–105	49–103	53–99	56–96
11	49–111	51–109	55–105	59–101
12	51–117	53–115	58–110	62–106
13	53–123	56–120	60–116	64–112
14	54–130	58–126	62–122	67–117
15	56–136	60–132	65–127	69–123

continues

Table B–8 continued

	$n_1 = 9$			
n_2	*0.005*	*0.01*	*0.025*	*0.05*
9	56–115	59–112	62–109	66–105
10	58–122	61–119	65–115	69–111
11	61–128	63–126	68–121	72–117
12	63–135	66–132	71–127	75–123
13	65–142	68–139	73–134	78–129
14	67–149	71–145	76–140	81–135
15	69–156	73–152	79–146	84–141

	$n_1 = 10$			
n_2	*0.005*	*0.01*	*0.025*	*0.05*
10	71–139	74–136	78–132	82–128
11	73–147	77–143	81–139	86–134
12	76–154	79–151	84–146	89–141
13	79–161	82–158	88–152	92–148
14	81–169	85–165	91–159	96–154
15	84–176	88–172	94–166	99–161

	$n_1 = 11$			
n_2	*0.005*	*0.01*	*0.025*	*0.05*
11	87–166	91–162	96–157	100–153
12	90–174	94–170	99–165	104–160
13	93–182	97–178	103–172	108–167
14	96–190	100–186	106–180	112–174
15	99–198	103–194	110–187	116–181

	$n_1 = 12$			
n_2	*0.005*	*0.01*	*0.025*	*0.05*
12	105–195	109–191	115–185	120–180
13	109–203	113–199	119–193	125–187
14	112–212	116–208	123–201	129–195
15	115–221	120–216	127–209	133–203

	$n_1 = 13$			
n_2	*0.005*	*0.01*	*0.025*	*0.05*
13	125–166	130–221	136–215	142–209
14	129–235	134–230	141–223	147–217
15	133–244	138–239	145–232	152–225

	$n_1 = 14$			
n_2	*0.005*	*0.01*	*0.025*	*0.05*
14	147–259	152–254	160–246	166–240
15	151–269	156–264	164–256	171–249

	$n_1 = 15$			
n_2	*0.005*	*0.0871*	*0.025*	*0.05*
15	171–294	176–289	184–281	192–273

Source: E.W. Minium, *Statistical Reasoning in Psychology and Education,* pp. 549–550. © Reprinted by permission of John Wiley & Sons, Inc.

Table B–9 Kruskal-Wallis Distribution

	Right-Tail Probability for Q				
n_1, n_2, n_3	*0.100*	*0.050*	*0.020*	*0.0101*	*0.001*
2, 2, 2	4.571				
3, 2, 1	4.286				
3, 2, 2	4.500	4.714			
3, 3, 1	4.571	5.143			
3, 3, 2	4.556	5.361	6.250		
3, 3, 3	4.622	5.600	6.489	7.200	
4, 2, 1	4.500				
4, 2, 2	4.458	5.333	6.000		
4, 3, 1	4.056	5.208			
4, 3, 2	4.511	5.444	6.144	6.444	
4, 3, 3	4.709	5.791	6.564	6.745	
4, 4, 1	4.167	4.967	6.667	6.667	
4, 4, 2	4.555	5.455	6.600	7.036	
4, 4, 3	4.545	5.598	6.712	7.144	8.909
4, 4, 4	4.654	5.692	6.962	7.654	9.269
5, 2, 1	4.200	5.000			
5, 2, 2	4.373	5.160	6.000	6.533	
5, 3, 1	4.018	4.960	6.044		
5, 3, 2	4.651	5.251	6.124	6.909	
5, 3, 3	4.533	5.648	6.533	7.079	8.727
5, 4, 1	3.987	4.985	6.431	6.955	
5, 4, 2	4.541	5.273	6.505	7.205	8.591
5, 4, 3	4.549	5.656	6.676	7.445	8.795
5, 4, 4	4.668	5.657	6.953	7.760	9.168
5, 5, 1	4.109	5.127	6.145	7.309	
5, 5, 2	4.623	5.338	6.446	7.338	8.938
5, 5, 3	4.545	5.705	6.866	7.578	9.284
5, 5, 4	4.523	5.666	7.000	7.823	9.606
5, 5, 5	4.560	5.780	7.220	8.000	9.920

Note: Each table entry is the smallest value of the Kruskal-Wallis Q such that its right-tail probability is less than or equal to the value given on the top row for $k = 3$, each sample size less than or equal to 5. For $k > 3$, right-tail probabilities for Q are found from Table B–6 with $k - 1$ degrees of freedom.

Source: Adapted with permission from R.L. Iman, D. Quade, & D.A. Alexander, Exact Probability Levels for the Kruskal-Wallis Test, in *Selected Tables in Mathematical Statistics,* Vol. 3, Institute of Mathematical Statistics, ed., pp. 329–384, © 1975, American Mathematical Society.

APPENDIX C

Frequency Distribution of Discharges by DRG, Critical Care Hospital, 2004: An Index to the Number of Cases by DRG

DRG	*Frequency*	*DRG*	*Frequency*	*DRG*	*Frequency*
001	41	046	3	094	5
002	23	047	4	096	7
004	2	049	9	097	4
006	1	050	3	099	3
007	8	051	2	100	2
009	3	053	9	101	8
010	19	055	2	102	3
011	5	057	2	103	2
012	10	063	31	104	21
013	10	064	11	105	33
014	38	065	3	106	11
015	12	066	1	107	28
016	1	067	1	108	17
017	1	068	3	109	44
018	7	069	4	110	33
019	8	073	2	111	7
020	7	075	76	112	5
021	6	076	35	113	7
023	2	077	3	115	8
024	28	078	22	116	45
025	24	079	26	117	3
026	1	080	1	118	1
027	5	082	47	120	6
028	6	083	2	121	14
029	5	084	2	122	10
032	1	085	22	123	9
034	10	086	3	124	84
035	4	087	14	125	41
036	1	088	27	126	3
037	3	089	77	127	61
040	1	090	4	128	2
042	3	091	1	130	36
044	4	092	8	131	6
045	3	093	2	132	17

continues

DRG	Frequency	DRG	Frequency	DRG	Frequency
133	14	185	11	253	7
134	12	188	34	254	1
135	1	189	10	256	5
136	2	191	23	257	15
137	1	192	2	258	12
138	42	193	4	260	1
139	18	197	6	261	1
140	3	198	1	262	1
141	8	199	6	263	2
142	3	200	3	264	4
143	26	201	3	265	4
144	50	202	45	266	2
145	10	203	28	268	4
146	5	204	41	269	9
147	4	205	22	270	2
148	59	206	1	271	3
149	14	207	11	272	3
150	13	208	3	274	4
151	6	209	17	275	1
152	2	210	11	276	4
153	5	211	8	277	27
154	27	212	1	278	15
155	11	213	6	280	6
157	9	216	13	281	5
158	3	217	8	282	1
159	13	218	11	283	6
160	19	219	28	284	3
161	5	220	1	286	5
162	1	223	2	287	2
164	1	224	3	288	75
165	7	225	3	289	11
166	3	226	6	290	17
167	21	227	8	292	2
168	7	228	2	293	1
169	3	229	3	294	14
170	10	231	4	295	12
171	4	233	17	296	22
172	36	234	5	297	8
173	9	236	1	299	2
174	41	238	4	300	6
175	7	239	15	301	8
176	11	240	16	302	36
177	5	241	3	303	27
178	1	242	2	304	25
179	42	243	23	305	31
180	24	244	4	308	3
181	17	247	6	310	5
182	43	248	5	311	2
183	21	249	1	313	1

DRG	Frequency	DRG	Frequency	DRG	Frequency
315	9	388	69	455	1
316	77	389	11	460	2
318	5	390	61	461	4
319	1	391	443	462	168
320	26	392	3	463	1
321	3	394	3	466	1
322	1	395	59	467	20
323	6	397	7	468	28
324	1	398	15	472	1
325	2	399	4	473	25
326	1	353	42	475	61
331	46	400	1	477	12
332	5	401	4	478	39
334	4	402	3	479	25
335	11	403	44	480	5
336	1	404	6	481	13
337	1	357	6	482	52
338	1	406	5	483	51
341	3	407	4	485	6
344	1	408	11	486	9
346	1	409	1	487	7
352	2	410	148	489	18
354	6	413	8	490	7
355	4	415	29	492	28
356	4	416	31	493	14
357	6	418	43	494	11
358	21	419	4	495	2
359	63	421	5	496	1
360	3	423	14	498	9
365	4	424	1	499	3
366	5	425	7	500	12
368	7	426	23	503	2
369	5	427	4	504	2
370	116	428	1	506	2
371	96	429	13	507	1
372	86	430	417	509	2
373	337	434	2	510	8
374	27	435	5	511	7
376	11	439	2	512	5
377	2	440	7	515	24
378	4	442	13	516	21
379	28	443	7	517	26
380	3	444	1	518	22
381	7	445	4	519	2
383	83	449	17	520	4
384	22	450	4	521	10
385	26	451	2	523	10
386	20	452	18	524	10
387	1	453	2	525	1

continues

DRG	*Frequency*	*DRG*	*Frequency*	*DRG*	*Frequency*
526	33	532	4	537	7
527	71	533	11	538	5
528	4	534	21	539	5
529	1	535	8	540	4
531	4	536	10	Total	6824

APPENDIX D

Answers and Solutions

CHAPTER 1

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
See glossary (Appendix A).

3. Outline the procedure for age-adjusting crude mortality rates by the direct standardization method.
To adjust the crude mortality rate by age, first calculate the age-specific death rate (ASDR) for each age group in the two populations that are being compared. Second, combine the two populations by age groups. Third, multiply the ASDR for each population age group times the combined population total for each group. This results in an expected number of deaths for each age group for each population. Finally, sum the ASDRs for each age group in each population; this results in the age-adjusted mortality rate for each group.

5. Describe the differences between neonatal mortality rate, post-neonatal mortality rate, and infant mortality rate.
The *neonatal mortality rate* is the number of deaths of newborns under 28 days of age for a given time period compared to the number of live births for the same time period.

The *post-neonatal mortality rate* is the number of deaths of infants age 28 days up to and not including one year of age for a given time period compared to the number of live births for the same time period.

The *infant mortality rate* is the number of deaths of infants under one year of age for a given time period compared to the number of live births for the same time period. The infant mortality rate combines the neonatal and post-neonatal mortality rates.

MULTIPLE CHOICE

1. b

$$[(20 + 55 + 155)/(15{,}000 + 17{,}000 + 6{,}000) \times 1{,}000 = 6.1 \text{ per } 1{,}000$$

PROBLEMS

1. Review the hypothetical data on deaths in the MICU in Table 1–A–1 and answer the questions that follow.

a. What is the ratio of male deaths to female deaths?

There were a total of 44 MICU deaths: 24 men and 20 women; the ratio is 24:20 or 6:5. The interpretation is that for every six male deaths there were five female deaths.

b. What proportion of the patients who died were admitted from the Emergency Department (ED)? What proportion were transfers from other hospitals?

Seven of the patients who died were admitted from the ED; the proportion of patient who were admitted from the ED is 7/44 or .16.

Eleven of the patients who died were transfers from other hospitals; the proportion of patient who were transfers from other hospitals is 11/44 or .25.

c. The total number of patients discharged from DRG 475 was 61. What is the case fatality rate for DRG 475?

Fifteen patients who died in the MICU fell in DRG 475. The case fatality rate is:

$$(15/61) \times 100 = 24.6\%$$

d. The total number of patients discharged from DRG 483 was 51. What is the case fatality rate for DRG 483?

Five patients who died in the MICU fell in DRG 483. The case fatality rate is:

$$(5/51) \times 100 = 9.8\%$$

e. What is the relative risk of death for patients discharged from DRG 475 compared to discharges from DRG 483?

$$\frac{.245}{.098} = 2.5$$

The risk of death is 2.5 times greater for DRG 475 than DRG 483.

3. Review the data in Table 1–A–3 and answer the questions that follow.

Table 1–A–3 Ohio AIDS Cases by Age, Race, and Sex, as of June 30, 2003; U.S. AIDS Cases 1981–1999

Demographics	*Total Ohio*[1]	*Total U.S.*[2]
Age		
<13	96	8,718
13–19	72	3,725
20–24	331	25,904
25–29	776	97,676
30–39	4,686	329,066
40–49	5,362	190,087
50–64	2,254	68,196
65+	217	10,002
Subtotal	13,794	733,374
Race/Ethnicity		
White	6,943	318,354
Black	5,742	272,881
Hispanic	642	133,703
Other	74	7,479
Unknown	393	957
Subtotal	13,794	733,374
Sex		
Male	10,766	609,329
Female	2,634	124,045
Unknown	394	
Subtotal	13,794	733,374

[1]*Source:* Ohio HIV/AIDS Statistical Summary, HIV Infection and AIDS Cases Diagnosed through June 2003, Ohio Department of Health, www.odh.state.oh.us.

[2]US DHHS, Public Health Service, CDC, National Center for HIV, STD, and TB Prevention, AIDS Public Information Data Set, CDC WONDER on-line database, wonder.cdc.gov.

a. What is the male-to-female ratio for AIDS in Ohio? In the United States?
The male to female ratio in Ohio is 10,766 to 2,634 or 4.1 to 1.

The male to female ratio in the United States is 609,329 to 124,045 or 4.9:1.

b. Out of the total number of AIDS cases in Ohio, what proportion are women? Of the total cases in the United States, what proportion are women?
Proportion of cases in Ohio that are **known** to be women is
$2{,}634/(2{,}634 + 10{,}766 + 394) = .19$

Proportion of cases in the US that are women is $124{,}045/(124{,}045 + 609{,}329) = .20$

c. What proportion of the total AIDS cases in Ohio are ages 30 to 39? What proportion in the United States are ages 30 to 39?

The proportion of cases in Ohio that fall into the age group 30 to 39 is 4,686/13,794 = .34.

The proportion of cases in the US that fall into the age group 30 to 39 is 329,066/733,374 = .45.

d. Calculate the proportion of AIDS cases in Ohio by race. Calculate the proportion of AIDS cases in the United States by race.

Race/Ethnicity	*Ohio*	*p*	*US*	*p**
White	6,943	.50	318,354	.43
Black	5,742	.42	272,881	.37
Hispanic	642	.04	133,703	.18
Other	74	.01	7,479	.01
Unknown	393	.03	957	.001

*Does not total to 1.0 due to rounding

e. How do the above ratios and proportions, Ohio versus the United States, compare?

The ratio of males to females is slightly higher in the United States than in Ohio. The proportion of cases that fall in the 30–39 age group is much greater in the United States than in Ohio, 45% versus 34%. The proportion of whites and blacks affected is greater in Ohio than in the United States as a whole. This is probably because the overall racial proportions in the state of Ohio differ from the total US population proportions.

5. Calculate the odds ratio for the data in Table 1–13. Interpret the results.

Table 1–13 Relative Risk of Death Due to Malignancies, Women versus Men Aged 65+, State of Michigan, 2001

	Death Due to Malignancies		
Sex	*Yes*	*No*	*Total*
Men	7,153	21,507	28,660
	(a)	(b)	(a + b)
Women	6,565	28,890	35,455
	(c)	(d)	(c + d)

Risk of illness among men:
$a/(a + b) = 7{,}153/(7{,}15 + 21{,}507) = 0.2496$
Risk of illness among women
$c/(c + d) = 6{,}565/(6{,}565 + 28{,}890) = 0.1852$
Risk ratio, men to women: .2496/1852 = 1.34
Thus, the risk of death due to malignancy among men aged 65+ is 1.3 times greater than the risk of death due to malignancy in women in the same age group.

Source: Data from Centers for Disease Control and Prevention, CDC WONDER database, wonder.cdc.gov.

$$\text{Odds ratio} = \frac{a \times d}{b \times c}$$

$$= \frac{7{,}153 \times 28{,}890}{21{,}507 \times 6{,}565} = \frac{206{,}650{,}170}{141{,}193{,}455} = 1.46$$

The odds of dying from a malignancy is almost 1.5 times greater in men than women in the state of Michigan.

7. The overall mortality rate for patients who have had a cerebrovascular accident (CVA) is 15.8% at City General Hospital. You have been asked to compare the hospital's mortality rate to that of the state. Using the data provided in Table 1–A–6, calculate the age-adjusted death rate and the standard mortality ratio (SMR) for the hospital, using the indirect method of standardization. Explain the results.

Table 1–A–6 Mortality Rates for CVAs, State versus City General Hospital

Severity of Illness	*State Mortality Rate*	*Hospital Discharges for CVA*	*Observed Deaths*	*Expected Deaths*
1	4.2	55	2	2.31
2	5.9	116	8	6.84
3	7.8	195	20	15.21
4	20.9	147	29	30.72
5	34.6	62	32	21.45
		575	91	76.53

Observed mortality rate is $(91 \times 100)/575 = 15.8\%$
Expected mortality rate is $(76.53 \times 100)/575 = 13.3\%$.

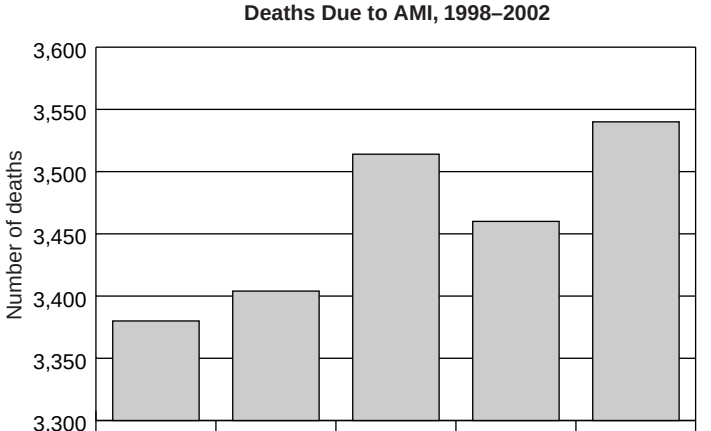
$$\text{SMR} = \frac{\text{Observed Complication Rate}}{\text{Expected Complication Rate}} = \frac{.158}{.133} = 1.19$$

After indirect standardization, the mortality rate for CGH is 19% higher than the expected mortality rate.

QUESTIONS (all data gathered from the state of Utah)

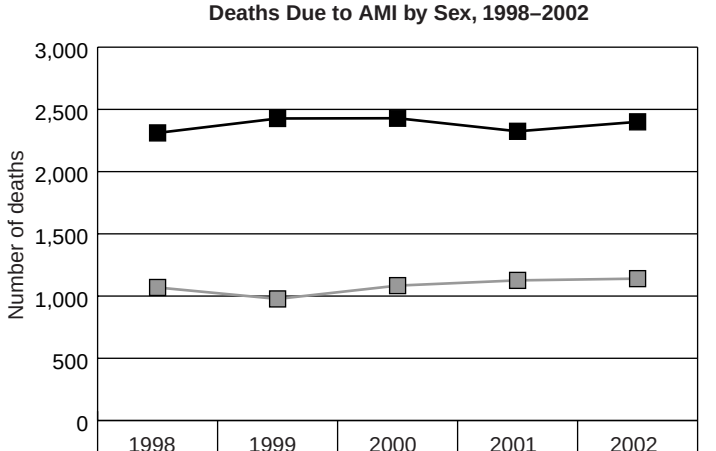
1. For the diagnosis of acute myocardial infarction, ICD-9-CM category 410:

a. Prepare a bar graph that displays the number of deaths due to AMI by year, 1998 through 2002.



	1998	1999	2000	2001	2002
Deaths	3,380	3,404	3,514	3,450	3,540

b. Prepare a line graph that displays the number of deaths by gender for the years 1998 through 2002. What are your conclusions?



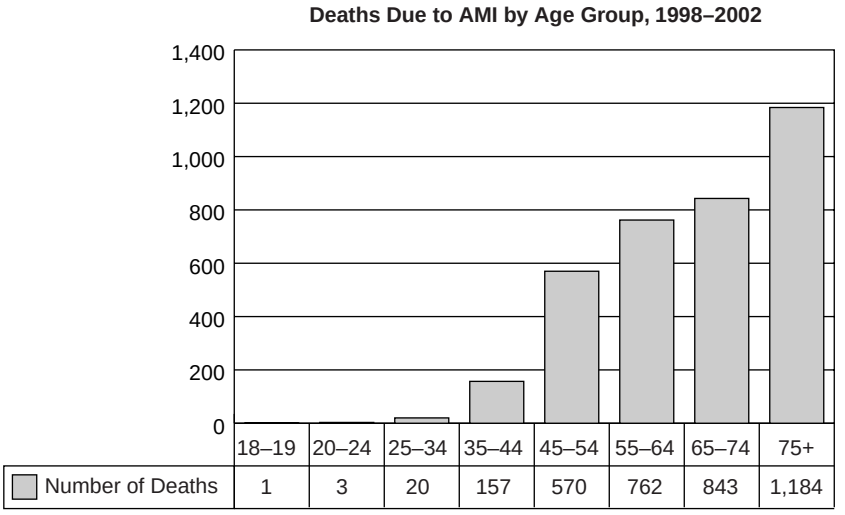
	1998	1999	2000	2001	2002
Men	2,311	2,427	2,429	2,324	2,400
Women	1,069	977	1,084	1,126	1,140

In raw numbers, there are more men who die as a result of an AMI than women. This is true for each year 1998–2002.

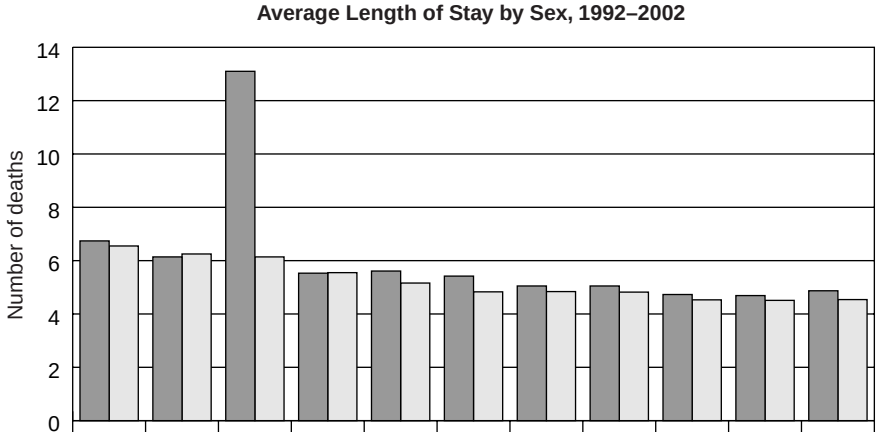
c. Prepare a table that displays the number of deaths due to AMI by age group in the state of Utah. Use the table to prepare a bar graph of the same information.

Deaths Due to AMI by Age Group 2002

Age Group	*Deaths*
18–19	1
20–24	3
25–34	20
35–44	157
45–54	570
55–64	762
65–74	843
75+	1,184
	3,540



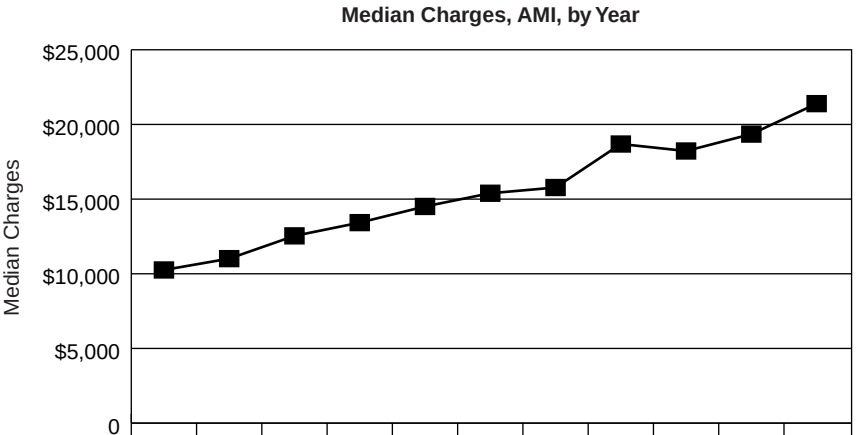
d. Prepare a bar graph that displays the average length of stay by gender for the years 1992 through 2002. What are your conclusions after reviewing the data?



	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Women	5.74	5.14	13.1	5.53	5.61	5.42	5.05	5.05	4.73	4.69	4.87
Men	6.56	6.25	6.14	5.55	5.16	4.83	4.84	4.82	4.53	4.51	4.54

In general, the average length of stay for women slightly exceeds that for men. There appears to be an aberration in 1994 which requires further investigation.

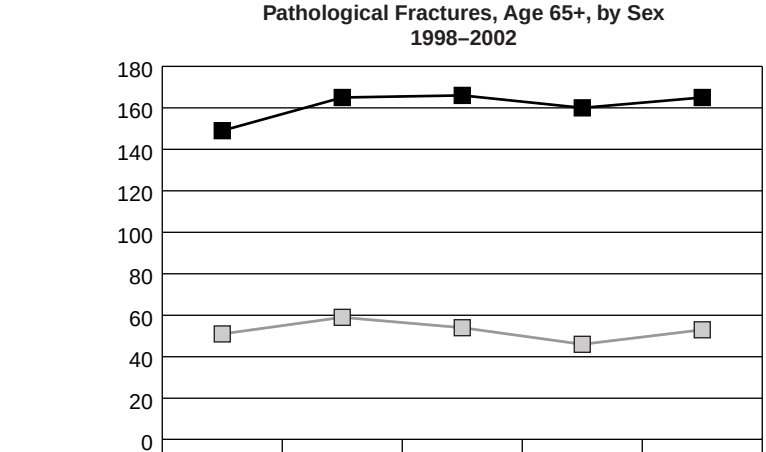
e. Prepare a line graph that displays the median charges by year, 1992 through 2002. What does the graph indicate?



	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02
Median Charges $	10,255	11,014	12,538	13,425	14,508	15,393	15,776	18,689	18,222	19,354	21,393

The median charges have increased each year in the time frame 1992–2002. The sharpest increase occurred in the time period 1998–1999.

3. Determine the number of patient discharges with pathological fractures, ICD-9-CM code733.1, by year, 1998 through 2002, and by gender. You are interested in patients aged 65 years and over. Prepare a line graph displaying the number of discharges by year and by gender. Discuss your findings.

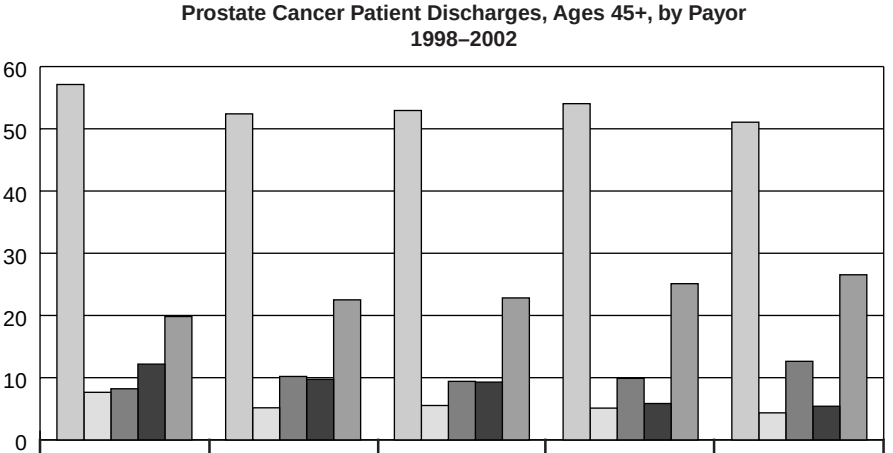


	1998	1999	2000	2001	2002
Female	149	165	166	160	165
Male	51	59	54	46	53

For the time frame, more women are affected with pathological fractures than men. This may be because women are more likely to experience this condition following menopause.

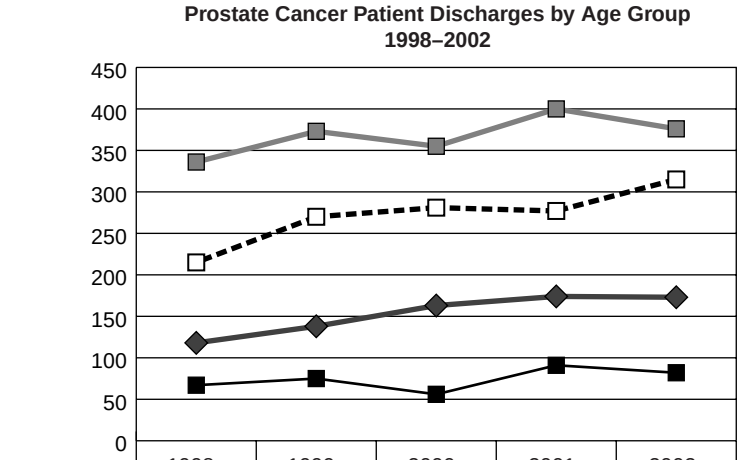
5. For ICD-9-CM code 185, for the years 1998 through 2002:

a. Prepare a bar graph or pie chart, by third-party payer, of men, aged 45 and older, discharged with a diagnosis of prostate cancer.



	1998	1999	2000	2001	2002
Medicare	52.12%	52.40%	52.94%	54.04%	51.06%
Other Govt	7.65%	5.16%	5.53%	5.11%	4.35%
Blue Cross	8.22%	10.20%	9.41%	9.89%	12.63%
Other Comm	12.18%	9.73%	9.29%	5.85%	5.41%
Mngd Care	19.83%	22.51%	22.82%	25.11%	26.54%

b. Prepare a bar graph that displays the number of men, by age group, discharged with prostate cancer.



	1998	1999	2000	2001	2002
45–54	67	75	56	91	82
55–64	215	270	281	277	315
65–74	336	373	355	400	378
75+	118	138	163	174	173

c. Discuss your findings.
It appears that the number of patients diagnosed with prostate cancer is increasing in each age category. The greatest number of cases is in the age 65–74 age group.

CHAPTER 2

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
See glossary (Appendix A)

3. What questions should be answered in the title of a table, chart, or graph?
The title for a table, chart, or graph should be as complete as possible. It should answer the following questions:

- What are the data? (counts, percentages, etc.)
- To whom does the data relate? (males and females with a certain condition, diseases by race, etc.)

- Where are the data from? (hospital, community, state, county, etc.)
- When? To what time frame does the data apply? (day, week, month, year)

5. Describe the differences between a stacked bar chart and a 100% component bar chart.

In a *stacked bar chart,* the segments of the bar for each data category are stacked like building blocks on top of one another to form a single bar. The bar represents the total number of cases in the data category; the segments of the bar represent the frequency of certain types of cases within the category.

In a *100% component bar chart,* all bars in the display are of the same height—each representing 100% of the cases in the category. The bar is divided into segments that represent the percentage of certain types of cases within categories. For example, a 100% component chart may display types of cancer by site; the segments within each bar may represent the percentage of males and females affected by that particular type of cancer.

The stacked bar chart displays frequencies; the 100% component bar chart displays percentages.

MULTIPLE CHOICE

1. a. bar graph

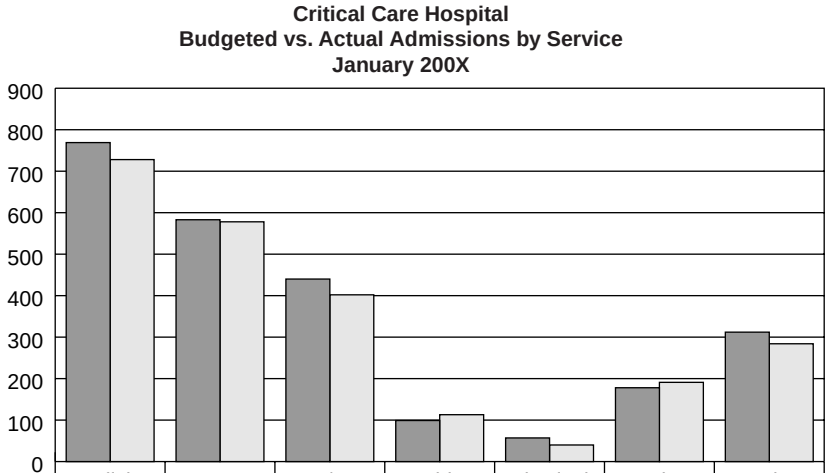
3. b. percentage of discharges by third-party payer

5. a. bar chart

PROBLEMS

Prepare the appropriate charts and graphs for the following problems. Include a title for each and identify the data source when indicated.

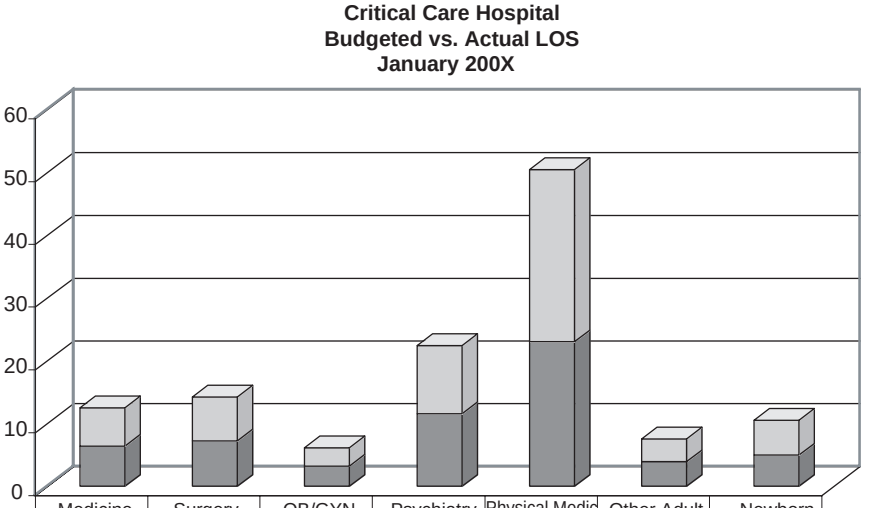
1. **The admissions data in Table 2–A–1 compare actual admissions by hospital service with the budgeted number of hospital admissions for the month of January for Critical Care Hospital. Using computer graphic software, construct a bar chart that compares budgeted admissions with actual admissions. Write a short summary of the results.**



	Medicine	Surgery	OB/Gyn	Psychiatry	Physical Medicine & Rehab	Other Adult	Newborn
Budgeted Admissions	769	583	440	99	57	178	312
Actual Admissions	728	578	402	113	48	191	294

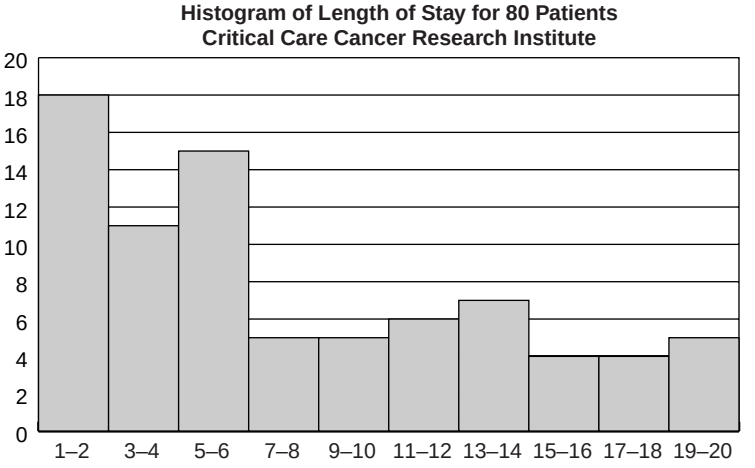
Actual admissions did not quite meet the budgeted projections except for Psychiatry Service and Other Adult. The differences do not appear to be significant.

3. Table 2–A–3 contains length-of-stay data by service for the month of January for Critical Care Hospital. Construct a stacked bar chart that compares actual average length of stay with the budgeted average length of stay.

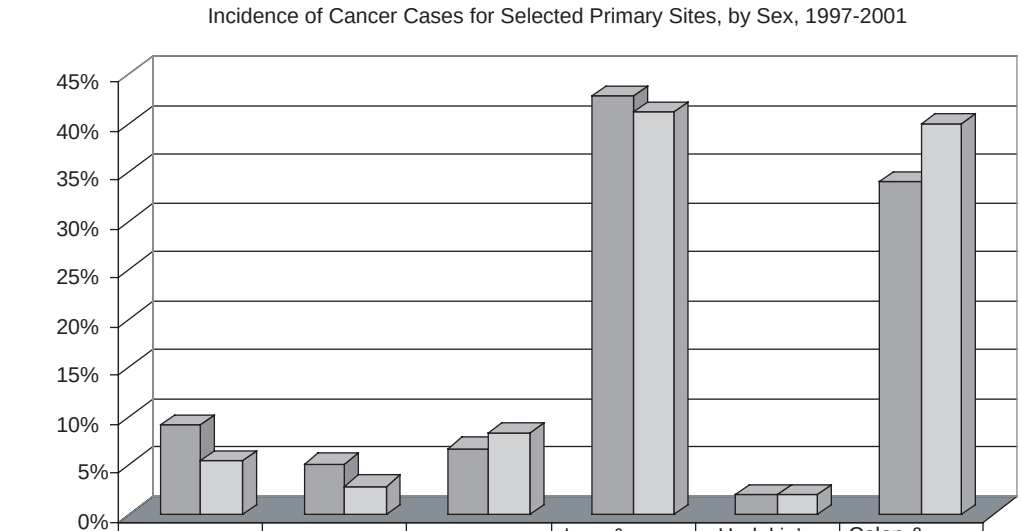


	Medicine	Surgery	OB/GYN	Psychiatry	Physical Medic and Rehab	Other Adult	Newborn
Actual	6.09	6.98	2.91	10.82	27.46	3.6	5.55
Budgeted	6.39	7.23	3.22	11.56	22.98	3.93	4.97

5. Exhibit 2–A–1 displays the lengths of stay for 80 patients at the Critical Care Cancer Research Institute. Construct a histogram of these data.



7. Review the data in Table 2–4. Determine the percentage of total male and female cancer cases for each site. Prepare a bar chart to display your results.



	Oral & Pharynx	Liver	Pancreas	Lung & Bronchus	Hodgkin's Lymphoma	Colon & Rectum
Male	9.17%	5.17%	6.78%	42.84%	2.05%	33.99%
Female	5.44%	2.95%	8.38%	41.23%	2.03%	39.97%

CHAPTER 3

KNOWLEDGE QUESTIONS

1. Define key terms listed at the beginning of this chapter.
See glossary (Appendix A).

3. Why are validity and reliability requirements of accuracy in the measurement process?
To have confidence in the data that we collect, we must assess the validity and reliability of the measures. *Validity* tells us that we are truly measuring the characteristic or property of interest, and *reliability* tells us that the measurement results are consistent over time or repeated measures.

5. You weighed your dog this morning on your bathroom scale. His weight was 15 lb. You decided to weigh him again in the evening, and his weight had increased to 20 lb. This is most likely what type of measurement error? Explain your answer.
This is a problem with aspects of *reliability.* In this case, it is a matter of *stability, or test-retest* reliability. If we obtain a different result each time a property is measured, we cannot be confident in our results.

7. Relate the importance of determining the sensitivity, specificity, and predictive value of a screening measure.

Sensitivity, specificity, and predictive value further assist us in evaluating the validity of a measure. *Sensitivity* is the extent to which the measure identifies the characteristic when it is truly present. If the instrument is not sensitive, it will not detect the characteristic of interest when it is present. *Specificity* is the extent to which a measure excludes the characteristic of interest when it is truly absent. An instrument that is non-specific will falsely detect the characteristic of interest when it is not present. A pilot test is conducted to assess the sensitivity and specificity of an instrument.

Predictive value is how well the instrument measures the characteristic of interest. For example, if there are 25 patients who suffer a diabetic coma, and the measure correctly identifies 15 of the cases, the predictive validity is 60%. That is, it correctly measures the characteristic of interest 60% of the time.

9. What is the difference between a discrete variable and a continuous variable?

Discrete variables have gaps between successive measures, or the gaps between measures are not equal. Discrete variables fall on the nominal and ordinal scales of measurement. We cannot multiply or divide variables that are discrete. In contrast, *continuous* variables fall on the interval and ratio scales of measurement. The term *continuous* implies that there are no gaps between successive numbers. Results of these measures may be fractional, such as a weight of 125.3 lbs. We can use multiplication and division with continuous variables.

MULTIPLE CHOICE

1. a. nominal

3. b. ordinal

5. d. interval

7. b. test-retest reliability

9. a. age

11. e. all of the above

PROBLEMS

1. The hospital readmission rate is often considered an indicator of an undesirable patient outcome. The quality improvement team is interested in reducing the number of readmissions among patients discharged with a principal diagnosis of congestive heart failure (CHF). The team believes that the high readmission rate is due to the difficulty that these patients have in controlling the number of drugs that they typically take. The team believes that by improving patient/family education regarding

drug administration, the readmission rate could be reduced. Thus, they have developed the screen "CHF patients taking three or more drugs" to identify these patients before discharge. To evaluate the effectiveness of the measure, the team conducts a study on all CHF patients discharged the previous year. The results appear in Table 3–A–1:

Table 3–A–1 Readmissions of CHF Patients

	CHF Patients		
No. of Drugs Administered	*Readmitted*	*Not Readmitted*	*Total*
≥ 3 drugs	200	40	240
< 3 drugs	100	900	1,000
Total	300	940	1,240

a. Calculate the sensitivity, specificity, and predictive value for this measure.

$$\begin{aligned} \text{Sensitivity} &= a/a + c \\ &= 200/(200 + 100) \\ &= 0.67 \end{aligned}$$

$$\begin{aligned} \text{Specificity} &= d/b + d \\ &= 900/(40 + 900) \\ &= 0.96 \end{aligned}$$

$$\begin{aligned} \text{Predictive value} &= a/a + b \\ &= 200/(200 + 40) \\ &= 0.83 \end{aligned}$$

b. On the basis of your results, is this an effective measure? Why or why not?
Even though the measure results in a number of false positives, it may be a useful measure for identifying this group of patients. There are few false negatives indicating good specificity (0.96) and fairly good predictive value (0.83).

3. At Werethebest Hospital, 34 Caesarean sections were performed during January and at Weresosick Hospital, 54 Caesarean sections were performed. During January, Werethebest Hospital had 200 deliveries; Weresosick Hospital had 1,100 deliveries. The national benchmark for the C-section rate is 15%.

The head of obstetrics at Werethebest Hospital claims that their OB service provides better care than that provided at the rival hospital. Do you agree with this assessment?

$$\text{C-section rate at Werethebest Hospital} = (34/200) \times 100 = 17.0\%$$
$$\text{C-section rate at Weresosick Hospital} = (54/1{,}100) \times 100 = 4.9\%$$

The C-section rate at Werethebest Hospital exceeds the national benchmark. The comparison should be made on the basis of the percentage of C-sections performed out of the total number of deliveries rather than the actual number of C-sections.

5. **As part of the quality improvement team, you have prepared a report on acute myocardial infarction (AMI) mortality, which is displayed in Table 3–A–2. You query DRG 121, Circulatory Disorders with AMI and Cardiovascular (CV) complications, Discharged Alive; DRG 122, Circulatory Disorders with AMI without CV Complications, Discharged Alive; and DRG 123, Circulatory Disorders with AMI, Expired. You want to include only those cases where AMI is the principal diagnosis.**

Table 3–A–2 Acute Myocardial Infarctions, DRGs 121, 122, and 123

Principal Diagnosis	*DRG 121*	*DRG 122*	*DRG 123*
410.01	3	1	1
410.11	9	3	4
410.31	2	1	0
410.41	11	11	1
410.61	1	0	0
410.71	66	31	9
410.91	25	11	5
421.0	1	0	0
428.0	31	0	0
Total	119	58	20

   a. **Assess the report below for validity and reliability. What corrections should be made to calculate the AMI mortality rate?**
   All diagnosis codes for the last two cases should be reviewed to determine if the correct principal diagnosis was assigned. If these two patients were admitted for conditions other than AMI, they should be excluded from the analysis.
   b. **What is the AMI mortality rate?**
   Assuming that the latter two cases are excluded from the analysis, the AMI mortality rate is calculated as:

$$\text{AMI mortality rate} = (20/(117 + 58 + 20) \times 100 = 10.3\%$$

c. Review Table 3–A–3, which displays the average length of stay for each DRG. What factors should also be considered when presenting the AMI mortality rate? What is the net AMI mortality rate?

Table 3–A–3 Average Length of Stay (ALOS) and Numbers of Patients with Length of Stay (LOS) Two Days or Less, DRGs 121, 122, and 123

	DRG 121	*DRG 122*	*DRG 123*
ALOS	5.6 days	3.6 days	3.5 days
LOS $\leq$ 2 days	26	18	10

The net AMI mortality rate may be a better reflection of the care provided by the hospital as it corrects for deaths for which the hospital may not have adequate time to treat the patient. The net mortality rate corrects for patients who died within 48 hours of admission. The net mortality rate is almost half the gross mortality rate.

$$\text{Net AMI Mortality Rate} = (20 - 10)/([117 + 58 + 20] - 10) \times 100 = 5.4\%$$

CHAPTER 4

KNOWLEDGE QUESTIONS

1. Define key terms listed at the beginning of this chapter.
See glossary (Appendix A).

3. Compare and contrast the following measures of central tendency: mean, median, and mode.
The *mean* is the arithmetic average of a frequency distribution. All observations in the frequency distribution are used to calculate the mean; the mean is considered a *nonresistant* statistic because it is influenced by the extreme values in the distribution. The mean is appropriate for interval and ratio level data.

The *median* is the "middlemost" value in a frequency distribution. Fifty percent of the observations in the distribution lie above the median and 50% lie below the median. The median is considered a *resistant* statistic because it is not influenced by extreme values in the distribution. The median is appropriate for ordinal, interval, and ratio level data.

The *mode* is the most frequently occurring value in a frequency distribution. The mode is considered a resistant statistic because it is not influenced by extreme values in the distribution. A limitation of the mode is that a frequency distribution may have more than one mode. A frequency distribution may be bimodal or multi-modal. The mode can fluctuate widely from sample to sample. For nominal level data, the mode is the most fre-

quently occurring category. The mode is the only measure of central tendency appropriate for nominal level data. The mode may also be reported for ordinal, interval, and ratio level data.

5. **Why do measures of central tendency and variation of ungrouped frequency distributions differ from those of grouped frequency distributions?**
   Measures of central tendency and variation are calculated from each observation in an ungrouped frequency distribution. Statistics calculated from an ungrouped frequency distribution are more precise than statistics calculated from a grouped frequency distribution.

   In a grouped frequency distribution, the frequency distribution is divided into class intervals. It is assumed that the observations are evenly distributed evenly throughout each class interval—even though this may not be the case. Statistics calculated from a grouped frequency distribution will vary depending upon how the distribution is grouped. Because the midpoints of the class intervals and not each observation in the distribution are used to calculate the statistics that describe the grouped frequency distribution, these statistics will be less precise.

MULTIPLE CHOICE

1. a. 63.3 $[(60 \times .67) + (70 \times .30)] = 63.3$; or $[(60 \times 40) = (70 \times 20)]/2 = 63.3$

3. d. 25 $(50 - 25 = 25)$

5. c. $6(10.5 - 4.5)$

7. b. 7.5

9. d. between 9.5 days and 14.5 days

11. b. 19.5 days

13. a. 14.5 days

15. a. mode

17. d. 25 $(5^2 = 25)$

19. b. mode—most frequently occurring observation

21. d. 103

PROBLEMS

1. Review the data in Tables 4–A–1 and 4–A–2 and answer the questions that follow. Use an electronic spreadsheet to assist you in preparing the answers.

Table 4–A–1 Male Deaths Due to Leukemia (ICD-9-CM Codes 200.0–200.9) in the state of Ohio, 1998

Age Group	*Leukemia Deaths in Men*	*p*	*Cum. p*	*M*	*f(M)*
5–14	16	0.012	0.012	9.5	152.00
15–24	20	0.015	0.026	19.5	390.00
25–34	27	0.020	0.046	29.5	796.50
35–44	63	0.046	0.092	39.5	2,488.50
45–54	118	0.086	0.178	49.5	5,841.00
55–64	194	0.142	0.320	59.5	11,543.00
65–74	388	0.284	0.604	69.5	26,966.00
75–84	418	0.306	0.909	79.5	33,231.00
85+	124	0.091	1.000	89.5	11,098.00
Total	1,368	1.000			92,506.00

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC Wonder on-line database, wonder.cdc.gov.

Mean	67.6
Crude Mode	79.5
Crude Median	69.5

Table 4–A–2 Female Deaths Due to Leukemia (ICD-9-CM Codes 200.0–200.9) in the state of Ohio, 1998

Age Group	*Leukemia Deaths in Women*	*p*	*Cum. p*	*M*	*f(M)*
5–14	6	0.005	0.006	9.5	57.00
15–24	10	0.008	0.013	19.5	195.00
25–34	9	0.008	0.021	29.5	265.50
35–44	26	0.022	0.043	39.5	1,027.00
45–54	61	0.051	0.094	49.5	3,019.50
55–64	132	0.110	0.204	59.5	7,854.00
65–74	296	0.247	0.451	69.5	20.572.00
75–84	463	0.387	0.838	79.5	36,808.50
85+	194	0.162	1.000	89.5	17,363.00
Total	1,197	1.000			87,161.50

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC Wonder on-line database, wonder.cdc.gov.

Mean	72.82
Crude Mode	79.5
Crude Median	79.5

a. What is the mean age of death for men? For women?
The mean age of death for men is 67.6 years (92,506/1,368); the mean age for women is 72.8 years (87,161.5/1,197).

b. What are the crude modes and median ages of death for men? For women?
The crude mode for men is 79.5 years; the crude mode for women is 79.5 years. (The midpoints of the most frequently occurring intervals.)

The crude median for men is 69.5 years; the crude median for women is 79.5 years.

c. What are the refined mode and the refined median for men? For women?
The refined mode for men:

$$\begin{aligned} Mode &= L + [w(f_{mo} - f_b)]/[f_{mo} - f_b) + (f_{mo} - f_a)] \\ &= 74.5 + [10(418 - 388)]/(418 - 388) + (418 - 124) \\ &= 74.5 + (300/324) \\ &= 74.5 + .9 \\ &= 75.4 \end{aligned}$$

The refined median for men:

$$\begin{aligned} Mdn &= L + w(1/2n - c)/f_{mdn} \\ &= 64.5 + [10(684 - 438)]/388 \\ &= 64.5 + [10(246)/388 \\ &= 54.5 + 6.3 \\ &= 70.8 \end{aligned}$$

The refined mode for women:

$$\begin{aligned} Mode &= L + [w(f_{mo} - f_b)]/[f_{mo} - f_b) + (f_{mo} - f_a)] \\ &= 74.5 + [10(463 - 296)]/(463 - 296) + (463 - 194) \\ &= 74.5 + (1670/436) \\ &= 74.5 + 3.8 \\ &= 78.3 \end{aligned}$$

The refined median for women:

$$\begin{aligned} Mdn &= L + w(1/2n - c)/f_{mdn} \\ &= 74.5 + [10(598.5 - 540)]/463 \\ &= 74.5 + [10(58.5)/463 \\ &= 74.5 + 1.3 \\ &= 75.8 \end{aligned}$$

d. Compare and contrast the crude and refined results for each group. Explain any disparities that may exist.

The crude median for men (69.5) is similar to the refined median (70.8). The crude mode for men (79.5), however, is somewhat higher than the refined mode (75.4). This is because there are more observations in the interval that falls below the interval that containing the mode than the interval that falls above the interval containing the mode, 388 and 124 respectively. The mode is pulled in the direction of the interval that contains more observations.

The crude median for women (79.5) is somewhat higher than the refined median (75.8). A greater proportion of deaths for women occur before age 74.5 than after age 84.5; thus the median is being pulled in the direction of the lesser age. This also better matches the mean age, which is 72.8. The crude mode for women (79.5) is not much different from the refined mode (78.3).

e. In analyzing the results of your data for men and women, what conclusions can you draw?

Since we do not have information regarding age at diagnosis and/or survival times, we cannot say that women live longer with the disease than men. We can only conclude that the average age of death due to leukemia is higher for women, 72.8, than the average age for men, 67.6. For both data sets, the median better represents the most typical age at death because the mean is influenced by extreme values.

3. The lengths of stay for a group of patients discharged DRG 127 are presented in Exhibit 4–A–2.

a. Use statistical software to calculate the mean, median, mode, variance, and standard deviation for the ungrouped frequency distribution and to prepare a frequency table.

Statistics

LOS

N	Valid	61
	Missing	0
Mean		6.31
Median		4.00
Mode		3
Std. Deviation		6.220
Variance		38.685

LOS

		Frequency	*Percent*	*Valid Percent*	*Cumulative Percent*
Valid	1	6	9.8	9.8	9.8
	2	6	9.8	9.8	19.7
	3	16	26.2	26.2	45.9
	4	5	8.2	8.2	54.1
	5	5	8.2	8.2	62.3
	6	4	6.6	6.6	68.9
	7	1	1.6	1.6	70.5
	8	4	6.6	6.6	77.0
	10	4	6.6	6.6	83.6
	11	3	4.9	4.9	88.5
	13	1	1.6	1.6	90.2
	14	1	1.6	1.6	91.8
	15	1	1.6	1.6	93.4
	16	1	1.6	1.6	95.1
	17	1	1.6	1.6	96.7
	27	1	1.6	1.6	98.4
	36	1	1.6	1.6	100.0
Total		61	100.0	100.0	

b. Group LOS into class intervals. Prepare a table that displays the frequencies for each class interval, the cumulative frequency, the relative proportion, and the cumulative percent.

To determine the number of class intervals needed:

$$
\begin{aligned}
K &= 1 + 3.3 \log (n) \\
&= 1 + 3.3 \log (61) \\
&= 1 + 3.3 (1.785) \\
&= 1 + 5.89 \\
&= 6.89 \text{ or } 7 \text{ class intervals (eight class intervals were used in order to cover the entire distribution)}
\end{aligned}
$$

To determine the width of the class intervals:

$$
\begin{aligned}
W &= (\max - \min)/k \\
&= (36 - 1)/7 \\
&= 5 \text{ (width of the class intervals is 5)}
\end{aligned}
$$

Class Interval	f	cf	rel f	Cum %
1–5	38	38	0.623	0.623
6–10	13	51	0.213	0.836
11–15	6	57	0.098	0.934
16–20	2	59	0.033	0.967
21–25	0	59	0.000	0.967
26–30	1	60	0.016	0.984
31–35	0	60	0.000	0.984
36–40	1	61	0.016	1.000
	61		1.000	

Class Interval	f	M	fM	$f(M^2)$
1–5	38	3	114	342
6–10	13	8	104	832
11–15	6	13	78	1,014
16–20	2	18	36	648
21–25	0	23	0	0
26–30	1	28	28	784
31–35	0	33	0	0
36–40	1	38	38	1,444
	61		398	5,064

c. Compute the mean, median, mode, variance, and standard deviation for the grouped data.

$$\overline{X} = 398/61$$
$$= 6.5$$
$$\text{Crude mode} = 3$$
$$\text{Crude median} = 3$$

$$\text{Re } finedMode = L + [w(f_{mo} - f_b)]/[f_{mo} - f_b) + (f_{mo} - f_a)]$$
$$= 0.5 + [5(38 - 0)]/(38 - 0) + (38 - 13)$$
$$= 0.5 + (190/63)$$
$$= 0.5 + 3.0$$
$$= 3.5$$

$$\text{Re } finedMdn = L + w(1/2n - c)/f_{mdn}$$
$$= 0.5 + [5(30.5 - 0)/38$$
$$= 0.5 + [5(30.5)/38$$
$$= 0.5 + 4.0$$
$$= 4.5$$

Variance and standard deviation:

$$\begin{aligned}\sum x^2 &= \sum f(M^2) - \sum (fM)^2/n \\ &= 5{,}064 - (398)^2/61 \\ &= 5{,}064 - 2{,}596.79 \\ &= 2{,}467.21\end{aligned}$$

$$\begin{aligned}s^2 &= \sum x^2/n - 1 \\ &= 2{,}467.21/61 - 1 \\ &= 41.1\end{aligned}$$

$$\begin{aligned}s &= \sqrt{\sum x^2/n - 1} \\ &= \sqrt{41.1} \\ &= 6.4\end{aligned}$$

d. Compare the results of the grouped and ungrouped frequency distributions.

	Ungrouped	*Grouped (Raw)*	*Grouped (Refined)*
Mean	6.31	6.5	6.5
Median	4	3	4.5
Mode	3	3	3.5
Variance	38.7		41.1
S.D.	6.2		6.4

CHAPTER 5

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
See glossary (Appendix A)

3. What is the difference between the standard deviation and the standard normal deviate?
The *standard deviation* is a measure of variation that is used to describe the number of units that an observation in a normal distribution is from the mean. The *standard normal deviate* is actually a *z* value in the standard normal distribution but has the same interpretation as the standard deviation. A standard normal deviate is the number of standard deviation units that an observed value lies away from the population mean, μ.

5. You have been analyzing hospital discharges from DRG 15, Transient Ischemic Attack and Precerebral Occlusions. The average length of stay (ALOS) for patients discharged from DRG 15 is 2.2 days. The national length of stay for DRG 15 is 4.1

days. You are interested in determining whether the hospital's length of stay for DRG 15 is significantly different from the national ALOS.

a. State the null and alternative hypotheses.

$$H_0: \mu_1 = \mu_2$$
$$H_A: \mu_1 \neq \mu_2$$

b. Set the alpha level.

$$\alpha = .05$$

7. Explain the differences between the alpha level and the *p* value.

The alpha level or level of significance is set in advance of conducting the statistical test. The alpha level is the level at which we will reject the null hypothesis. The alpha level also states the probability of making a type I error. The *p* value is obtained as a result of conducting the statistical test. The *p* value tells us how rare the result of the statistical test actually is. If the *p* value is less than or equal to the alpha level we reject the null hypothesis.

9. The mean length of stay for patients discharged from DRG 005, Extracranial Vascular Procedures, is 3.33. The standard deviation for the group is 3.18, and the number of patients discharged is 21. Calculate the 95% confidence interval for the mean length of stay.

$$\text{Standard error of the mean (S.E.)} = s/\sqrt{n}$$
$$= 3.18/\sqrt{21}$$
$$= 3.18/4.58$$
$$= 0.69$$

$$CI_{95} = \overline{X} \pm 1.96(s_{\overline{x}})$$
$$= 3.33 \pm 1.96(0.69)$$
$$= 3.33 \pm 1.35$$
$$= 1.98, 4.68$$

MULTIPLE CHOICE

1. a. $\pm 1\ \sigma$ of the mean

3. d. all of the above

5. b. 16%

7. c. positively skewed

9. b. 10 and 30

11. d. a and b

13. b. rejected 5% of the time when it is true

15. d. accept the null hypothesis when it is false

17. d. all of the above

19. c. 55 men and 45 women

21. b. two-stage random sampling

PROBLEMS

1. Review the data on length of stay that appear in Table 5-A-2 to answer the questions below.

Table 5–A–2 Critical Care Hospital, Length of Stay of Patients by Sex, DRG 127, Heart Failure and Shock

	LOS Days	*Female*	*Male*	*Total*
	1	1	5	6
	2	1	5	6
	3	3	13	16
	4	4	1	5
	5	0	5	5
	6	3	1	4
	7	0	1	1
	8	2	2	4
	10	1	3	4
	11	1	2	3
	13	1	0	1
	14	0	1	1
	15	1	0	1
	16	0	1	1
	17	0	1	1
	27	1	0	1
	36	1	0	1
Total		20	41	61

a. Using a microcomputer statistical package, calculate the average length of stay for the entire group, for men, and for women.

LOS

Gender	*Mean*	*N*	*Std. Deviation*
Female	8.70	20	8.761
Male	5.15	41	4.163
Total	6.31	61	6.220

b. You are interested in determining if there is a difference in the average length of stay by sex. State the null and alternative hypotheses; state the a priori alpha level.

$$H_0: \mu_1 = \mu_2$$
$$H_A: \mu_1 \neq \mu_2$$
$$\alpha = .05$$

c. Calculate the standard error of the mean length of stay for the entire group, for men, and for women.

Referring to the SPSS output below, the mean length of stay is 6.3 for the entire group, 5.2 for men, and 8.7 for women.

d. Calculate the 95% confidence interval for length of stay for the entire group, for men, and for women.

Referring to the SPSS output below, the 95% confidence intervals for each group are:

Group	*95% Confidence Interval*
Entire Group	4.72 to 7.90
Men	3.83 to 6.46
Women	4.60 to 12.8

LOS	*Female*	*Male*	*Total*
N	20	41	61
Mean	8.70	5.15	6.31
Standard Deviation	8.761	4.163	6.220
Standard Error	1.959	.650	.796
Lower bound 95% CI	4.60	3.83	4.72
Upper bound 95% CI	12.80	6.46	7.90
Minimum	1	1	1
Maximum	36	17	36

CHAPTER 6

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
See glossary (Appendix A).

3. Describe situations in which we would use one-tailed tests; describe situations in which we would use two-tailed tests.
We use a one-tailed test when we are interested in determining whether one population mean is significantly greater or less than another population mean. We use a two-tailed test when we are interested in determining whether two population means are significantly different from each other regardless of direction.

5. In hypothesis testing, what is meant by the term *effect?*
An *effect* is a change in one variable that may be due to another variable. The variable that displays the effect is referred to as the *dependent variable.* The variable considered to be responsible for the variability in the dependent variable is referred to as the *independent variable.*

MULTIPLE CHOICE

1. b. one-tailed test

3. b. non-directional

5. d. reject the H_0 at $\alpha = .05$ and reject the H_0 at $\alpha = .01$

7. b. variance

9. a. 99

PROBLEMS

1. As an HIM DRG analyst, you are interested in comparing the mean length of stay (LOS) for Critical Care Hospital and the national mean for DRG 002, Craniotomy, Age Greater Than 17 without CC. The hospital mean LOS is 4.17, and the standard deviation is 2.57. The national average LOS for DRG 005 is 5.2 days; the hypothetical standard deviation is 2.56. The summary data for Critical Care Hospital appear in Table 6–A–1 and Exhibit 6–A–1.

a. State the null and alternative hypotheses and the a priori alpha level.

$$H_0: \mu_1 = \mu_2$$
$$H_A: \mu_1 \neq \mu_2$$
$$\alpha = .05$$

b. Calculate the difference between the hospital mean and the national mean using the one sample *t* test. What is the number of degrees of freedom? What is the resultant *t* statistic? Is it statistically significant?

The SPSS output for the one sample *t* test:

One-Sample Statistics

	N	*Mean*	*Std. Deviation*	*Std. Error Mean*
LOS	23	4.17	2.570	.536

One-Sample Test

Test Value = 5.2

					95% Confidence Interval of the Difference	
	t	*df*	*Sig. (2-tailed)*	*Mean Difference*	*Lower*	*Upper*
LOS	−1.915	22	.069	−1.026	−2.14	.09

c. What are your conclusions?

The calculated value of t is -1.915, $df = 22$, $p = 0.069$. The calculated value of t is not statistically significant; we therefore conclude that it appears that the ALOS for DRG 002 for Critical Care Hospital is not significantly different from the national ALOS for DRG 002.

3. You have been monitoring the lengths of stay for two of your physicians who discharge the most patients from DRG 410, Chemotherapy without Acute Leukemia as a Secondary Diagnosis. The relevant statistics appear in Tables 6–A–4 and 6–A–5. You specifically want to know if the observed difference in the lengths of stay for physicians 1460 and 8210 is statistically significant.

Table 6–A–4 Frequency Distribution of Length of Stay by Physician, DRG 410, in 2004 at Critical Care Hospital (SPSS Output)

LOS * Physician Crosstabulation

		Physician		
		1460	*8210*	*Total*
LOS	2	13	21	34
	3	1	1	2
	4	1	0	1
	5	1	1	2
	13	1	0	1
	14	0	1	1
	15	1	0	1
	23	1	0	1
Total		19	24	43

Table 6–A–5 Mean and Standard Deviation for Length of Stay by Physician, DRG 410, in 2004 at Critical Care Hospital (SPSS Output)

Report

Physician	*Mean*	*N*	*Std. Deviation*
1460	4.68	19	5.812
8210	2.67	24	2.496
Total	3.56	43	4.350

a. State the null and alternative hypotheses and the a priori alpha level.

$$H_0: \mu_1 = \mu_2$$
$$H_A: \mu_1 \neq \mu_2$$
$$\alpha = .05$$

b. Use the *t* test for two independent sample means to determine if the observed difference between the two means is statistically significant. What are the number of degrees of freedom, the resultant *t* statistic, and the significance level?

Group Statistics

	LOS	
	Physician	
	1460	*8210*
N	19	24
Mean	4.68	2.67
Std. Deviation	5.812	2.496
Std. Error Mean	1.333	.510

Independent Samples Test

			LOS	
			Equal variances assumed	*Equal variances not assumed*
Levene's Test for Equality of Variances	F		7.743	
	Sig.		.008	
t-test for Equality of Means	*t*		1.535	1.413
	df		41	23.253
	Sig. (2-tailed)		.133	.171
	Mean Difference		2.018	2.018
	Std. Error Difference		1.315	1.427
	95% Confidence Interval of the Difference	Lower	−.637	−.934
		Upper	4.673	4.969

Since the hypothesis is one of inequality, we conduct a two-tailed *t*-test. The calculated *t* is 1.535, the degrees of freedom is 41 (43 − 2), $p = .008$. The calculated value of *t* is statistically significant.

c. What are your conclusions?

We reject the null hypothesis. It appears that the observed difference in the average lengths of stay is statistically significant.

CHAPTER 7

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.

See glossary (Appendix A).

3. To conduct the ANOVA procedure, the dependent variable must fall upon which scale of measurement? The grouping variable or independent variable falls upon which scale of measurement?
For the ANOVA procedure, the dependent variable must be continuous; it must fall on either the ratio or the interval scale of measurement. The independent or grouping variable is discrete; it must fall on the nominal scale of measurement.

5. What is the purpose of conducting post hoc procedures?
When we are using ANOVA to compare three or more group means, a significant *F* ratio does not tell us which of the group means are significantly different from each other. We conduct post hoc procedures to determine which group means differ. All or some of the group means may be significantly different from each other.

MULTIPLE CHOICE

1. c. one-way ANOVA

3. a. 34

5. d. all of the above

7. d. a and c

9. d. a and b

PROBLEMS

1. You have been analyzing hospital discharges from DRG 14, Intracranial Hemorrhage and Stroke with Infarction. You want to know if there is a difference in the average age of men and women discharged from DRG 14. The frequency distribution for discharges by sex appears in Table 7–A–1. You have decided to use the ANOVA procedure to calculate your results.

Table 7–A–1 Frequency Distribution of Age at Discharge by Gender, DRG 14, in 2004 at Critical Care Hospital (SPSS Output)

Age * Gender Crosstabulation

		Gender		
		Female	*Male*	*Total*
Age	22	0	1	1
	23	0	1	1
	26	0	1	1
	39	1	0	1
	44	0	1	1
	46	1	0	1
	47	0	1	1
	49	1	0	1
	50	1	0	1
	52	0	1	1
	55	0	1	1
	56	0	1	1
	57	1	0	1
	57	0	1	1
	57	0	1	1
	58	1	0	1
	58	1	0	1
	60	0	1	1
	60	0	1	1
	64	0	1	1
	68	1	0	1
	68	1	0	1
	70	0	1	1
	70	0	1	1
	71	0	1	1
	71	0	1	1
	72	1	0	1
	72	0	1	1
	73	0	1	1
	75	0	1	1
	76	1	0	1
	77	1	0	1
	77	1	0	1
	78	0	1	1
	83	0	1	1
	84	0	1	1
	86	1	0	1
	86	1	0	1
Total		15	23	38

a. State the null and alternative hypotheses and the alpha level that you will use.

H_0: $\mu_1 = \mu_2$
H_A: $\mu_1 \neq \mu_2$
$\alpha = .05$

b. What is the mean age for men? What is the mean age for women?

Descriptives

Age

Gender	*Mean*	*N*	*Std. Deviation*
Female	64.47	15	14.721
Male	59.57	23	17.758
Total	61.50	38	16.595

The mean age for women is 64.5, and the mean age for men is 59.6.

c. What is the calculated value of *F*? Is it statistically significant?

ANOVA Table

		Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Age * Gender	Between Groups (Combined)	218.114	1	218.114	.787	.381
	Within Groups	9971.386	36	276.983		
	Total	10189.500	37			

The calculated value of F is .787, df = 1, 36, p = .381. The calculated value of F is not statistically significant.

d. What is your conclusion?

We fail to reject the null hypothesis. It appears that the observed difference between the mean age for men and women is not statistically significant.

3. Physicians 2170 and 8060 have the most patients discharged from DRG 14. You want to know if there is a difference in the average age and length of stay of patients of these two physicians. The frequency distribution for age and length of stay for these two physicians appears in Tables 7–A–3 and 7–A–4. You have decided to use the ANOVA procedure to calculate your results.

Table 7–A–3 Frequency Distribution of Age at Discharge, DRG 14, Physicians 2170 and 8060, in 2004 at Critical Care Hospital (SPSS Output)

Age * Physician Crosstabulation

		Physician		
		2170	*8060*	*Total*
Age	26	0	1	1
	44	1	0	1
	46	0	1	1
	50	1	0	1
	52	1	0	1
	56	1	0	1
	57	0	1	1
	57	0	1	1
	57	0	1	1
	68	1	0	1
	68	0	1	1
	70	1	0	1
	75	1	0	1
	76	1	0	1
	77	0	1	1
	84	1	0	1
	86	0	1	1
Total		9	8	17

Table 7–A–4 Frequency Distribution of Length of Stay, Physicians 2170 and 8060, DRG 14, in 2004 at Critical Care Hospital (SPSS Output)

LOS * Physician Crosstabulation

		Physician		
		2170	*8060*	*Total*
LOS	2	3	1	4
	3	4	3	7
	5	1	1	2
	6	0	1	1
	7	1	0	1
	12	0	2	2
Total		9	8	17

a. State the null and alternative hypotheses and the alpha level that you will use.
The null and alternative hypotheses for testing the differences between the means for both age and length of stay is:

$$H_0: \mu_1 = \mu_2$$
$$H_A: \mu_1 \neq \mu_2$$
$$\alpha = .05$$

b. What is the average age for patients of physician 2170? What is the average age of patients for physician 8060?

Report

Age

Physician	*Mean*	*N*	*Std. Deviation*
2170	63.89	9	13.788
8060	59.25	8	18.530
Total	61.71	17	15.842

The average age of patients of physician 2170 is 63.9; the average age of patients of physician 8060 is 59.3.

c. What is the average length of stay for patients of physician 2170? What is the average length of patients for physician 8060?

LOS

Physician	*Mean*	*N*	*Std. Deviation*
2170	3.33	9	1.658
8060	5.75	8	4.062
Total	4.47	17	3.184

The average length of stay for patients of physician 2170 is 3.3; the average length of patients for patients of physician 8060 is 5.75.

d. What is the calculated value of *F* for each variable? Are they statistically significant?

ANOVA Table

		Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Age * Physician	Between Groups (Combined)	91.141	1	91.141	.348	.564
	Within Groups	3924.389	15	261.626		
	Total	4015.529	16			

For age, the calculated F is .348, df = 1, 15, p= .564. The calculated F for age is not statistically significant.

ANOVA Table

		Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
LOS * Physician	Between Groups (Combined)	24.735	1	24.735	2.698	.121
	Within Groups	137.500	15	9.167		
	Total	162.235	16			

For ALOS, the calculated F is 2.698, df = 1, 15, p = .121. The calculated F for ALOS is not statistically significant.

e. What is your conclusion? What factors may be influencing your results?
We fail to reject the null hypotheses. It appears that the observed difference between the mean age and ALOS for physicians 2170 and 8060 are not statistically significant. The sample size may not be large enough to detect a difference that is statistically significant.

CHAPTER 8

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
See Glossary (Appendix A).

3. What is the range for the Pearson *r*? How is the Pearson *r* statistic interpreted? Explain the concepts of positive linear relationship and negative linear relationship.
The range of the Pearson *r* is −1.0 to +1.0. If the Pearson *r* correlation coefficient is either +1.0 or −1.0, the relationship between the two variables is perfect; if the Pearson *r* is equal to zero, there is no relationship between the two variables. If the relationship between the two variables is positive, as one variable increases, so does the second. If the relationship is negative, as one variable increases, the second variable decreases.

5. What is the interpretation of the regression line in a scatter diagram?
The *regression line* is an indicator of the strength of the relationship between two linear variables. The steeper the slope of the regression line, the stronger the relationship between the two variables. If the slope of the regression line is in an upward direction from left to right, the relationship between the two variables is positive. If the slope of the regression line is downward from left to right, the relationship between the two variables is negative. If the regression line is flat, there is not relationship between the two variables.

MULTIPLE CHOICE

1. d. all of the above

3. c. there is a perfect positive relationship between x and y

5. c. 0.0

7. d. we have made an error in our calculations

PROBLEMS

1. You are studying DRG 105, Cardiac Valve Procedures and Other Major Cardiothoracic Procedures without Cardiac Catheterization, for Critical Care Hospital. Using the data provided in Table 8–A–1, calculate the Pearson *r* for each of the following pairs:

- **Age and length of stay**
- **Total charges and length of stay***
- **Age and total charges**

a. State the null and alternative hypotheses and alpha level for each.

Table 8–A–1 Case Summaries for DRG 105, Cardiac Valve Procedures and Other Major Cardiothoracic Procedures without Cardiac Catheterization

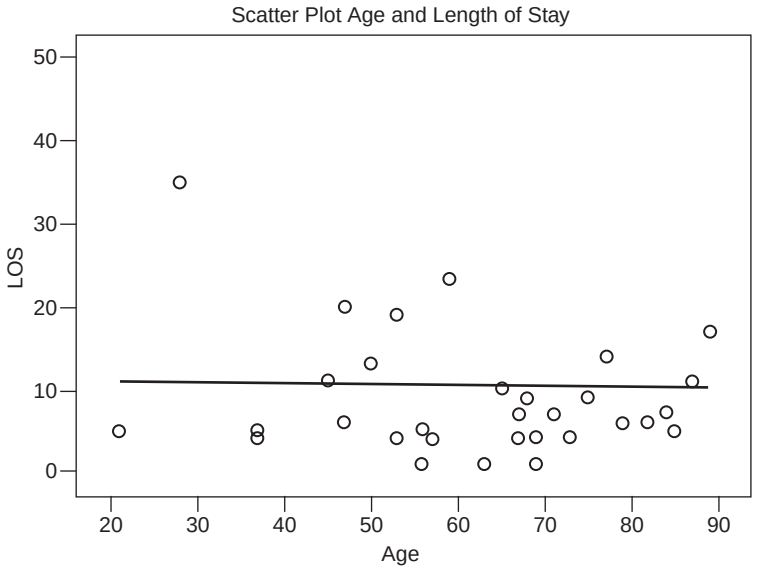
	Gender	*Age*	*LOS*	*Charges*	*Payor*
1	Female	47	20	$91,683	Medicaid
2	Female	75	43	$93,708	Medicare
3	Female	84	7	$21,446	Medicare
4	Female	50	13	$37,797	Medicare
5	Male	77	14	$54,364	Medicare
6	Male	57	4	$17,626	Medicare
7	Male	73	4	$12,832	Medicare
8	Female	56	1	$36,153	Medicaid
9	Male	69	1	$14,907	Medicaid
10	Female	81	23	$104,148	Medicare
11	Male	21	5	$21,423	Medicaid
12	Female	37	5	$24,971	Medicaid
13	Female	69	4	$17,022	Medicare
14	Female	89	17	$50,652	Medicare
15	Male	28	35	$186,496	Medicaid
16	Male	47	6	$24,441	Medicaid
17	Male	87	11	$35,349	Medicare
18	Female	85	5	$22,155	Medicare
19	Male	56	5	$24,455	Managed Care
20	Male	45	11	$36,401	Medicaid
21	Male	82	6	$25,783	Medicare
22	Female	65	10	$37,055	Managed Care
23	Male	67	4	$19,236	Medicare
24	Male	59	23	$60,132	Other
25	Female	67	7	$35,777	Medicare
26	Male	53	4	$19,972	Managed Care
27	Male	71	7	$25,409	Medicare
28	Female	79	6	$281,140	Medicare
29	Male	63	1	$41,283	Medicaid
30	Male	53	19	$71,439	Medicaid
31	Female	75	9	$33,735	Medicare
32	Female	68	9	$37,830	Gov Mngd Care
33	Male	37	4	$22,311	Medicaid
Total *N*	33	33	33	33	33

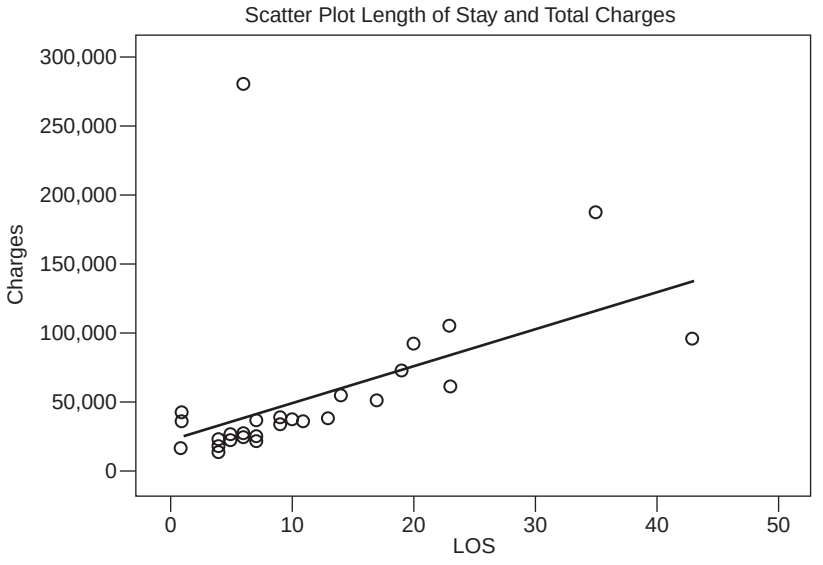
$H_0: \rho = 0$

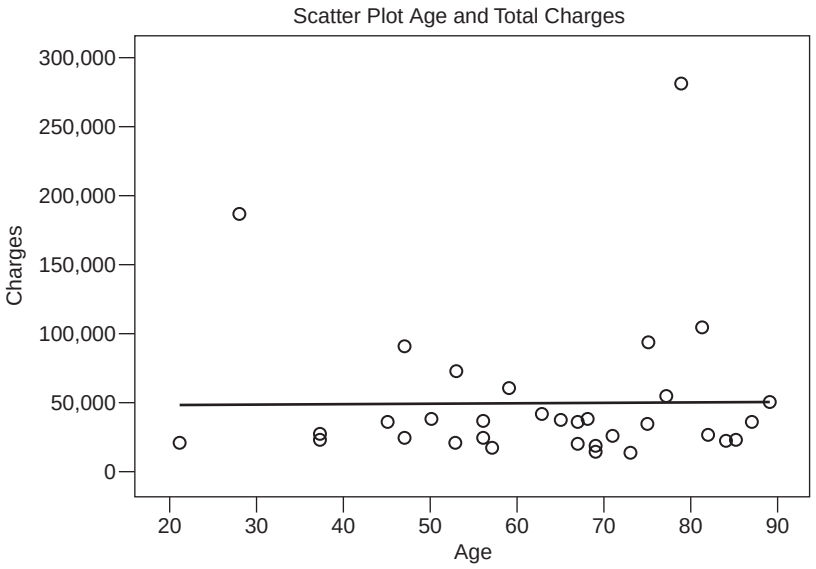
$H_A: \rho \neq 0$

$\alpha = .05$

b. Construct a scatter diagram with regression line for each.







For the variables age and length of stay, and age and total charges, the regression lines in the scatter plots are flat indication no relationship. For length of stay and total charges, the regression line originates in the lower left and moves to the upper right, indicating a positive linear relationship. The Pearson *r* is calculated for each; the results are:

Descriptive Statistics

	Mean	*Std. Deviation*	*N*
Age	62.79	17.315	33
LOS	10.39	9.585	33
Charges	49670.7715	54114.32276	33

Correlations

		Age	*LOS*	*Charges*
Age	Pearson Correlation	1	−.024	.009
	Sig. (2-tailed)		.896	.961
	N	33	33	33
LOS	Pearson Correlation	−.024	1	.480(**)
	Sig. (2-tailed)	.896		.005
	N	33	33	33
Charges	Pearson Correlation	.009	.480(**)	1
	Sig. (2-tailed)	.961	.005	
	N	33	33	33

** Correlation is significant at the 0.01 level (2-tailed).

c. State your conclusions for each.
For age and length of stay, $r = -.024$, and $p = .896$. We therefore fail to reject the null hypothesis and conclude that it appears that the relationship between age and length of stay is not statistically significant.

For age and total charges, $r = -.009$, and p = .961. We therefore fail to reject the null hypothesis and conclude that it appears that the relationship between age and total charges is not statistically significant.

For length of stay and total charges, $r = -.48$, and p = .005. We therefore reject the null hypothesis and conclude that it appears that the relationship between length of stay and total charges is statistically significant.

CHAPTER 9

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
See glossary (Appendix A).

3. Describe the circumstances under which it would be appropriate to use the chi-square test.
We can use the chi-square test of independence when we are interested in determining if two variables, such as age and sex, are related. We can use the chi-square test to compare frequencies on a variable from two or more independent populations. We can use the chi-square goodness-of-fit test when we want to compare an observed frequency distribution to a theoretical frequency distribution. And, we can use McNemar's chi square when we want to analyze frequencies for paired samples.

MULTIPLE CHOICE

1. e. c and d

3. a. observed frequencies are similar to the expected frequencies

5. b. 18

7. c. 30

9. a. reject the null hypothesis

PROBLEMS

1. You are assisting Dr. Hartman in studying the number of deaths due to acute myocardial infarctions (AMIs). Dr. Hartman is particularly interested in knowing if more men died from AMIs than women. To answer this question, you review discharges by sex for DRG123, Circulatory Disorders with AMI, Expired. Use the nonparametric procedure for chi-square to determine if there is an association between sex and deaths due to AMI at Critical Care Hospital. A frequency distribution of discharges by sex and age from DRG 123 appears in Table 9–A–1.

Table 9–A–1 Frequency Distribution of Discharges by Age and Gender, DRG 123 (SPSS Output)

Age * Gender Crosstabulation

		Gender		
		Female	*Male*	*Total*
Age	49	0	1	1
	50	0	1	1
	61	0	1	1
	66	0	1	1
	75	1	0	1
	76	0	1	1
	77	0	1	1
	88	1	0	1
	88	0	1	1
Total		2	7	9

a. State the null and alternative hypotheses.

$$H_0: p_1 = p_2$$
$$H_A: p_1 \neq p_2$$

b. State the alpha level.

$$\alpha = .05$$

c. What is the result of the chi-square test?

Gender

	Observed N	*Expected N*	*Residual*
Female	2	4.5	−2.5
Male	7	4.5	2.5
Total	9		

Test Statistics

	Gender
Chi-Square(*)	2.778
df	1
Asymp. Sig.	.096

* 2 cells (100.0%) have expected frequencies less than 5. The minimum expected cell frequency is 4.5.

d. State your conclusions.
$\chi^2 = 2.778$, df = 1, p = .096; We fail to reject the null hypothesis and conclude that it appears there is no difference in the proportion of males and females who expired as a result of an acute myocardial infarction. The result of the chi square test is not statistically significant. The results should be viewed with caution because of the number of cells that have expected frequency counts of less than five.

CHAPTER 10

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
See glossary (Appendix A).

3. Under what conditions is it appropriate to use the sign test?
The sign test is analogous to the *t*-test procedure for a single sample. It is used when the sample size is small and the underlying population distribution is not normal.

5. Under what conditions is it appropriate to use the Mann-Whitney Wilcoxon test?
The Mann-Whitney Wilcoxon test is used as an alternative to the *t*-test for two independent samples when the assumption for the *t*-test are violated. The Mann-Whitney-Wilcoxon test compares the medians of two independent samples.

MULTIPLE CHOICE

1. e. all of the above

3. d. all of the above

5. d. analysis of variance (ANOVA)

PROBLEMS

1. You are analyzing length of stay by physician for DRG 124, Circulatory Disorders, Except Acute Myocardial Infarction with Cardiac Catheterization and Complex Diagnosis. You are focusing on physicians 1630, 1830, and 3220. The lengths of stay for the patients of these three physicians appear in Table 10–A–1. Since the sample size for each physician is small, you decide to conduct the Kruskal-Wallis test to compare the mean lengths of stay.

Table 10–A–1 LOS Physician Crosstabulation (SPSS Output)

Count		*Physician*			
		2050	*2210*	*8290*	*Total*
LOS	1	1	2	2	5
	2	7	4	2	13
	3	1	2	1	4
	4	4	2	2	8
	5	0	1	0	1
	6	0	1	2	3
	8	0	1	0	1
Total		13	13	9	35

a. State the null and alternative hypotheses.

$$H_0: M_1 = M_2 = M_3$$
$$H_A: M_1 \neq M_2 \neq M_3$$

b. State the alpha level.

$$\alpha = .05$$

c. What is the result of the Kruskal-Wallis test?

Ranks

	Physician	*N*	*Mean Rank*
LOS	2050	13	16.42
	2210	13	19.00
	8290	9	18.83
	Total	35	

Test Statistics(*,†)

	LOS
Chi-Square	.527
df	2
Asymp. Sig.	.768

* Kruskal Wallis Test
† Grouping Variable: Physician

d. State your conclusions.

We fail to reject the null hypothesis. It appears that there is no significant difference in the average lengths of stay for the three physicians.

3. Review Exhibits 10–A–1 and 10–A–2 for discharges from DRG 127, Heart Failure and Shock. Use the Mann-Whitney *U* test to determine if there is a difference in age by sex and length of stay by sex for discharges from DRG 127.

a. State the null and alternative hypotheses.

$$H_0: M_Y = M_Y$$
$$H_A: M_Y \neq M_Y$$

b. State the alpha level.

$$\alpha = .05$$

c. What are the results of the Mann-Whitney *U* tests?

Mann-Whitney U for age and sex:

Ranks

	Gender	*N*	*Mean Rank*	*Sum of Ranks*
Age	Female	20	35.28	705.50
	Male	41	28.91	1185.50
	Total	61		

Test Statistics(*)

	Age
Mann-Whitney U	324.500
Wilcoxon W	1185.500
Z	−1.314
Asymp. Sig. (2-tailed)	.189

* Grouping Variable: Gender

Mann-Whitney U for length of stay and sex:

Ranks

	Gender	N	Mean Rank	Sum of Ranks
LOS	Female	20	37.33	746.50
	Male	41	27.91	1144.50
	Total	61		

Test Statistics(*)

	LOS
Mann-Whitney U	283.500
Wilcoxon W	1144.500
Z	−1.965
Asymp. Sig. (2-tailed)	.049

* Grouping Variable: Gender

d. State your conclusions.

For age and sex, the Mann-Whitney U statistic is 1185.5. The calculated z is −1.314, $p = .189$. We therefore fail to reject the null hypothesis and conclude that it appears that observed difference in median age by sex is not statistically significant for patients discharged from DRG 127.

For length of stay and sex, the Mann-Whitney U statistic is 1144.5. The calculated z is −1.965, $p = .049$. We therefore reject the null hypothesis and conclude that it appears that observed difference in median length of stay by sex is statistically significant for patients discharged from DRG 127.

e. Use the ANOVA procedure to run the same analyses. Compare the ANOVA results with the Mann-Whitney *U* test results.

						95% Confidence Interval for Mean			
		N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Age	Female	20	69.40	11.440	2.558	64.05	74.75	54	88
	Male	41	63.71	15.202	2.374	58.91	68.51	37	87
	Total	61	65.57	14.240	1.823	61.93	69.22	37	88
LOS	Female	20	8.70	8.761	1.959	4.60	12.80	1	36
	Male	41	5.15	4.163	.650	3.83	6.46	1	17
	Total	61	6.31	6.220	.796	4.72	7.90	1	36

ANOVA

		Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Age	Between Groups	435.630	1	435.630	2.191	.144
	Within Groups	11731.288	59	198.835		
	Total	12166.918	60			
LOS	Between Groups	169.760	1	169.760	4.656	.035
	Within Groups	2151.322	59	36.463		
	Total	2321.082	60			

The results are similar. For age and sex, $F = 2.191$, $df = 1, 59$, $p = .144$. We fail to reject the null hypothesis and conclude that the observed difference in mean age for men and women is not statistically significant for patients discharged from DRG 127.

For length of stay and sex, $F = 4.656$, $df = 1, 59$, $p = .035$. We reject the null hypothesis and conclude that the observed difference in mean length of stay for men and women is statistically significant for patients discharged from DRG 127.

Index

Q

R

S

T