EXERCISES

1.4.1 Using the table of random numbers, select a new random starting point, and draw another simple

random sample of size 10 from the data in Table 1.4.1. Record the ages of the subjects in this new

sample. Save your data for future use. What is the variable of interest in this exercise? What measurement

scale was used to obtain the measurements?

1.4.2 Select another simple random sample of size 10 from the population represented in Table 1.4.1.

Compare the subjects in this sample with those in the sample drawn in Exercise 1.4.1. Are there

any subjects who showed up in both samples? How many? Compare the ages of the subjects in

the two samples. How many ages in the first sample were duplicated in the second sample?

1.4.3 Using the table of random numbers, select a random sample and a systematic sample, each of size

15, from the data in Table 1.4.1. Visually compare the distributions of the two samples. Do they

appear similar? Which appears to be the best representation of the data?

1.4.4 Construct an example where it would be appropriate to use stratified sampling. Discuss how you

would use stratified random sampling and stratified sampling proportional to size with this example.

Which do you think would best represent the population that you described in your example?

Why?

1.5.1 Using the example of weight loss as an endpoint, discuss how you would use the scientific method

to test the observation that change in diet is related to weight loss. Include all of the steps, including

the hypothesis to be tested and the design of your experiment.

1.5.2 Continuing with Exercise 1.5.1, consider how you would use the scientific method to test the observation

that both exercise and change in diet are related to weight loss. Include all of the steps,

paying particular attention to how you might design the experiment and which hypotheses would

be testable given your design.

**REVIEW QUESTIONS AND EXERCISES**

1. Explain what is meant by descriptive statistics.

2. Explain what is meant by inferential statistics.

3. Define:

(a) Statistics (b) Biostatistics

(c) Variable (d) Quantitative variable

(e) Qualitative variable (f) Random variable

(g) Population (h) Finite population

(i) Infinite population ( j) Sample

(k) Discrete variable (l) Continuous variable

(m) Simple random sample (n) Sampling with replacement

(o) Sampling without replacement

4. Define the word measurement.

5. List, describe, and compare the four measurement scales.

6. For each of the following variables, indicate whether it is quantitative or qualitative and specify

the measurement scale that is employed when taking measurements on each:

(a) Class standing of the members of this class relative to each other

(b) Admitting diagnosis of patients admitted to a mental health clinic

(c) Weights of babies born in a hospital during a year

(d) Gender of babies born in a hospital during a year

(e) Range of motion of elbow joint of students enrolled in a university health sciences

curriculum

(f) Under-arm temperature of day-old infants born in a hospital

7. For each of the following situations, answer questions a through e:

(a) What is the sample in the study?

(b) What is the population?

(c) What is the variable of interest?

(d) How many measurements were used in calculating the reported results?

(e) What measurement scale was used?

Situation A. A study of 300 households in a small southern town revealed that 20 percent had at

least one school-age child present.

Situation B. A study of 250 patients admitted to a hospital during the past year revealed that, on

the average, the patients lived 15 miles from the hospital.

8. Consider the two situations given in Exercise 7. For Situation A describe how you would use a

stratified random sample to collect the data. For Situation B describe how you would use systematic

sampling of patient records to collect the data.

**EXERCISES**

2.3.1 In a study of the oral home care practice and reasons for seeking dental care among individuals

on renal dialysis, Atassi (A-1) studied 90 subjects on renal dialysis. The oral hygiene status of all

subjects was examined using a plaque index with a range of 0 to 3 10 = no soft plaque deposits,

abundance of soft plaque deposits). The following table shows the plaque index scores for

all 90 subjects

(a) Use these data to prepare:

A frequency distribution

A relative frequency distribution

A cumulative frequency distribution

A cumulative relative frequency distribution

A histogram

A frequency polygon

(b) What percentage of the measurements are less than 2.00?

(c) What proportion of the subjects have measurements greater than or equal to 1.50?

(d) What percentage of the measurements are between 1.50 and 1.99 inclusive?

(e) How many of the measurements are greater than 2.49?

(f) What proportion of the measurements are either less than 1.0 or greater than 2.49?

(g) Someone picks a measurement at random from this data set and asks you to guess the value.

What would be your answer? Why?

(h) Frequency distributions and their histograms may be described in a number of ways depending

on their shape. For example, they may be symmetric (the left half is at least approximately a

mirror image of the right half ), skewed to the left (the frequencies tend to increase as the measurements

increase in size), skewed to the right (the frequencies tend to decrease as the measurements

increase in size), or U-shaped (the frequencies are high at each end of the distribution and

small in the center). How would you describe the present distribution?

2.3.2 Janardhan et al. (A-2) conducted a study in which they measured incidental intracranial aneurysms

(IIAs) in 125 patients. The researchers examined postprocedural complications and concluded that

IIAs can be safely treated without causing mortality and with a lower complications rate than previously

reported. The following are the sizes (in millimeters) of the 159 IIAs in the sample.

(a) Use these data to prepare:

A frequency distribution

A relative frequency distribution

A cumulative frequency distribution

A cumulative relative frequency distribution

A histogram

A frequency polygon

(b) What percentage of the measurements are between 10 and 14.9 inclusive?

(c) How many observations are less than 20?

(d) What proportion of the measurements are greater than or equal to 25?

(e) What percentage of the measurements are either less than 10.0 or greater than 19.95?

(f) Refer to Exercise 2.3.1, part h. Describe the distribution of the size of the aneurysms in this

sample.

2.3.3 Hoekema et al. (A-3) studied the craniofacial morphology of patients diagnosed with obstructive

sleep apnea syndrome (OSAS) in healthy male subjects. One of the demographic variables the

researchers collected for all subjects was the Body Mass Index (calculated by dividing weight in kg

by the square of the patient’s height in cm). The following are the BMI values of 29 OSAS subjects.

(a) Use these data to construct:

A frequency distribution

A relative frequency distribution

A cumulative frequency distribution

A cumulative relative frequency distribution

A histogram

A frequency polygon

(b) What percentage of the measurements are less than 30?

(c) What percentage of the measurements are between 40.0 and 49.99 inclusive?

(d) What percentage of the measurements are greater than 34.99?

(e) Describe these data with respect to symmetry and skewness as discussed in Exercise 2.3.1,

part h.

(f) How many of the measurements are less than 40?

2.3.4 David Holben (A-4) studied selenium levels in beef raised in a low selenium region of the United

States. The goal of the study was to compare selenium levels in the region-raised beef to selenium

levels in cooked venison, squirrel, and beef from other regions of the United States. The data below

are the selenium levels calculated on a dry weight basis in for a sample of 53 regionraised

cattle.

(a) Use these data to construct:

A frequency distribution

A relative frequency distribution

A cumulative frequency distribution

A cumulative relative frequency distribution

A histogram

A frequency polygon

(b) Describe these data with respect to symmetry and skewness as discussed in Exercise 2.3.1,

part h.

(c) How many of the measurements are greater than 40?

(d) What percentage of the measurements are less than 25?

2.3.5 The following table shows the number of hours 45 hospital patients slept following the administration

of a certain anesthetic.

(a) From these data construct:

A frequency distribution

A relative frequency distribution

A histogram

A frequency polygon

(b) Describe these data relative to symmetry and skewness as discussed in Exercise 2.3.1, part h.

2.3.6 The following are the number of babies born during a year in 60 community hospitals.

(a) From these data construct:

A frequency distribution

A relative frequency distribution

A frequency polygon

(b) Describe these data relative to symmetry and skewness as discussed in Exercise 2.3.1, part h.

2.3.7 In a study of physical endurance levels of male college freshman, the following composite

endurance scores based on several exercise routines were collected.

(a) From these data construct:

A frequency distribution

A relative frequency distribution

A frequency polygon

A histogram

(b) Describe these data relative to symmetry and skewness as discussed in Exercise 2.3.1, part h.

2.3.8 The following are the ages of 30 patients seen in the emergency room of a hospital on a Friday

night. Construct a stem-and-leaf display from these data. Describe these data relative to symmetry

and skewness as discussed in Exercise 2.3.1, part h.

2.3.9 The following are the emergency room charges made to a sample of 25 patients at two city hospitals.

Construct a stem-and-leaf display for each set of data. What does a comparison of the two

displays suggest regarding the two hospitals? Describe the two sets of data with respect to symmetry

and skewness as discussed in Exercise 2.3.1, part h.

2.3.10 Refer to the ages of patients discussed in Example 1.4.1 and displayed in Table 1.4.1.

(a) Use class interval widths of 5 and construct:

A frequency distribution

A relative frequency distribution

A cumulative frequency distribution

A cumulative relative frequency distribution

A histogram

A frequency polygon

(b) Describe these data with respect to symmetry and skewness as discussed in Exercise 2.3.1,

part h.

2.3.11 The objectives of a study by Skjelbo et al. (A-5) were to examine (a) the relationship between

chloroguanide metabolism and efficacy in malaria prophylaxis and (b) the mephenytoin metabolism

and its relationship to chloroguanide metabolism among Tanzanians. From information provided

by urine specimens from the 216 subjects, the investigators computed the ratio of unchanged

S-mephenytoin to R-mephenytoin (S/R ratio). The results were as follows:

(a) From these data construct the following distributions: frequency, relative frequency, cumulative

frequency, and cumulative relative frequency; and the following graphs: histogram, frequency

polygon, and stem-and-leaf plot.

(b) Describe these data with respect to symmetry and skewness as discussed in Exercise 2.3.1,

part h.

(c) The investigators defined as poor metabolizers of mephenytoin any subject with an S/R mephenytoin

ratio greater than .9. How many and what percentage of the subjects were poor metabolizers?

(d) How many and what percentage of the subjects had ratios less than .7? Between .3 and .6999

inclusive? Greater than .4999?

2.3.12 Schmidt et al. (A-6) conducted a study to investigate whether autotransfusion of shed mediastinal

blood could reduce the number of patients needing homologous blood transfusion and reduce the

amount of transfused homologous blood if fixed transfusion criteria were used. The following table

shows the heights in centimeters of the 109 subjects of whom 97 were males.

(a) For these data construct the following distributions: frequency, relative frequency, cumulative

frequency, and cumulative relative frequency; and the following graphs: histogram, frequency polygon,

and stem-and-leaf plot.

(b) Describe these data with respect to symmetry and skewness as discussed in Exercise 2.3.1,

part h.

(c) How do you account for the shape of the distribution of these data?

(d) How tall were the tallest 6.42 percent of the subjects?

(e) How tall were the shortest 10.09 percent of the subjects?

**EXERCISES**

For each of the data sets in the following exercises compute (a) the mean, (b) the median, (c) the

mode, (d) the range, (e) the variance, (f) the standard deviation, (g) the coefficient of variation,

and (h) the interquartile range. Treat each data set as a sample. For those exercises for which you

think it would be appropriate, construct a box-and-whisker plot and discuss the usefulness in understanding

the nature of the data that this device provides. For each exercise select the measure of

central tendency that you think would be most appropriate for describing the data. Give reasons

to justify your choice.

2.5.1 Porcellini et al. (A-8) studied 13 HIV-positive patients who were treated with highly active antiretroviral

therapy (HAART) for at least 6 months. The CD4 T cell counts at baseline for the

13 subjects are listed below.

2.5.2 Shair and Jasper (A-9) investigated whether decreasing the venous return in young rats would affect

ultrasonic vocalizations (USVs). Their research showed no significant change in the number of

ultrasonic vocalizations when blood was removed from either the superior vena cava or the carotid

artery. Another important variable measured was the heart rate (bmp) during the withdrawal of

blood. The table below presents the heart rate of seven rat pups from the experiment involving the

carotid artery.

2.5.3 Butz et al. (A-10) evaluated the duration of benefit derived from the use of noninvasive positivepressure

ventilation by patients with amyotrophic lateral sclerosis on symptoms, quality of life,

and survival. One of the variables of interest is partial pressure of arterial carbon dioxide (PaCO2).

The values below (mm Hg) reflect the result of baseline testing on 30 subjects as established by

arterial blood gas analyses.

2.5.4 According to Starch et al. (A-11), hamstring tendon grafts have been the “weak link” in anterior

cruciate ligament reconstruction. In a controlled laboratory study, they compared two techniques

for reconstruction: either an interference screw or a central sleeve and screw on the tibial side. For

eight cadaveric knees, the measurements below represent the required force (in newtons) at which

initial failure of graft strands occurred for the central sleeve and screw technique.

2.5.5 Cardosi et al. (A-12) performed a 4-year retrospective review of 102 women undergoing radical

hysterectomy for cervical or endometrial cancer. Catheter-associated urinary tract infection was

observed in 12 of the subjects. Below are the numbers of postoperative days until diagnosis of the

infection for each subject experiencing an infection.

2.5.6 The purpose of a study by Nozawa et al. (A-13) was to evaluate the outcome of surgical repair

of a pars interarticularis defect by segmental wire fixation in young adults with lumbar spondylolysis.

The authors found that segmental wire fixation historically has been successful in the

treatment of nonathletes with spondylolysis, but no information existed on the results of this type

of surgery in athletes. In a retrospective study, the authors found 20 subjects who had the surgery

between 1993 and 2000. For these subjects, the data below represent the duration in months

of follow-up care after the operation.

2.5.7 See Exercise 2.3.1.

2.5.8 See Exercise 2.3.2.

2.5.9 See Exercise 2.3.3.

2.5.10 See Exercise 2.3.4.

2.5.11 See Exercise 2.3.5.

2.5.12 See Exercise 2.3.6.

2.5.13 See Exercise 2.3.7.

2.5.14 In a pilot study, Huizinga et al. (A-14) wanted to gain more insight into the psychosocial consequences

for children of a parent with cancer. For the study, 14 families participated in semistructured

interviews and completed standardized questionnaires. Below is the age of the sick parent

with cancer (in years) for the 14 families.

**REVIEW QUESTIONS AND EXERCISES**

1. Define:

(a) Stem-and-leaf display (b) Box-and-whisker plot

(c) Percentile (d) Quartile

(e) Location parameter (f) Exploratory data analysis

(g) Ordered array (h) Frequency distribution

(i) Relative frequency distribution ( j) Statistic

(k) Parameter (l) Frequency polygon

(m) True class limits (n) Histogram

2. Define and compare the characteristics of the mean, the median, and the mode.

3. What are the advantages and limitations of the range as a measure of dispersion?

4. Explain the rationale for using to compute the sample variance.

5. What is the purpose of the coefficient of variation?

6. What is the purpose of Sturges’s rule?

7. What is another name for the 50th percentile (second or middle quartile)?

8. Describe from your field of study a population of data where knowledge of the central tendency

and dispersion would be useful. Obtain real or realistic synthetic values from this population and

compute the mean, median, mode, variance, and standard deviation.

9. Collect a set of real, or realistic, data from your field of study and construct a frequency distribution,

a relative frequency distribution, a histogram, and a frequency polygon.

10. Compute the mean, median, mode, variance, and standard deviation for the data in Exercise 9.

11. Find an article in a journal from your field of study in which some measure of central tendency

and dispersion have been computed.

12. The purpose of a study by Tam et al. (A-15) was to investigate the wheelchair maneuvering in

individuals with lower-level spinal cord injury (SCI) and healthy controls. Subjects used a modified

wheelchair to incorporate a rigid seat surface to facilitate the specified experimental measurements.

Interface pressure measurement was recorded by using a high-resolution pressure-sensitive

mat with a spatial resolution of 4 sensors per square centimeter taped on the rigid seat support.

During static sitting conditions, average pressures were recorded under the ischial tuberosities. The

data for measurements of the left ischial tuberosity (in mm Hg) for the SCI and control groups

are shown below.

(a) Find the mean, median, variance, and standard deviation for the controls.

(b) Find the mean, median variance, and standard deviation for the SCI group.

(c) Construct a box-and-whisker plot for the controls.

(d) Construct a box-and-whisker plot for the SCI group.

(e) Do you believe there is a difference in pressure readings for controls and SCI subjects in this

study?

13. Johnson et al. (A-16) performed a retrospective review of 50 fetuses that underwent open fetal

myelomeningocele closure. The data below show the gestational age in weeks of the 50 fetuses

undergoing the procedure.

(a) Construct a stem-and-leaf plot for these gestational ages.

(b) Based on the stem-and-leaf plot, what one word would you use to describe the nature of the

data?

(c) Why do you think the stem-and-leaf plot looks the way it does?

(d) Compute the mean, median, variance, and standard deviation.

14. The following table gives the age distribution for the number of deaths in New York State due to

accidents for residents age 25 and older.

15. Krieser et al. (A-17) examined glomerular filtration rate (GFR) in pediatric renal transplant recipients.

GFR is an important parameter of renal function assessed in renal transplant recipients. The

following are measurements from 19 subjects of GFR measured with diethylenetriamine pentaacetic

acid. (Note: some subjects were measured more than once.)

(a) Compute mean, median, variance, standard deviation, and coefficient of variation.

(b) Construct a stem-and-leaf display.

(c) Construct a box-and-whisker plot.

(d) What percentage of the measurements is within one standard deviation of the mean? Two standard

deviations? Three standard deviations?

16. The following are the cystatin C levels (mg/L) for the patients described in Exercise 15 (A-17).

Cystatin C is a cationic basic protein that was investigated for its relationship to GFR levels. In

addition, creatinine levels are also given. (Note: Some subjects were measured more than

once.)

(a) For each variable, compute the mean, median, variance, standard deviation, and coefficient of

variation.

(b) For each variable, construct a stem-and-leaf display and a box-and-whisker plot.

(c) Which set of measurements is more variable, cystatin C or creatinine? On what do you base

your answer?

17. Give three synonyms for variation (variability).

18. The following table shows the age distribution of live births in Albany County, New York,

for 2000.

For these data construct a cumulative frequency distribution, a relative frequency distribution, and

a cumulative relative frequency distribution.

19. Spivack (A-18) investigated the severity of disease associated with C. difficile in pediatric inpatients.

One of the variables they examined was number of days patients experienced diarrhea. The

data for the 22 subjects in the study appear below. Compute the mean, median, variance, and standard

deviation.

20. Express in words the following properties of the sample mean:

(a)

(b)

(c)

21. Your statistics instructor tells you on the first day of class that there will be five tests during the

term. From the scores on these tests for each student, the instructor will compute a measure of

central tendency that will serve as the student’s final course grade. Before taking the first test, you

must choose whether you want your final grade to be the mean or the median of the five test

scores. Which would you choose? Why?

22. Consider the following possible class intervals for use in constructing a frequency distribution of

serum cholesterol levels of subjects who participated in a mass screening:

Which set of class intervals do you think is most appropriate for the purpose? Why? State specifically

for each one why you think the other two are less desirable.

23. On a statistics test students were asked to construct a frequency distribution of the blood creatine

levels (units/liter) for a sample of 300 healthy subjects. The mean was 95, and the standard deviation

was 40. The following class interval widths were used by the students:

(a) 1 (d) 15

(b) 5 (e) 20

(c) 10 (f) 25

Comment on the appropriateness of these choices of widths.

24. Give a health sciences–related example of a population of measurements for which the mean would

be a better measure of central tendency than the median.

25. Give a health sciences–related example of a population of measurements for which the median

would be a better measure of central tendency than the mean.

26. Indicate for the following variables which you think would be a better measure of central tendency,

the mean, the median, or mode, and justify your choice:

(a) Annual incomes of licensed practical nurses in the Southeast.

(b) Diagnoses of patients seen in the emergency department of a large city hospital.

(c) Weights of high-school male basketball players.

27. Refer to Exercise 2.3.11. Compute the mean, median, variance, standard deviation, first quartile, third

quartile, and interquartile range. Construct a boxplot of the data. Are the mode, median, and mean

equal? If not, explain why. Discuss the data in terms of variability. Compare the IQR with the range.

What does the comparison tell you about the variability of the observations?

28. Refer to Exercise 2.3.12. Compute the mean, median, variance, standard deviation, first quartile, third

quartile, and interquartile range. Construct a boxplot of the data. Are the mode, median, and mean

equal? If not, explain why. Discuss the data in terms of variability. Compare the IQR with the range.

What does the comparison tell you about the variability of the observations?

29. Thilothammal et al. (A-19) designed a study to determine the efficacy of BCG (bacillus Calmette-

Gu rin) vaccine in preventing tuberculous meningitis. Among the data collected on each subject

was a measure of nutritional status (actual weight expressed as a percentage of expected weight for

actual height). The following table shows the nutritional status values of the 107 cases studied.

(a) For these data compute the following descriptive measures: mean, median, mode, variance,

standard deviation, range, first quartile, third quartile, and IQR.

(b) Construct the following graphs for the data: histogram, frequency polygon, stem-and-leaf plot,

and boxplot.

(c) Discuss the data in terms of variability. Compare the IQR with the range. What does the comparison

tell you about the variability of the observations?

(d) What proportion of the measurements are within one standard deviation of the mean? Two

standard deviations of the mean? Three standard deviations of the mean?

(e) What proportion of the measurements are less than 100?

(f) What proportion of the measurements are less than 50?

Exercises for Use with Large Data Sets Available on the Following Website:

www.wiley.com/college/daniel

Refer to the dataset NCBIRTH800. The North Carolina State Center for Health Statistics and

Howard W. Odum Institute for Research in Social Science at the University of North Carolina at

Chapel Hill (A-20) make publicly available birth and infant death data for all children born in the

state of North Carolina. These data can be accessed at www.irss.unc.edu/ncvital/bfd1down.html.

Records on birth data go back to 1968. This comprehensive data set for the births in 2001 contains

120,300 records. The data represents a random sample of 800 of those births and selected

variables. The variables are as follows:

Variable Label Description

PLURALITY Number of children born of the pregnancy

SEX Sex of child

MAGE Age of mother (years)

WEEKS Completed weeks of gestation (weeks)

MARITAL Marital status

RACEMOM Race of mother non-White,

Indian,

Asian or Pacific Islander)

HISPMOM Mother of Hispanic origin -Hispanic,

and unknown Hispanic,

American, classifiable)

GAINED Weight gained during pregnancy (pounds)

SMOKE did not smoke during pregnancy

did smoke during pregnancy

DRINK did not consume alcohol during pregnancy

did consume alcohol during pregnancy

TOUNCES Weight of child (ounces)

TGRAMS Weight of child (grams)

LOW was not low birth weight

was low birth weight

PREMIE was not premature

was premature

Premature defined at 36 weeks or sooner

For the variables of MAGE, WEEKS, GAINED, TOUNCES, and TGRAMS:

1. Calculate the mean, median, standard deviation, IQR, and range.

2. For each, construct a histogram and comment on the shape of the distribution.

3. Do the histograms for TOUNCES and TGRAMS look strikingly similar? Why?

4. Construct box-and-whisker plots for all four variables.

5. Construct side-by-side box-and-whisker plots for the variable of TOUNCES for women who admitted

to smoking and women who did not admit to smoking. Do you see a difference in birth weight

in the two groups? Which group has more variability?

6. Construct side-by-side box-and-whisker plots for the variable of MAGE for women who are and

are not married. Do you see a difference in ages in the two groups? Which group has more variability?

Are the results surprising?

7. Calculate the skewness and kurtosis of the data set. What do they indicate?

**EXERCISES**

3.4.1 In a study of violent victimization of women and men, Porcerelli et al. (A-2) collected information

from 679 women and 345 men aged 18 to 64 years at several family practice centers in the metropolitan

Detroit area. Patients filled out a health history questionnaire that included a question about

victimization. The following table shows the sample subjects cross-classified by sex and the type

of violent victimization reported. The victimization categories are defined as no victimization, partner

victimization (and not by others), victimization by persons other than partners (friends, family

members, or strangers), and those who reported multiple victimization.

(a) Suppose we pick a subject at random from this group. What is the probability that this subject

will be a woman?

(b) What do we call the probability calculated in part a?

(c) Show how to calculate the probability asked for in part a by two additional methods.

(d) If we pick a subject at random, what is the probability that the subject will be a woman and

have experienced partner abuse?

(e) What do we call the probability calculated in part d?

(f) Suppose we picked a man at random. Knowing this information, what is the probability that

he experienced abuse from nonpartners?

(g) What do we call the probability calculated in part f?

(h) Suppose we pick a subject at random. What is the probability that it is a man or someone

who experienced abuse from a partner?

(i) What do we call the method by which you obtained the probability in part h?

3.4.2 Fernando et al. (A-3) studied drug-sharing among injection drug users in the South Bronx in

New York City. Drug users in New York City use the term “split a bag” or “get down on a

bag” to refer to the practice of dividing a bag of heroin or other injectable substances. A common

practice includes splitting drugs after they are dissolved in a common cooker, a procedure

with considerable HIV risk. Although this practice is common, little is known about the prevalence

of such practices. The researchers asked injection drug users in four neighborhoods in

the South Bronx if they ever “got down on” drugs in bags or shots. The results classified by

gender and splitting practice are given below:

(a) How many marginal probabilities can be calculated from these data? State each in probability

notation and do the calculations.

(b) How many joint probabilities can be calculated? State each in probability notation and do the

calculations.

(c) How many conditional probabilities can be calculated? State each in probability notation and

do the calculations.

(d) Use the multiplication rule to find the probability that a person picked at random never split

drugs and is female.

(e) What do we call the probability calculated in part d?

(f) Use the multiplication rule to find the probability that a person picked at random is male,

given that he admits to splitting drugs.

(g) What do we call the probability calculated in part f?

3.4.3 Refer to the data in Exercise 3.4.2. State the following probabilities in words and calculate:

(a) P(Male Split Drugs)

(b) P(Male Split Drugs)

(c) P(Male Split Drugs)

(d) P(Male)

3.4.4 Laveist and Nuru-Jeter (A-4) conducted a study to determine if doctor–patient race concordance was

associated with greater satisfaction with care. Toward that end, they collected a national sample of

African-American, Caucasian, Hispanic, and Asian-American respondents. The following table classifies

the race of the subjects as well as the race of their physician:

Greater Satisfaction with Care?” Journal of Health and Social Behavior, 43 (2002), 296–306.

(a) What is the probability that a randomly selected subject will have an Asian/Pacific-Islander

physician?

(b) What is the probability that an African-American subject will have an African-American physician?

(c) What is the probability that a randomly selected subject in the study will be Asian-American

and have an Asian/Pacific-Islander physician?

(d) What is the probability that a subject chosen at random will be Hispanic or have a Hispanic

physician?

(e) Use the concept of complementary events to find the probability that a subject chosen at random

in the study does not have a white physician.

3.4.5 If the probability of left-handedness in a certain group of people is .05, what is the probability of

right-handedness (assuming no ambidexterity)?

3.4.6 The probability is .6 that a patient selected at random from the current residents of a certain hospital

will be a male. The probability that the patient will be a male who is in for surgery is .2. A

patient randomly selected from current residents is found to be a male; what is the probability that

the patient is in the hospital for surgery?

3.4.7 In a certain population of hospital patients the probability is .35 that a randomly selected patient

will have heart disease. The probability is .86 that a patient with heart disease is a smoker. What

is the probability that a patient randomly selected from the population will be a smoker and have

heart disease?

**EXERCISES**

3.5.1 A medical research team wishes to assess the usefulness of a certain symptom (call it S) in the

diagnosis of a particular disease. In a random sample of 775 patients with the disease, 744 reported

having the symptom. In an independent random sample of 1380 subjects without the disease,

21 reported that they had the symptom.

(a) In the context of this exercise, what is a false positive?

(b) What is a false negative?

(c) Compute the sensitivity of the symptom.

(d) Compute the specificity of the symptom.

(e) Suppose it is known that the rate of the disease in the general population is .001. What is the

predictive value positive of the symptom?

(f) What is the predictive value negative of the symptom?

(g) Find the predictive value positive and the predictive value negative for the symptom for the

following hypothetical disease rates: .0001, .01, and .10.

(h) What do you conclude about the predictive value of the symptom on the basis of the results

obtained in part g?

3.5.2 In an article entitled “Bucket-Handle Meniscal Tears of the Knee: Sensitivity and Specificity of

MRI signs,” Dorsay and Helms (A-6) performed a retrospective study of 71 knees scanned by

MRI. One of the indicators they examined was the absence of the “bow-tie sign” in the MRI as

evidence of a bucket-handle or “bucket-handle type” tear of the meniscus. In the study, surgery

confirmed that 43 of the 71 cases were bucket-handle tears. The cases may be cross-classified by

“bow-tie sign” status and surgical results as follows:

Source: Theodore A. Dorsay and Clyde A. Helms, “Bucket-Handle Meniscal Tears of the Knee: Sensitivity

and Specificity of MRI Signs,” Skeletal Radiology, 32 (2003), 266–272.

(a) What is the sensitivity of testing to see if the absent bow-tie sign indicates a meniscal tear?

(b) What is the specificity of testing to see if the absent bow-tie sign indicates a meniscal tear?

(c) What additional information would you need to determine the predictive value of the test?

3.5.3 Oexle et al. (A-7) calculated the negative predictive value of a test for carriers of X-linked ornithine

transcarbamylase deficiency (OTCD—a disorder of the urea cycle). A test known as the “allopurinol

test” is often used as a screening device of potential carriers whose relatives are OTCD patients.

They cited a study by Brusilow and Horwich (A-8) that estimated the sensitivity of the allopurinol

test as .927. Oexle et al. themselves estimated the specificity of the allopurinol test as .997.

Also they estimated the prevalence in the population of individuals with OTCD as 1!32000. Use

this information and Bayes’s theorem to calculate the predictive value negative of the allopurinol

screening test.

**REVIEW QUESTIONS AND EXERCISES**

1. Define the following:

(a) Probability (b) Objective probability

(c) Subjective probability (d) Classical probability

(e) The relative frequency concept of probability (f) Mutually exclusive events

(g) Independence (h) Marginal probability

(i) Joint probability ( j) Conditional probability

(k) The addition rule (l) The multiplication rule

(m) Complementary events (n) False positive

(o) False negative (p) Sensitivity

(q) Specificity (r) Predictive value positive

(s) Predictive value negative (t) Bayes’s theorem

2. Name and explain the three properties of probability.

3. Coughlin et al. (A-9) examined the breast and cervical screening practices of Hispanic and non-Hispanic

women in counties that approximate the U.S. southern border region. The study used data from the

Behavioral Risk Factor Surveillance System surveys of adults age years or older conducted in

1999 and 2000. The table below reports the number of observations of Hispanic and non-Hispanic

women who had received a mammogram in the past 2 years cross-classified with marital status.

(a) We select at random a subject who had a mammogram. What is the probability that she is

divorced or separated?

(b) We select at random a subject who had a mammogram and learn that she is Hispanic. With

that information, what is the probability that she is married?

(c) We select at random a subject who had a mammogram. What is the probability that she is

non-Hispanic and divorced or separated?

(d) We select at random a subject who had a mammogram. What is the probability that she is

Hispanic or she is widowed?

(e) We select at random a subject who had a mammogram. What is the probability that she is not

married?

4. Swor et al. (A-10) looked at the effectiveness of cardiopulmonary resuscitation (CPR) training in

people over 55 years old. They compared the skill retention rates of subjects in this age group

who completed a course in traditional CPR instruction with those who received chest-compression

only cardiopulmonary resuscitation (CC-CPR). Independent groups were tested 3 months after

training. The table below shows the skill retention numbers in regard to overall competence as

assessed by video ratings done by two video evaluators.

(a) Find the following probabilities and explain their meaning:

1. A randomly selected subject was enrolled in the CC-CPR class.

2. A randomly selected subject was rated competent.

3. A randomly selected subject was rated competent and was enrolled in the CPR course.

4. A randomly selected subject was rated competent or was enrolled in CC-CPR.

5. A randomly selected subject was rated competent given that he or she enrolled in the

CC-CPR course.

(b) We define the following events to be

A a subject enrolled in the CPR course

B a subject enrolled in the CC-CPR course

C a subject was evaluated as competent

D a subject was evaluated as not competent

Then explain why each of the following equations is or is not a true statement:

1. 2.

3. 4.

5. 6.

7. 8.

9. P1A   D2 = P1A2 P1A ƒ D2

P1A   B2 = 0 P1C   B2 = P1B2 P1C ƒ B2

P1D ƒ A2 = P1D2 P1C   B2 = P1C2P1B2

P1A2 = P1A   C2 + P1A   D2 P1B   C2 = P1B2 + P1C2

P1A   C2 = P1C   A2 P1A   B2 = P1B   A2

5. Pillmann et al. (A-11) studied patients with acute brief episodes of psychoses. The researchers

classified subjects into four personality types: obsessoid, asthenic/low self-confident, asthenic/high

self-confident, nervous/tense, and undeterminable. The table below cross-classifies these personality

types with three groups of subjects—those with acute and transient psychotic disorders (ATPD),

those with “positive” schizophrenia (PS), and those with bipolar schizo-affective disorder (BSAD):

Find the following probabilities if a subject in this study is chosen at random:

(a) (b) (c) (d)

(e) (f) (g) (h)

6. A certain county health department has received 25 applications for an opening that exists for a

public health nurse. Of these applicants 10 are over 30 and 15 are under 30. Seventeen hold bachelor’s

degrees only, and eight have master’s degrees. Of those under 30, six have master’s degrees.

If a selection from among these 25 applicants is made at random, what is the probability that a

person over 30 or a person with a master’s degree will be selected?

7. The following table shows 1000 nursing school applicants classified according to scores made on

a college entrance examination and the quality of the high school from which they graduated, as

rated by a group of educators:

(a) Calculate the probability that an applicant picked at random from this group:

1. Made a low score on the examination.

2. Graduated from a superior high school.

3. Made a low score on the examination and graduated from a superior high school.

4. Made a low score on the examination given that he or she graduated from a superior high

school.

5. Made a high score or graduated from a superior high school.

(b) Calculate the following probabilities:

1. 2. 3.

4. 5. 6.

8. If the probability that a public health nurse will find a client at home is .7, what is the probability

(assuming independence) that on two home visits made in a day both clients will be home?

9. For a variety of reasons, self-reported disease outcomes are frequently used without verification in epidemiologic

research. In a study by Parikh-Patel et al. (A-12), researchers looked at the relationship

between self-reported cancer cases and actual cases. They used the self-reported cancer data from a

California Teachers Study and validated the cancer cases by using the California Cancer Registry data.

The following table reports their findings for breast cancer:

(a) Let A be the event of reporting breast cancer in the California Teachers Study. Find the probability

of A in this study.

(b) Let B be the event of having breast cancer confirmed in the California Cancer Registry. Find

the probability of B in this study.

(c) Find

(d) Find

(e) Find

(f) Find the sensitivity of using self-reported breast cancer as a predictor of actual breast cancer

in the California registry.

(g) Find the specificity of using self-reported breast cancer as a predictor of actual breast cancer

in the California registry.

10. In a certain population the probability that a randomly selected subject will have been exposed

to a certain allergen and experience a reaction to the allergen is .60. The probability is .8 that a

subject exposed to the allergen will experience an allergic reaction. If a subject is selected at random

from this population, what is the probability that he or she will have been exposed to the

allergen?

11. Suppose that 3 percent of the people in a population of adults have attempted suicide. It is also

known that 20 percent of the population are living below the poverty level. If these two events

are independent, what is the probability that a person selected at random from the population

will have attempted suicide and be living below the poverty level?

12. In a certain population of women 4 percent have had breast cancer, 20 percent are smokers, and

3 percent are smokers and have had breast cancer. A woman is selected at random from the population.

What is the probability that she has had breast cancer or smokes or both?

13. The probability that a person selected at random from a population will exhibit the classic symptom

of a certain disease is .2, and the probability that a person selected at random has the disease

is .23. The probability that a person who has the symptom also has the disease is .18. A person

selected at random from the population does not have the symptom. What is the probability that

the person has the disease?

14. For a certain population we define the following events for mother’s age at time of giving birth:

A under 20 years; B 20–24 years; C 25–29 years; D 30–44 years. Are the events A, B,

C, and D pairwise mutually exclusive?

15. Refer to Exercise 14. State in words the event

16. Refer to Exercise 14. State in words the event

17. Refer to Exercise 14. Comment on the event

18. For a certain population we define the following events with respect to plasma lipoprotein levels

(mg"dl): A (10–15); Are the events A and B mutually exclusive? A

and C? B and C? Explain your answer to each question.

19. Refer to Exercise 18. State in words the meaning of the following events:

(a) (b) (c) (d)

20. Refer to Exercise 18. State in words the meaning of the following events:

(a) (b) (c)

21. Rothenberg et al. (A-13) investigated the effectiveness of using the Hologic Sahara Sonometer, a

portable device that measures bone mineral density (BMD) in the ankle, in predicting a fracture.

They used a Hologic estimated bone mineral density value of .57 as a cutoff. The results of the

investigation yielded the following data:

(a) Calculate the sensitivity of using a BMD value of .57 as a cutoff value for predicting fracture

and interpret your results.

(b) Calculate the specificity of using a BMD value of .57 as a cutoff value for predicting fracture

and interpret your results.

22. Verma et al. (A-14) examined the use of heparin-PF4 ELISA screening for heparin-induced thrombocytopenia

(HIT) in critically ill patients. Using C-serotonin release assay (SRA) as the way of

validating HIT, the authors found that in 31 patients tested negative by SRA, 22 also tested negative

by heparin-PF4 ELISA.

(a) Calculate the specificity of the heparin-PF4 ELISA testing for HIT.

(b) Using a “literature-derived sensitivity” of 95 percent and a prior probability of HIT occurrence

as 3.1 percent, find the positive predictive value.

(c) Using the same information as part (b), find the negative predictive value.

23. The sensitivity of a screening test is .95, and its specificity is .85. The rate of the disease for which

the test is used is .002. What is the predictive value positive of the test?

Exercises for Use with Large Data Sets Available on the Following Website:

www.wiley.com/college/daniel

Refer to the random sample of 800 subjects from the North Carolina birth registry we investigated

in the Chapter 2 review exercises.

1. Create a table that cross-tabulates the counts of mothers in the classifications of whether the baby

was premature or not (PREMIE) and whether the mother admitted to smoking during pregnancy

(SMOKE) or not.

(a) Find the probability that a mother in this sample admitted to smoking.

(b) Find the probability that a mother in this sample had a premature baby.

(c) Find the probability that a mother in the sample had a premature baby given that the mother

admitted to smoking.

(d) Find the probability that a mother in the sample had a premature baby given that the mother

did not admit to smoking.

(e) Find the probability that a mother in the sample had a premature baby or that the mother did

not admit to smoking.

2. Create a table that cross-tabulates the counts of each mother’s marital status (MARITAL) and

whether she had a low birth weight baby (LOW).

(a) Find the probability a mother selected at random in this sample had a low birth weight baby.

(b) Find the probability a mother selected at random in this sample was married.

(c) Find the probability a mother selected at random in this sample had a low birth weight child

given that she was married.

(d) Find the probability a mother selected at random in this sample had a low birth weight child

given that she was not married.

(e) Find the probability a mother selected at random in this sample had a low birth weight child

and the mother was married.

**EXERCISES**

4.2.1 In a study by Cross et al. (A-2), patients who were involved in problem gambling treatment were

asked about co-occurring drug and alcohol addictions. Let the discrete random variable X represent

the number of co-occurring addictive substances used by the subjects. Table 4.2.4 summarizes the

frequency distribution for this random variable.

(a) Construct a table of the relative frequency and the cumulative frequency for this discrete

distribution.

(b) Construct a graph of the probability distribution and a graph representing the cumulative

probability distribution for these data.

4.2.2 Refer to Exercise 4.2.1.

(a) What is probability that an individual selected at random used five addictive substances?

(b) What is the probability that an individual selected at random used fewer than three addictive

substances?

(c) What is the probability that an individual selected at random used more than six addictive

substances?

(d) What is the probability that an individual selected at random used between two and five

addictive substances, inclusive?

4.2.3 Refer to Exercise 4.2.1. Find the mean, variance, and standard deviation of this frequency distribution.

**EXERCISES**

In each of the following exercises, assume that N is sufficiently large relative to n that the binomial

distribution may be used to find the desired probabilities.

4.3.1 Based on data collected by the National Center for Health Statistics and made available to the

public in the Sample Adult database (A-5), an estimate of the percentage of adults who have at

some point in their life been told they have hypertension is 23.53 percent. If we select a simple

random sample of 20 U.S. adults and assume that the probability that each has been told that he

or she has hypertension is .24, find the probability that the number of people in the sample who

have been told that they have hypertension will be:

(a) Exactly three (b) Three or more

(c) Fewer than three (d) Between three and seven, inclusive

4.3.2 Refer to Exercise 4.3.1. How many adults who have been told that they have hypertension would

you expect to find in a sample of 20?

4.3.3 Refer to Exercise 4.3.1. Suppose we select a simple random sample of five adults. Use Equation

4.3.2 to find the probability that, in the sample, the number of people who have been told that

they have hypertension will be:

(a) Zero (b) More than one

(c) Between one and three, inclusive (d) Two or fewer

(e) Five

4.3.4 The same survey database cited in exercise 4.3.1 (A-5) shows that 32 percent of U.S. adults indicated

that they have been tested for HIV at some point in their life. Consider a simple random

sample of 15 adults selected at that time. Find the probability that the number of adults who have

been tested for HIV in the sample would be:

(a) Three (b) Less than five

(c) Between five and nine, inclusive (d) More than five, but less than 10

(e) Six or more

4.3.5 Refer to Exercise 4.3.4. Find the mean and variance of the number of people tested for HIV in

samples of size 15.

4.3.6 Refer to Exercise 4.3.4. Suppose we were to take a simple random sample of 25 adults today and

find that two have been tested for HIV at some point in their life. Would these results be surprising?

Why or why not?

4.3.7 Coughlin et al. (A-6) estimated the percentage of women living in border counties along the southern

United States with Mexico (designated counties in California, Arizona, New Mexico, and

Texas) who have less than a high school education to be 18.7. Assume the corresponding probability

is .19. Suppose we select three women at random. Find the probability that the number with

less than a high-school education is:

(a) Exactly zero (b) Exactly one

(c) More than one (d) Two or fewer

(e) Two or three (f) Exactly three

4.3.8 In a survey of nursing students pursuing a master’s degree, 75 percent stated that they expect

to be promoted to a higher position within one month after receiving the degree. If this percentage

holds for the entire population, find, for a sample of 15, the probability that the number

expecting a promotion within a month after receiving their degree is:

(a) Six (b) At least seven

(c) No more than five (d) Between six and nine, inclusive

4.3.9 Given the binomial parameters and show by means of the binomial expansion given

EXERCISES

4.4.1 Singh et al. (A-8) looked at the occurrence of retinal capillary hemangioma (RCH) in patients with

von Hippel–Lindau (VHL) disease. RCH is a benign vascular tumor of the retina. Using a retrospective

consecutive case series review, the researchers found that the number of RCH tumor

incidents followed a Poisson distribution with tumors per eye for patients with VHL. Using

this model, find the probability that in a randomly selected patient with VHL:

(a) There are exactly five occurrences of tumors per eye.

(b) There are more than five occurrences of tumors per eye.

(c) There are fewer than five occurrences of tumors per eye.

(d) There are between five and seven occurrences of tumors per eye, inclusive.

4.4.2 Tubert-Bitter et al. (A-9) found that the number of serious gastrointestinal reactions reported to the

British Committee on Safety of Medicine was 538 for 9,160,000 prescriptions of the anti-inflammatory

drug piroxicam. This corresponds to a rate of .058 gastrointestinal reactions per 1000 prescriptions

written. Using a Poisson model for probability, with find the probability of

(a) Exactly one gastrointestinal reaction in 1000 prescriptions

(b) Exactly two gastrointestinal reactions in 1000 prescriptions

(c) No gastrointestinal reactions in 1000 prescriptions

(d) At least one gastrointestinal reaction in 1000 prescriptions

4.4.3 If the mean number of serious accidents per year in a large factory (where the number of employees

remains constant) is five, find the probability that in the current year there will be:

(a) Exactly seven accidents (b) Ten or more accidents

(c) No accidents (d) Fewer than five accidents

4.4.4 In a study of the effectiveness of an insecticide against a certain insect, a large area of land was

sprayed. Later the area was examined for live insects by randomly selecting squares and counting

the number of live insects per square. Past experience has shown the average number of live

insects per square after spraying to be .5. If the number of live insects per square follows a Poisson

distribution, find the probability that a selected square will contain:

(a) Exactly one live insect (b) No live insects

(c) Exactly four live insects (d) One or more live insects

4.4.5 In a certain population an average of 13 new cases of esophageal cancer are diagnosed each year.

If the annual incidence of esophageal cancer follows a Poisson distribution, find the probability

that in a given year the number of newly diagnosed cases of esophageal cancer will be:

(a) Exactly 10 (b) At least eight

(c) No more than 12 (d) Between nine and 15, inclusive

(e) Fewer than seven

**EXERCISES**

Given the standard normal distribution find:

4.6.1 The area under the curve between and

4.6.2 The probability that a z picked at random will have a value between and

4.6.3 4.6.4

4.6.5 4.6.6

4.6.7 4.6.8

4.6.9 4.6.10

Given the following probabilities, find z1:

4.6.11 4.6.12

4.6.13 4.6.14

4.6.15

4.7 NORMAL DISTRIBUTION APPLICATIONS

Although its importance in the field of statistics is indisputable, one should realize that

the normal distribution is not a law that is adhered to by all measurable characteristics

occurring in nature. It is true, however, that many of these characteristics are

P1-z 1 … z … z 12 = .8132

P1z 7 z 12 = .0384 P1z 1 … z … 2.982 = .1117

P1z … z 12 = .0055 P1-2.67 … z … z 12 = .9718

P1-1.65 … z … 1.652 P1z = .742

P1-1.96 … z … 1.962 P1-2.58 … z … 2.582

P1z 6 -2.332 P1z 6 2.332

P1z   .552 P1z   -.552

**EXERCISES**

4.7.1 For another subject (a 29-year-old male) in the study by Diskin et al. (A-11), acetone levels were

normally distributed with a mean of 870 and a standard deviation of 211 ppb. Find the probability

that on a given day the subject’s acetone level is:

(a) Between 600 and 1000 ppb

(b) Over 900 ppb

(c) Under 500 ppb

(d) Between 900 and 1100 ppb

4.7.2 In the study of fingerprints, an important quantitative characteristic is the total ridge count for the

10 fingers of an individual. Suppose that the total ridge counts of individuals in a certain population

are approximately normally distributed with a mean of 140 and a standard deviation of 50.

Find the probability that an individual picked at random from this population will have a ridge

count of:

(a) 200 or more

(b) Less than 100

(c) Between 100 and 200

(d) Between 200 and 250

(e) In a population of 10,000 people how many would you expect to have a ridge count of 200

or more?

4.7.3 One of the variables collected in the North Carolina Birth Registry data (A-3) is pounds gained

during pregnancy. According to data from the entire registry for 2001, the number of pounds gained

during pregnancy was approximately normally distributed with a mean of 30.23 pounds and a standard

deviation of 13.84 pounds. Calculate the probability that a randomly selected mother in North

Carolina in 2001 gained:

(a) Less than 15 pounds during pregnancy (b) More than 40 pounds

(c) Between 14 and 40 pounds (d) Less than 10 pounds

(e) Between 10 and 20 pounds

4.7.4 Suppose the average length of stay in a chronic disease hospital of a certain type of patient is

60 days with a standard deviation of 15. If it is reasonable to assume an approximately normal

distribution of lengths of stay, find the probability that a randomly selected patient from this group

will have a length of stay:

(a) Greater than 50 days (b) Less than 30 days

(c) Between 30 and 60 days (d) Greater than 90 days

4.7.5 If the total cholesterol values for a certain population are approximately normally distributed with a

mean of 200 mg/100 ml and a standard deviation of 20 mg/100 ml, find the probability that an individual

picked at random from this population will have a cholesterol value:

(a) Between 180 and 200 mg/100 ml (b) Greater than 225 mg/100 ml

(c) Less than 150 mg/100 ml (d) Between 190 and 210 mg/100 ml

4.7.6 Given a normally distributed population with a mean of 75 and a variance of 625, find:

(a) (b)

(c) (d)

(e)

4.7.7 The weights of a certain population of young adult females are approximately normally distributed

with a mean of 132 pounds and a standard deviation of 15. Find the probability that a subject

selected at random from this population will weigh:

(a) More than 155 pounds (b) 100 pounds or less

(c) Between 105 and 145 pounds

REVIEW QUESTIONS AND EXERCISES

1. What is a discrete random variable? Give three examples that are of interest to the health professional.

2. What is a continuous random variable? Give three examples of interest to the health professional.

3. Define the probability distribution of a discrete random variable.

4. Define the probability distribution of a continuous random variable.

5. What is a cumulative probability distribution?

6. What is a Bernoulli trial?

7. Describe the binomial distribution.

8. Give an example of a random variable that you think follows a binomial distribution.

9. Describe the Poisson distribution.

10. Give an example of a random variable that you think is distributed according to the Poisson law.

11. Describe the normal distribution.

12. Describe the standard normal distribution and tell how it is used in statistics.

13. Give an example of a random variable that you think is, at least approximately, normally distributed.

14. Using the data of your answer to Question 13, demonstrate the use of the standard normal distribution

in answering probability questions related to the variable selected.

15. Kanjanarat et al. (A-12) estimate the rate of preventable adverse drug events (ADEs) in hospitals

to be 35.2 percent. Preventable ADEs typically result from inappropriate care or medication errors,

which include errors of commission and errors of omission. Suppose that 10 hospital patients experiencing

an ADE are chosen at random. Let and calculate the probability that:

(a) Exactly seven of those drug events were preventable

(b) More than half of those drug events were preventable

(c) None of those drug events were preventable

(d) Between three and six inclusive were preventable

16. In a poll conducted by the Pew Research Center in 2003 (A-13), a national sample of adults answered

the following question, “All in all, do you strongly favor, favor, oppose, or strongly oppose . . . making

it legal for doctors to give terminally ill patients the means to end their lives?” The results showed

that 43 percent of the sample subjects answered “strongly favor” or “favor” to this question. If 12

subjects represented by this sample are chosen at random, calculate the probability that:

(a) Exactly two of the respondents answer “strongly favor” or “favor”

(b) No more than two of the respondents answer “strongly favor” or “favor”

(c) Between five and nine inclusive answer “strongly favor” or “favor”

17. In a study by Thomas et al. (A-14) the Poisson distribution was used to model the number of

patients per month referred to an oncologist. The researchers use a rate of 15.8 patients per month

that are referred to the oncologist. Use Table C in the Appendix and a rate of 16 patients per month

to calculate the probability that in a month:

(a) Exactly 10 patients are referred to an oncologist

(b) Between five and 15 inclusive are referred to an oncologist

(c) More than 10 are referred to an oncologist

(d) Less than eight are referred to an oncologist

(e) Less than 12, but more than eight are referred to an oncologist

18. On the average, two students per hour report for treatment to the first-aid room of a large elementary

school.

(a) What is the probability that during a given hour three students come to the first-aid room for

treatment?

(b) What is the probability that during a given hour two or fewer students will report to the firstaid

room?

(c) What is the probability that during a given hour between three and five students, inclusive,

will report to the first-aid room?

19. A Harris Interactive poll conducted in Fall, 2002 (A-15) via a national telephone survey of adults

asked, “Do you think adults should be allowed to legally use marijuana for medical purposes if

their doctor prescribes it, or do you think that marijuana should remain illegal even for medical

purposes.” The results showed that 80 percent of respondents answered “Yes” to the above question.

Assuming 80 percent of Americans would say “Yes” to the above question, find the probability

when eight Americans are chosen at random that:

(a) Six or fewer said “Yes” (b) Seven or more said “Yes”

(c) All eight said “Yes” (d) Fewer than four said “Yes”

(e) Between four and seven inclusive said “Yes”

20. In a study of the relationship between measles vaccination and Guillain-Barr  syndrome (GBS), Silveira

et al. (A-16) used a Poisson model in the examination of the occurrence of GBS during latent

periods after vaccinations. They conducted their study in Argentina, Brazil, Chile, and Colombia.

They found that during the latent period, the rate of GBS was 1.28 cases per day. Using this estimate

rounded to 1.3, find the probability on a given day of:

(a) No cases of GBS (b) At least one case of GBS

(c) Fewer than five cases of GBS

21. The IQs of individuals admitted to a state school for the mentally retarded are approximately normally

distributed with a mean of 60 and a standard deviation of 10.

(a) Find the proportion of individuals with IQs greater than 75.

(b) What is the probability that an individual picked at random will have an IQ between 55 and 75?

(c) Find

22. A nurse supervisor has found that staff nurses, on the average, complete a certain task in 10 minutes.

If the times required to complete the task are approximately normally distributed with a standard

deviation of 3 minutes, find:

(a) The proportion of nurses completing the task in less than 4 minutes

(b) The proportion of nurses requiring more than 5 minutes to complete the task

(c) The probability that a nurse who has just been assigned the task will complete it within 3

minutes

23. Scores made on a certain aptitude test by nursing students are approximately normally distributed

with a mean of 500 and a variance of 10,000.

(a) What proportion of those taking the test score below 200?

(b) A person is about to take the test. What is the probability that he or she will make a score of

650 or more?

(c) What proportion of scores fall between 350 and 675?

24. Given a binomial variable with a mean of 20 and a variance of 16, find n and p.

25. Suppose a variable X is normally distributed with a standard deviation of 10. Given that .0985 of

the values of X are greater than 70, what is the mean value of X?

26. Given the normally distributed random variable X, find the numerical value of k such that

27. Given the normally distributed random variable X with mean 100 and standard deviation 15, find

the numerical value of k such that:

(a)

(b)

(c)

(d) where k' and k are equidistant from

28. Given the normally distributed random variable X with and find

29. Given the normally distributed random variable X with and find

30. Given the normally distributed random variable X with and find

31. Given the normally distributed random variable X with and find

32. Given the normally distributed random variable X with and find

33. Explain why each of the following measurements is or is not the result of a Bernoulli trial:

(a) The gender of a newborn child

(b) The classification of a hospital patient’s condition as stable, critical, fair, good, or poor

(c) The weight in grams of a newborn child

34. Explain why each of the following measurements is or is not the result of a Bernoulli trial:

(a) The number of surgical procedures performed in a hospital in a week

(b) A hospital patient’s temperature in degrees Celsius

(c) A hospital patient’s vital signs recorded as normal or not normal

35. Explain why each of the following distributions is or is not a probability distribution:

cap5

EXERCISES

5.3.1 The National Health and Nutrition Examination Survey of 1988–1994 (NHANES III, A-1) estimated

the mean serum cholesterol level for U.S. females aged 20–74 years to be 204 mg/dl. The

estimate of the standard deviation was approximately 44. Using these estimates as the mean and

standard deviation for the U.S. population, consider the sampling distribution of the sample mean

based on samples of size 50 drawn from women in this age group. What is the mean of the sampling

distribution? The standard error?

5.3.2 The study cited in Exercise 5.3.1 reported an estimated mean serum cholesterol level of 183 for

women aged 20–29 years. The estimated standard deviation was approximately 37. Use these estimates

as the mean and standard deviation for the U.S. population. If a simple random sample

of size 60 is drawn from this population, find the probability that the sample mean serum cholesterol

level will be:

(a) Between 170 and 195 (b) Below 175

(c) Greater than 190

5.3.3 If the uric acid values in normal adult males are approximately normally distributed with a mean

and standard deviation of 5.7 and 1 mg percent, respectively, find the probability that a sample of

size 9 will yield a mean:

(a) Greater than 6 (b) Between 5 and 6

(c) Less than 5.2

5.3.4 Wright et al. (A-2) used the 1999–2000 National Health and Nutrition Examination Survey (NHANES)

to estimate dietary intake of 10 key nutrients. One of those nutrients was calcium (mg). They found

in all adults 60 years or older a mean daily calcium intake of 721 mg with a standard deviation of

454. Using these values for the mean and standard deviation for the U.S. population, find the probability

that a random sample of size 50 will have a mean:

(a) Greater than 800 mg (b) Less than 700 mg

(c) Between 700 and 850 mg

5.3.5 In the study cited in Exercise 5.3.4, researchers found the mean sodium intake in men and women

60 years or older to be 2940 mg with a standard deviation of 1476 mg. Use these values for the

mean and standard deviation of the U.S. population and find the probability that a random sample

of 75 people from the population will have a mean:

(a) Less than 2450 mg (b) Over 3100 mg

(c) Between 2500 and 3300 mg (d) Between 2500 and 2900 mg

m s

s

m

= .9818

= .9909 - .0091

= P1-2.36 … z … 2.362

P1115 … x … 1252 = Pa

115 - 120

2.12 … z …

125 - 120

2.12

b

15>150 = 2.12132

x

EXERCISES 145

5.3.6 Given a normally distributed population with a mean of 100 and a standard deviation of 20, find

the following probabilities based on a sample of size 16:

(a) (b)

(c)

5.3.7 Given and find:

(a) (b)

(c) (d)

5.3.8 Suppose a population consists of the following values: 1, 3, 5, 7, 9. Construct the sampling distribution

of based on samples of size 2 selected without replacement. Find the mean and variance

of the sampling distribution.

5.3.9 Use the data of Example 5.3.1 to construct the sampling distribution of based on samples of size 3

selected without replacement. Find the mean and variance of the sampling distribution.

5.3.10 Use the data cited in Exercise 5.3.1. Imagine we take samples of size 5, 25, 50, 100, and 500 from

the women in this age group.

(a) Calculate the standard error for each of these sampling scenarios.

(b) Discuss how sample size affects the standard error estimates calculated in part (a) and the

potential implications this may have in statistical practice.

EXERCISES

5.4.1 The study cited in Exercises 5.3.1 and 5.3.2 gives the following data on serum cholesterol levels

in U.S. females:

Use these estimates as the mean and standard deviation for the respective U.S. populations.

Suppose we select a simple random sample of size 50 independently from each population. What

is the probability that the difference between sample means will be more than 8?

5.4.2 In the study cited in Exercises 5.3.4 and 5.3.5, the calcium levels in men and women ages 60 years

or older are summarized in the following table:

Mean Standard Deviation

Men 797 482

Women 660 414

Use these estimates as the mean and standard deviation for the U.S. populations for these age

groups. If we take a random sample of 40 men and 35 women, what is the probability of obtaining

a difference between sample means of 100 mg or more?

5.4.3 Given two normally distributed populations with equal means and variances of and

what is the probability that samples of size and will yield a value of

greater than or equal to 8?

5.4.4 Given two normally distributed populations with equal means and variances of and

what is the probability that samples of size and will yield a value of

as large as or larger than 12?

5.4.5 For a population of 17-year-old boys and 17-year-old girls, the means and standard deviations,

respectively, of their subscapular skinfold thickness values are as follows: boys, 9.7 and 6.0; girls,

15.6 and 9.5. Simple random samples of 40 boys and 35 girls are selected from the populations.

What is the probability that the difference between sample means will be greater

than 10?

EXERCISES

5.5.1 Smith et al. (A-5) performed a retrospective analysis of data on 782 eligible patients admitted with

myocardial infarction to a 46-bed cardiac service facility. Of these patients, 248 (32 percent)

reported a past myocardial infarction. Use .32 as the population proportion. Suppose 50 subjects

are chosen at random from the population. What is the probability that over 40 percent would

report previous myocardial infarctions?

5.5.2 In the study cited in Exercise 5.5.1, 13 percent of the patients in the study reported previous

episodes of stroke or transient ischemic attack. Use 13 percent as the estimate of the prevalence

of stroke or transient ischemic attack within the population. If 70 subjects are chosen at random

from the population, what is the probability that 10 percent or less would report an incidence of

stroke or transient ischemic attack?

5.5.3 In the same 1999–2000 NHANES (A-3) report cited in Example 5.5.1, researchers estimated that

64 percent of U.S. adults ages 20–74 were overweight or obese (overweight: BMI 25-29, obese:

BMI 30 or greater). Use this estimate as the population proportion for U.S. adults ages 20–74. If

125 subjects are selected at random from the population, what is the probability that 70 percent

or more would be found to be overweight or obese?

5.5.4 Gallagher et al. (A-6) reported on a study to identify factors that influence women’s attendance at

cardiac rehabilitation programs. They found that by 12 weeks post-discharge, only 64 percent of

eligible women attended such programs. Using 64 percent as an estimate of the attendance percentage

of all eligible women, find the probability that in a sample of 45 women selected at random

from the population of eligible women less than 50 percent would attend programs.

5.5.5 Given a population in which and a random sample from this population of size 100, find:

(a) (b)

(c) P1.56 … pN … .632

P1p P1pN … .582

5.5.6 It is known that 35 percent of the members of a certain population suffer from one or more chronic

diseases. What is the probability that in a sample of 200 subjects drawn at random from this population

80 or more will have at least one chronic disease?

REVIEW QUESTIONS AND EXERCISES

1. What is a sampling distribution?

2. Explain how a sampling distribution may be constructed from a finite population.

3. Describe the sampling distribution of the sample mean when sampling is with replacement from

a normally distributed population.

4. Explain the central limit theorem.

5. How does the sampling distribution of the sample mean, when sampling is without replacement,

differ from the sampling distribution obtained when sampling is with replacement?

6. Describe the sampling distribution of the difference between two sample means.

7. Describe the sampling distribution of the sample proportion when large samples are drawn.

8. Describe the sampling distribution of the difference between two sample means when large samples

are drawn.

9. Explain the procedure you would follow in constructing the sampling distribution of the difference

between sample proportions based on large samples from finite populations.

10. Suppose it is known that the response time of healthy subjects to a particular stimulus is a normally

distributed random variable with a mean of 15 seconds and a variance of 16. What is the

probability that a random sample of 16 subjects will have a mean response time of 12 seconds

or more?

11. Janssen et al. (A-10) studied Americans ages 60 and over. They estimated the mean body mass index

of women over age 60 with normal skeletal muscle to be 23.1 with a standard deviation of 3.7. Using

these values as the population mean and standard deviation for women over age 60 with normal skeletal

muscle index, find the probability that 45 randomly selected women in this age range with normal

skeletal muscle index will have a mean BMI greater than 25.

12. In the study cited in Review Exercise 11, the researchers reported the mean BMI for men ages

60 and older with normal skeletal muscle index to be 24.7 with a standard deviation of 3.3.

Using these values as the population mean and standard deviation, find the probability that 50

randomly selected men in this age range with normal skeletal muscle index will have a mean

BMI less than 24.

13. Using the information in Review Exercises 11 and 12, find the probability that the difference in

mean BMI for 45 women and 50 men selected independently and at random from the respective

populations will exceed 3.

14. In the results published by Wright et al. (A-2) based on data from the 1999–2000 NHANES study

referred to in Exercises 5.4.1 and 5.4.2, investigators reported on their examination of iron levels. The

mean iron level for women ages 20–39 years was 13.7 mg with an estimated standard deviation of

8.9 mg. Using these as population values for women ages 20–39, find the probability that a random

sample of 100 women will have a mean iron level less than 12 mg.

15. Refer to Review Exercise 14. The mean iron level for men between the ages of 20 and 39 years

is 17.9 mg with an estimated standard deviation of 10.9 mg. Using 17.9 and 10.9 as population

parameters, find the probability that a random sample of 120 men will have a mean iron level

higher than 19 mg.

16. Using the information in Review Exercises 14 and 15, and assuming independent random samples

of size 100 and 120 for women and men, respectively, find the probability that the difference in

sample mean iron levels is greater than 5 mg.

17. The results of the 1999 National Health Interview Survey released in 2003 (A-7) showed that

among U.S. adults ages 60 and older, 19 percent had been told by a doctor or other health care

provider that they had some form of cancer. If we use this as the percentage for all adults 65 years

old and older living in the United States, what is the probability that among 65 adults chosen at

random more than 25 percent will have been told by their doctor or some other health care provider

that they have cancer?

18. Refer to Review Exercise 17. The reported cancer rate for women subjects ages 65 and older is 17

percent. Using this estimate as the true percentage of all females ages 65 and over who have been

told by a health care provider that they have cancer, find the probability that if 220 women are selected

at random from the population, more than 20 percent will have been told they have cancer.

19. Refer to Review Exercise 17. The cancer rate for men ages 65 and older is 23 percent. Use this

estimate as the percentage of all men ages 65 and older who have been told by a health care

provider that they have cancer. Find the probability that among 250 men selected at random that

fewer than 20 percent will have been told they have cancer.

20. Use the information in Review Exercises 18 and 19 to find the probability that the difference in

the cancer percentages between men and women will be less than 5 percent when 220 women and

250 men aged 65 and older are selected at random.

21. How many simple random samples (without replacement) of size 5 can be selected from a population

of size 10?

22. It is estimated by the 1999–2000 NHANES (A-7) that among adults 18 years old or older 53 percent

have never smoked. Assume the proportion of U.S. adults who have never smoked to be .53.

Consider the sampling distribution of the sample proportion based on simple random samples of

size 110 drawn from this population. What is the functional form of the sampling distribution?

23. Refer to Exercise 22. Compute the mean and variance of the sampling distribution.

24. Refer to Exercise 22. What is the probability that a single simple random sample of size 110 drawn

from this population will yield a sample proportion smaller than .50?

25. In a population of subjects who died from lung cancer following exposure to asbestos, it was found

that the mean number of years elapsing between exposure and death was 25. The standard deviation

was 7 years. Consider the sampling distribution of sample means based on samples of size

35 drawn from this population. What will be the shape of the sampling distribution?

26. Refer to Exercise 25. What will be the mean and variance of the sampling distribution?

27. Refer to Exercise 25. What is the probability that a single simple random sample of size 35 drawn

from this population will yield a mean between 22 and 29?

28. For each of the following populations of measurements, state whether the sampling distribution of

the sample mean is normally distributed, approximately normally distributed, or not approximately

normally distributed when computed from samples of size (A) 10, (B) 50, and (C) 200.

(a) The logarithm of metabolic ratios. The population is normally distributed.

(b) Resting vagal tone in healthy adults. The population is normally distributed.

(c) Insulin action in obese subjects. The population is not normally distributed.

29. For each of the following sampling situations indicate whether the sampling distribution of the

sample proportion can be approximated by a normal distribution and explain why or why not.

EXERCISES

For each of the following exercises construct 90, 95, and 99 percent confidence intervals for the

population mean, and state the practical and probabilistic interpretations of each. Indicate which

interpretation you think would be more appropriate to use when discussing confidence intervals with

someone who has not had a course in statistics, and state the reason for your choice. Explain why

the three intervals that you construct are not of equal width. Indicate which of the three intervals

you would prefer to use as an estimate of the population mean, and state the reason for your choice.

6.2.1 We wish to estimate the average number of heartbeats per minute for a certain population. The

average number of heartbeats per minute for a sample of 49 subjects was found to be 90. Assume

that these 49 patients constitute a random sample, and that the population is normally distributed

with a standard deviation of 10.

6.2.2 We wish to estimate the mean serum indirect bilirubin level of 4-day-old infants. The mean for a

sample of 16 infants was found to be 5.98 mg!100 cc. Assume that bilirubin levels in 4-day-old

infants are approximately normally distributed with a standard deviation of 3.5 mg!100 cc.

6.2.3 In a length of hospitalization study conducted by several cooperating hospitals, a random sample

of 64 peptic ulcer patients was drawn from a list of all peptic ulcer patients ever admitted to the

participating hospitals and the length of hospitalization per admission was determined for each.

The mean length of hospitalization was found to be 8.25 days. The population standard deviation

is known to be 3 days.

6.2.4 A sample of 100 apparently normal adult males, 25 years old, had a mean systolic blood pressure

of 125. It is believed that the population standard deviation is 15.

6.2.5 Some studies of Alzheimer’s disease (AD) have shown an increase in production in patients

with the disease. In one such study the following values were obtained from 16 neocortical

biopsy samples from AD patients.

1009 1280 1180 1255 1547 2352 1956 1080

1776 1767 1680 2050 1452 2857 3100 1621

Assume that the population of such values is normally distributed with a standard deviation of 350.

EXERCISES

6.3.1 Use the t distribution to find the reliability factor for a confidence interval based on the following

confidence coefficients and sample sizes:

a b c d

Confidence coefficient .95 .99 .90 .95

Sample size 15 24 8 30

6.3.2 In a study of the effects of early Alzheimer’s disease on nondeclarative memory, Reber et al. (A-2)

used the Category Fluency Test to establish baseline persistence and semantic memory and language

abilities. The eight subjects in the sample had Category Fluency Test scores of 11, 10, 6, 3, 11, 10,

9, 11. Assume that the eight subjects constitute a simple random sample from a normally distributed

population of similar subjects with early Alzheimer’s disease.

(a) What is the point estimate of the population mean?

(b) What is the standard deviation of the sample?

(c) What is the estimated standard error of the sample mean?

(d) Construct a 95 percent confidence interval for the population mean category fluency test score.

(e) What is the precision of the estimate?

(f) State the probabilistic interpretation of the confidence interval you constructed.

(g) State the practical interpretation of the confidence interval you constructed.

6.3.3 Pedroletti et al. (A-3) reported the maximal nitric oxide diffusion rate in a sample of 15 asthmatic

schoolchildren and 15 controls as mean ; standard error of the mean. For asthmatic children, they

reported (nanoliters per second) and for control subjects they reported

For each group, determine the following:

(a) What was the sample standard deviation?

(b) What is the 95 percent confidence interval for the mean maximal nitric oxide diffusion rate

of the population?

(c) What assumptions are necessary for the validity of the confidence interval you constructed?

(d) What are the practical and probabilistic interpretations of the interval you constructed?

(e) Which interpretation would be more appropriate to use when discussing confidence intervals

with someone who has not had a course in statistics? State the reasons for your choice.

(f) If you were to construct a 90 percent confidence interval for the population mean from the

information given here, would the interval be wider or narrower than the 95 percent confidence

interval? Explain your answer without actually constructing the interval.

(g) If you were to construct a 99 percent confidence interval for the population mean from the

information given here, would the interval be wider or narrower than the 95 percent confidence

interval? Explain your answer without actually constructing the interval.

6.3.4 The concern of a study by Beynnon et al. (A-4) were nine subjects with chronic anterior cruciate

ligament (ACL) tears. One of the variables of interest was the laxity of the anteroposterior,

where higher values indicate more knee instability. The researchers found that among

subjects with ACL-deficient knees, the mean laxity value was 17.4 mm with a standard deviation

of 4.3 mm.

(a) What is the estimated standard error of the mean?

(b) Construct the 99 percent confidence interval for the mean of the population from which the

nine subjects may be presumed to be a random sample.

(c) What is the precision of the estimate?

(d) What assumptions are necessary for the validity of the confidence interval you constructed?

6.3.5 A sample of 16 ten-year-old girls had a mean weight of 71.5 and a standard deviation of 12 pounds,

respectively. Assuming normality, find the 90, 95, and 99 percent confidence intervals for

6.3.6 The subjects of a study by Dugoff et al. (A-5) were 10 obstetrics and gynecology interns at the

University of Colorado Health Sciences Center. The researchers wanted to assess competence in

performing clinical breast examinations. One of the baseline measurements was the number of

such examinations performed. The following data give the number of breast examinations performed

for this sample of 10 interns.

Construct a 95 percent confidence interval for the mean of the population from which the study

subjects may be presumed to have been drawn.

EXERCISES

For each of the following exercises construct 90, 95, and 99 percent confidence intervals for the

difference between population means. Where appropriate, state the assumptions that make your

method valid. State the practical and probabilistic interpretations of each interval that you construct.

Consider the variables under consideration in each exercise, and state what use you think

researchers might make of your results.

6.4.1 Iannelo et al. (A-8) performed a study that examined free fatty acid concentrations in 18 lean subjects

and 11 obese subjects. The lean subjects had a mean level of 299 mEq/L with a standard

error of the mean of 30, while the obese subjects had a mean of 744 Eq/L with a standard error

of the mean of 62.

6.4.2 Chan et al. (A-9) developed a questionnaire to assess knowledge of prostate cancer. There was a

total of 36 questions to which respondents could answer “agree,” “disagree,” or “don’t know.”

Scores could range from 0 to 36. The mean scores for Caucasian study participants was 20.6 with

a standard deviation of 5.8, while the mean scores for African-American men was 17.4 with a

standard deviation of 5.8. The number of Caucasian study participants was 185, and the number

of African-Americans was 86.

6.4.3 The objectives of a study by van Vollenhoven et al. (A-10) were to examine the effectiveness of

etanercept alone and etanercept in combination with methotrexate in the treatment of rheumatoid

arthritis. The researchers conducted a retrospective study using data from the STURE database,

which collects efficacy and safety data for all patients starting biological treatments at the major

hospitals in Stockholm, Sweden. The researchers identified 40 subjects who were prescribed etanercept

only and 57 subjects who were given etanercept with methotrexate. Using a 100-mm visual

analogue scale (the higher the value, the greater the pain), researchers found that after 3 months

of treatment, the mean pain score was 36.4 with a standard error of the mean of 5.5 for subjects

taking etanercept only. In the sample receiving etanercept plus methotrexate, the mean score was

30.5 with a standard error of the mean of 4.6.

6.4.4 The purpose of a study by Nozawa et al. (A-11) was to determine the effectiveness of segmental

wire fixation in athletes with spondylolysis. Between 1993 and 2000, 20 athletes (6 women and

14 men) with lumbar spondylolysis were treated surgically with the technique. The following table

gives the Japanese Orthopaedic Association (JOA) evaluation score for lower back pain syndrome

for men and women prior to the surgery. The lower score indicates less pain.

Gender JOA scores

Female 14, 13, 24, 21, 20, 21

Male 21, 26, 24, 24, 22, 23, 18, 24, 13, 22, 25, 23, 21, 25

Source: Satoshi Nozawa, Katsuji Shimizu, Kei Miyamoto, and Mizuo

Tanaka, “Repair of Pars Interarticularis Defect by Segmental Wire Fixation

in Young Athletes with Spondylolysis,” American Journal of Sports

Medicine, 31 (2003), 359–364.

6.4.5 Krantz et al. (A-12) investigated dose-related effects of methadone in subjects with torsade de

pointes, a polymorphic ventricular tachycardia. In the study of 17 subjects, nine were being

treated with methadone for opiate dependency and eight for chronic pain. The mean daily dose

of methadone in the opiate dependency group was 541 mg/day with a standard deviation of

156, while the chronic pain group received a mean dose of 269 mg/day with a standard deviation

of 316.

6.4.6 Transverse diameter measurements on the hearts of adult males and females gave the following

results:

6.4.7 Twenty-four experimental animals with vitamin D deficiency were divided equally into two groups.

Group 1 received treatment consisting of a diet that provided vitamin D. The second group was

not treated. At the end of the experimental period, serum calcium determinations were made with

the following results:

Treated group:

Untreated group:

Assume normally distributed populations with equal variances.

6.4.8 Two groups of children were given visual acuity tests. Group 1 was composed of 11 children

who receive their health care from private physicians. The mean score for this group was 26

with a standard deviation of 5. Group 2 was composed of 14 children who receive their health

care from the health department, and had an average score of 21 with a standard deviation of

6. Assume normally distributed populations with equal variances.

6.4.9 The average length of stay of a sample of 20 patients discharged from a general hospital was

7 days with a standard deviation of 2 days. A sample of 24 patients discharged from a chronic

disease hospital had an average length of stay of 36 days with a standard deviation of 10 days.

Assume normally distributed populations with unequal variances.

6.4.10 In a study of factors thought to be responsible for the adverse effects of smoking on human reproduction,

cadmium level determinations (nanograms per gram) were made on placenta tissue of a

sample of 14 mothers who were smokers and an independent random sample of 18 nonsmoking

mothers. The results were as follows:

Nonsmokers: 10.0, 8.4, 12.8, 25.0, 11.8, 9.8, 12.5, 15.4, 23.5,

9.4, 25.1, 19.5, 25.5, 9.8, 7.5, 11.8, 12.2, 15.0

Smokers: 30.0, 30.1, 15.0, 24.1, 30.5, 17.8, 16.8, 14.8,

13.4, 28.5, 17.5, 14.4, 12.5, 20.4

Does it appear likely that the mean cadmium level is higher among smokers than nonsmokers?

Why do you reach this conclusion?

EXERCISES

For each of the following exercises state the practical and probabilistic interpretations of the interval

that you construct. Identify each component of the interval: point estimate, reliability coefficient,

and standard error. Explain why the reliability coefficients are not the same for all exercises.

6.5.1 Luna et al. (A-14) studied patients who were mechanically ventilated in the intensive care unit

of six hospitals in Buenos Aires, Argentina. The researchers found that of 472 mechanically

ventilated patients, 63 had clinical evidence of ventilator-associated pneumonia (VAP). Construct

a 95 percent confidence interval for the proportion of all mechanically ventilated patients at these

hospitals who may be expected to develop VAP.

6.5.2 Q waves on the electrocardiogram, according to Schinkel et al. (A-15), are often considered to be

reflective of irreversibly scarred myocardium. These researchers assert, however, that there are

some indications that residual viable tissue may be present in Q-wave-infarcted regions. Their study

of 150 patients with chronic electrocardiographic Q-wave infarction found 202 dysfunctional

Q-wave regions. With dobutamine stress echocardiography (DSE), they noted that 118 of these

202 regions were viable with information from the DSE testing. Construct a 90 percent confidence

interval for the proportion of viable regions that one might expect to find a population of dysfunctional

Q-wave regions.

6.5.3 In a study by von zur Muhlen et al. (A-16), 136 subjects with syncope or near syncope were studied.

Syncope is the temporary loss of consciousness due to a sudden decline in blood flow to the

brain. Of these subjects, 75 also reported having cardiovascular disease. Construct a 99 percent

confidence interval for the population proportion of subjects with syncope or near syncope who

also have cardiovascular disease.

6.5.4 In a simple random sample of 125 unemployed male high-school dropouts between the ages of 16

and 21, inclusive, 88 stated that they were regular consumers of alcoholic beverages. Construct a

95 percent confidence interval for the population proportion.

EXERCISES

For each of the following exercises state the practical and probabilistic interpretations of the interval

that you construct. Identify each component of the interval: point estimate, reliability coefficient,

and standard error. Explain why the reliability coefficients are not the same for all exercises.

6.6.1 Horwitz et al. (A-18) studied 637 persons who were identified by court records from 1967 to 1971

as having experienced abuse or neglect. For a control group, they located 510 subjects who as children

attended the same elementary school and lived within a five-block radius of those in the

abused / neglected group. In the abused/neglected group, and control group, 114 and 57 subjects,

respectively, had developed antisocial personality disorders over their lifetimes. Construct a 95 percent

confidence interval for the difference between the proportions of subjects developing antisocial

personality disorders one might expect to find in the populations of subjects from which the

subjects of this study may be presumed to have been drawn.

6.6.2 The objective of a randomized controlled trial by Adab et al. (A-19) was to determine whether providing

women with additional information on the pros and cons of screening for cervical cancer would

increase the willingness to be screened. A treatment group of 138 women received a leaflet on screening

that contained more information (average individual risk for cervical cancer, likelihood of positive

finding, the possibility of false positive/negative results, etc.) than the standard leaflet developed by

the British National Health Service that 136 women in a control group received. In the treatment group,

109 women indicated they wanted to have the screening test for cervical cancer while in the control

group, 120 indicated they wanted the screening test. Construct a 95 percent confidence interval for the

difference in proportions for the two populations represented by these samples.

6.6.3 Spertus et al. (A-20) performed a randomized single blind study for subjects with stable coronary

artery disease. They randomized subjects into two treatment groups. The first group had current

angina medications optimized, and the second group was tapered off existing medications and then

started on long-acting diltiazem at 180 mg/day. The researchers performed several tests to determine

if there were significant differences in the two treatment groups at baseline. One of the characteristics

of interest was the difference in the percentages of subjects who had reported a history

of congestive heart failure. In the group where current medications were optimized, 16 of 49 subjects

reported a history of congestive heart failure. In the subjects placed on the diltiazem, 12 of

the 51 subjects reported a history of congestive heart failure. State the assumptions that you think

are necessary and construct a 95 percent confidence interval for the difference between the proportions

of those reporting congestive heart failure within the two populations from which we presume

these treatment groups to have been selected.

6.6.4 To study the difference in drug therapy adherence among subjects with depression who received usual

care and those who received care in a collaborative care model was the goal of a study conducted

by Finley et al. (A-21). The collaborative care model emphasized the role of clinical pharmacists in

providing drug therapy management and treatment follow-up. Of the 50 subjects receiving usual care,

24 adhered to the prescribed drug regimen, while 50 out of 75 subjects in the collaborative care model

adhered to the drug regimen. Construct a 90 percent confidence interval for the difference in adherence

proportions for the populations of subjects represented by these two samples.

EXERCISES

6.7.1 A hospital administrator wishes to estimate the mean weight of babies born in her hospital. How

large a sample of birth records should be taken if she wants a 99 percent confidence interval that

is 1 pound wide? Assume that a reasonable estimate of is 1 pound. What sample size is required

if the confidence coefficient is lowered to .95?

6.7.2 The director of the rabies control section in a city health department wishes to draw a sample from

the department’s records of dog bites reported during the past year in order to estimate the mean

age of persons bitten. He wants a 95 percent confidence interval, he will be satisfied to let

and from previous studies he estimates the population standard deviation to be about 15 years.

How large a sample should be drawn?

6.7.3 A physician would like to know the mean fasting blood glucose value (milligrams per 100 ml) of

patients seen in a diabetes clinic over the past 10 years. Determine the number of records the

physician should examine in order to obtain a 90 percent confidence interval for if the desired

width of the interval is 6 units and a pilot sample yields a variance of 60.

6.7.4 For multiple sclerosis patients we wish to estimate the mean age at which the disease was first

diagnosed. We want a 95 percent confidence interval that is 10 years wide. If the population variance

is 90, how large should our sample be?

EXERCISES

6.8.1 An epidemiologist wishes to know what proportion of adults living in a large metropolitan area

have subtype ayr hepatitis B virus. Determine the sample size that would be required to estimate

the true proportion to within .03 with 95 percent confidence. In a similar metropolitan area the

proportion of adults with the characteristic is reported to be .20. If data from another metropolitan

area were not available and a pilot sample could not be drawn, what sample size would be

required?

6.8.2 A survey is planned to determine what proportion of the high-school students in a metropolitan

school system have regularly smoked marijuana. If no estimate of p is available from previous

studies, a pilot sample cannot be drawn, a confidence coefficient of .95 is desired, and is

to be used, determine the appropriate sample size. What sample size would be required if 99 percent

confidence were desired?

6.8.3 A hospital administrator wishes to know what proportion of discharged patients is unhappy with

the care received during hospitalization. How large a sample should be drawn if we let

the confidence coefficient is .95, and no other information is available? How large should the sample

be if p is approximated by .25?

6.8.4 A health planning agency wishes to know, for a certain geographic region, what proportion of

patients admitted to hospitals for the treatment of trauma die in the hospital. A 95 percent confidence

interval is desired, the width of the interval must be .06, and the population proportion, from

other evidence, is estimated to be .20. How large a sample is needed?

EXERCISES

6.9.1 A study by Aizenberg et al. (A-23) examined the efficacy of sildenafil, a potent phosphodiesterase

inhibitor, in the treatment of elderly men with erectile dysfunction induced by antidepressant treatment

for major depressive disorder. The ages of the 10 enrollees in the study were

74, 81, 70, 70, 74, 77, 76, 70, 71, 72

Assume that the subjects in this sample constitute a simple random sample drawn from a population

of similar subjects. Construct a 95 percent confidence interval for the variance of the ages of

subjects in the population.

6.9.2 Borden et al. (A-24) performed experiments on cadaveric knees to test the effectiveness of several

meniscal repair techniques. Specimens were loaded into a servohydraulic device and tension-loaded

to failure. The biomechanical testing was performed by using a slow loading rate to simulate the

stresses that the medial meniscus might be subjected to during early rehabilitation exercises and

activities of daily living. One of the measures is the amount of displacement that occurs. Of the

12 specimens receiving the vertical mattress suture and the FasT-FIX method, the displacement

values measured in millimeters are 16.9, 20.2, 20.1, 15.7, 13.9, 14.9, 18.0, 18.5, 9.2, 18.8, 22.8,

17.5. Construct a 90 percent confidence interval for the variance of the displacement in millimeters

for a population of subjects receiving these repair techniques.

6.9.3 Forced vital capacity determinations were made on 20 healthy adult males. The sample variance

was 1,000,000. Construct 90 percent confidence intervals for and

6.9.4 In a study of myocardial transit times, appearance transit times were obtained on a sample of

30 patients with coronary artery disease. The sample variance was found to be 1.03. Construct

99 percent confidence intervals for and

6.9.5 A sample of 25 physically and mentally healthy males participated in a sleep experiment in which

the percentage of each participant’s total sleeping time spent in a certain stage of sleep was

recorded. The variance computed from the sample data was 2.25. Construct 95 percent confidence

intervals for and

6.9.6 Hemoglobin determinations were made on 16 animals exposed to a harmful chemical. The following

observations were recorded: 15.6, 14.8, 14.4, 16.6, 13.8, 14.0, 17.3, 17.4, 18.6, 16.2, 14.7, 15.7,

16.4, 13.9, 14.8, 17.5. Construct 95 percent confidence intervals for and

6.9.7 Twenty air samples taken at the same site over a period of 6 months showed the following amounts

of suspended particulate matter (micrograms per cubic meter of air):

68 22 36 32

42 24 28 38

30 44 28 27

28 43 45 50

79 74 57 21

Consider these measurements to be a random sample from a population of normally distributed

measurements, and construct a 95 percent confidence interval for the population variance.

EXERCISES

6.10.1 The purpose of a study by Moneim et al. (A-26) was to examine thumb amputations from team

roping at rodeos. The researchers reviewed 16 cases of thumb amputations. Of these, 11 were complete

amputations while five were incomplete. The ischemia time is the length of time that insufficient

oxygen is supplied to the amputated thumb. The ischemia times (hours) for 11 subjects

experiencing complete amputations were

4.67, 10.5, 2.0, 3.18, 4.00, 3.5, 3.33, 5.32, 2.0, 4.25, 6.0

For five victims of incomplete thumb amputation, the ischemia times were

3.0, 10.25, 1.5, 5.22, 5.0

Treat the two reported sets of data as sample data from the two populations as described.

Construct a 95 percent confidence interval for the ratio of the two unknown population

variances.

6.10.2 The objective of a study by Horesh et al. (A-27) was to explore the hypothesis that some forms

of suicidal behavior among adolescents are related to anger and impulsivity. The sample consisted

of 65 adolescents admitted to a university-affiliated adolescent psychiatric unit. The researchers

used the Impulsiveness-Control Scale (ICS, A-28) where higher numbers indicate higher degrees

of impulsiveness and scores can range from 0 to 45. The 33 subjects classified as suicidal had an

ICS score standard deviation of 8.4 while the 32 nonsuicidal subjects had a standard deviation of

6.0. Assume that these two groups constitute independent simple random samples from two

populations of similar subjects. Assume also that the ICS scores in these two populations are normally

distributed. Find the 99 percent confidence interval for the ratio of the two population variances

of scores on the ICS.

6.10.3 Stroke index values were statistically analyzed for two samples of patients suffering from

myocardial infarction. The sample variances were 12 and 10. There were 21 patients in

each sample. Construct the 95 percent confidence interval for the ratio of the two population

variances.

6.10.4 Thirty-two adult asphasics seeking speech therapy were divided equally into two groups. Group 1

received treatment 1, and group 2 received treatment 2. Statistical analysis of the treatment effectiveness

scores yielded the following variances: Construct the 90 percent confidence

interval for

6.10.5 Sample variances were computed for the tidal volumes (milliliters) of two groups of patients suffering

from atrial septal defect. The results and sample sizes were as follows:

Construct the 95 percent confidence interval for the ratio of the two population variances.

6.10.6 Glucose responses to oral glucose were recorded for 11 patients with Huntington’s disease (group 1)

and 13 control subjects (group 2). Statistical analysis of the results yielded the following sample

variances: Construct the 95 percent confidence interval for the ratio of the

two population variances.

6.10.7 Measurements of gastric secretion of hydrochloric acid (milliequivalents per hour) in 16 normal

subjects and 10 subjects with duodenal ulcer yielded the following results:

Normal subjects: 6.3, 2.0, 2.3, 0.5, 1.9, 3.2, 4.1, 4.0, 6.2, 6.1,

3.5, 1.3, 1.7, 4.5, 6.3, 6.2

Ulcer subjects: 13.7, 20.6, 15.9, 28.4, 29.4, 18.4, 21.1, 3.0,

26.2, 13.0

Construct a 95 percent confidence interval for the ratio of the two population variances. What

assumptions must be met for this procedure to be valid?

REVIEW QUESTIONS AND EXERCISES

1. What is statistical inference?

2. Why is estimation an important type of inference?

3. What is a point estimate?

4. Explain the meaning of unbiasedness.

5. Define the following:

(a) Reliability coefficient (c) Precision (e) Estimator

(b) Confidence coefficient (d) Standard error (f) Margin of error

6. Give the general formula for a confidence interval.

7. State the probabilistic and practical interpretations of a confidence interval.

8. Of what use is the central limit theorem in estimation?

9. Describe the t distribution.

10. What are the assumptions underlying the use of the t distribution in estimating a single population

mean?

11. What is the finite population correction? When can it be ignored?

12. What are the assumptions underlying the use of the t distribution in estimating the difference

between two population means?

13. Arterial blood gas analyses performed on a sample of 15 physically active adult males yielded the

following resting values:

75, 80, 80, 74, 84, 78, 89, 72, 83, 76, 75, 87, 78, 79, 88

Compute the 95 percent confidence interval for the mean of the population.

14. What proportion of asthma patients are allergic to house dust? In a sample of 140, 35 percent

had positive skin reactions. Construct the 95 percent confidence interval for the population

proportion.

15. An industrial hygiene survey was conducted in a large metropolitan area. Of 70 manufacturing

plants of a certain type visited, 21 received a “poor” rating with respect to absence of safety hazards.

Construct a 95 percent confidence interval for the population proportion deserving a “poor”

rating.

16. Refer to the previous problem. How large a sample would be required to estimate the population

proportion to within .05 with 95 percent confidence (.30 is the best available estimate of p):

(a) If the finite population correction can be ignored?

(b) If the finite population correction is not ignored and N 1500?

17. In a dental survey conducted by a county dental health team, 500 adults were asked to give the

reason for their last visit to a dentist. Of the 220 who had less than a high-school education, 44

said they went for preventative reasons. Of the remaining 280, who had a high-school education

or better, 150 stated that they went for preventative reasons. Construct a 95 percent confidence

interval for the difference between the two population proportions.

18. A breast cancer research team collected the following data on tumor size:

Type of Tumor n s

A 21 3.85 cm 1.95 cm

B 16 2.80 cm 1.70 cm

Construct a 95 percent confidence interval for the difference between population means.

19. A certain drug was found to be effective in the treatment of pulmonary disease in 180 of 200

cases treated. Construct the 90 percent confidence interval for the population proportion.

20. Seventy patients with stasis ulcers of the leg were randomly divided into two equal groups. Each

group received a different treatment for edema. At the end of the experiment, treatment effectiveness

was measured in terms of reduction in leg volume as determined by water displacement. The

means and standard deviations for the two groups were as follows:

Group (Treatment) s

A 95 cc 25

B 125 cc 30

Construct a 95 percent confidence interval for the difference in population means.

21. What is the average serum bilirubin level of patients admitted to a hospital for treatment of hepatitis?

A sample of 10 patients yielded the following results:

20.5, 14.8, 21.3, 12.7, 15.2, 26.6, 23.4, 22.9, 15.7, 19.2

Construct a 95 percent confidence interval for the population mean.

22. Determinations of saliva pH levels were made in two independent random samples of seventhgrade

schoolchildren. Sample A children were caries-free while sample B children had a high

incidence of caries. The results were as follows:

A: 7.14, 7.11, 7.61, 7.98, 7.21, 7.16, 7.89 B: 7.36, 7.04, 7.19, 7.41, 7.10, 7.15, 7.36,

7.24, 7.86, 7.47, 7.82, 7.37, 7.66, 7.62, 7.65 7.57, 7.64, 7.00, 7.25, 7.19

Construct a 90 percent confidence interval for the difference between the population means.

Assume that the population variances are equal.

23. Drug A was prescribed for a random sample of 12 patients complaining of insomnia. An independent

random sample of 16 patients with the same complaint received drug B. The number of hours of

sleep experienced during the second night after treatment began were as follows:

A: 3.5, 5.7, 3.4, 6.9, 17.8, 3.8, 3.0, 6.4, 6.8, 3.6, 6.9, 5.7

B: 4.5, 11.7, 10.8, 4.5, 6.3, 3.8, 6.2, 6.6, 7.1, 6.4, 4.5,

5.1, 3.2, 4.7, 4.5, 3.0

Construct a 95 percent confidence interval for the difference between the population means.

Assume that the population variances are equal.

24. The objective of a study by Crane et al. (A-29) was to examine the efficacy, safety, and maternal

satisfaction of (a) oral misoprostol and (b) intravenous oxytocin for labor induction in women with

premature rupture of membranes at term. Researchers randomly assigned women to the two treatments.

For the 52 women who received oral misoprostol, the mean time in minutes to active labor

was 358 minutes with a standard deviation of 308 minutes. For the 53 women taking oxytocin,

the mean time was 483 minutes with a standard deviation of 144 minutes. Construct a 99 percent

confidence interval for the difference in mean time to active labor for these two different medications.

What assumptions must be made about the reported data? Describe the population about

which an inference can be made.

25. Over a 2-year period, 34 European women with previous gestational diabetes were retrospectively

recruited from West London antenatal databases for a study conducted by Kousta et al.

(A-30). One of the measurements for these women was the fasting nonesterified fatty acids concentration

(NEFA) measured in In the sample of 34 women, the mean NEFA level was

435 with a sample standard deviation of 215.0. Construct a 95 percent confidence interval for

the mean fasting NEFA level for a population of women with gestational diabetes. State all

necessary assumptions about the reported data and subjects.

26. Scheid et al. (A-31) questioned 387 women receiving free bone mineral density screening. The

questions focused on past smoking history. Subjects undergoing hormone replacement therapy

(HRT), and subjects not undergoing HRT, were asked if they had ever been a regular smoker. In

the HRT group, 29.3 percent of 220 women stated that they were at some point in their life a regular

smoker. In the non–HRT group, 17.3 percent of 106 women responded positively to being at

some point in their life a regular smoker. (Sixty-one women chose not to answer the question.)

Construct a 95 percent confidence interval for the difference in smoking percentages for the two

populations of women represented by the subjects in the study. What assumptions about the data

are necessary?

27. The purpose of a study by Elliott et al. (A-32) was to assess the prevalence of vitamin D deficiency

in women living in nursing homes. The sample consisted of 39 women in a 120-bed skilled

nursing facility. Women older than 65 years of age who were long-term residents were invited

to participate if they had no diagnosis of terminal cancer or metastatic disease. In the sample,

23 women had 25-hydroxyvitamin D levels of 20 ng/ml or less. Construct a 95 percent confidence

interval for the percent of women with vitamin D deficiency in the population presumed to be represented

by this sample.

28. In a study of the role of dietary fats in the etiology of ischemic heart disease the subjects were

60 males between 40 and 60 years of age who had recently had a myocardial infarction and 50

apparently healthy males from the same age group and social class. One variable of interest in

the study was the proportion of linoleic acid (L.A.) in the subjects’ plasma triglyceride fatty acids.

The data on this variable were as follows:

Construct the 95 percent confidence interval for the difference between population means. What do

these data suggest about the levels of linoleic acid in the two sampled populations?

29. The purpose of a study by Tahmassebi and Curzon (A-33) was to compare the mean salivary flow

rate among subjects with cerebral palsy and among subjects in a control group. Each group had 10

subjects. The following table gives the mean flow rate in ml/ minute as well as the standard error.

Group Sample Size Mean ml/minute Standard Error

Cerebral palsy 10 0.220 0.0582

Control 10 0.334 0.1641

Source: J. F. Tahmassebi and M. E. J. Curzon, “The Cause of Drooling in Children

with Cerebral Palsy—Hypersalivation or Swallowing Defect?” International Journal of

Paediatric Dentistry, 13 (2003), 106–111.

Construct the 90 percent confidence interval for the difference in mean salivary flow rate for the

two populations of subjects represented by the sample data. State the assumptions necessary for

this to be a valid confidence interval.

30. Culligan et al. (A-34) compared the long-term results of two treatments: (a) a modified Burch procedure,

and (b) a sling procedure for stress incontinence with a low-pressure urethra. Thirty-six

women took part in the study with 19 in the Burch treatment group and 17 in the sling procedure

treatment group. One of the outcome measures at three months post-surgery was maximum urethral

closure pressure (cm In the Burch group the mean and standard deviation were 16.4

and 8.2 cm, respectively. In the sling group, the mean and standard deviation were 39.8 and 23.0,

respectively. Construct the 99 percent confidence interval for the difference in mean maximum urethral

closure pressure for the two populations represented by these subjects. State all necessary

assumptions.

31. In general, narrow confidence intervals are preferred over wide ones. We can make an interval narrow

by using a small confidence coefficient. For a given set of other conditions, what happens to

the level of confidence when we use a small confidence coefficient? What would happen to the

interval width and the level of confidence if we were to use a confidence coefficient of zero?

32. In general, a high level of confidence is preferred over a low level of confidence. For a given set

of other conditions, suppose we set our level of confidence at 100 percent. What would be the

effect of such a choice on the width of the interval?

33. The subjects of a study by Borland et al. (A-35) were children in acute pain. Thirty-two children who

presented at an emergency room were enrolled in the study. Each child used the visual analogue scale

to rate pain on a scale from 0 to 100 mm. The mean pain score was 61.3 mm with a 95 percent confidence

interval of 53.2 mm–69.4 mm. Which would be the appropriate reliability factor for the interval,

z or t? Justify your choice. What is the precision of the estimate? The margin of error?

34. Does delirium increase hospital stay? That was the research question investigated by McCusker

et al. (A-36). The researchers sampled 204 patients with prevalent delirium and 118 without delirium.

The conclusion of the study was that patients with prevalent delirium did not have a higher

mean length of stay compared to those without delirium. What was the target population? The

sampled population?

35. Assessing driving self-restriction in relation to vision performance was the objective of a study

by West et al. (A-37). The researchers studied 629 current drivers ages 55 and older for 2 years.

The variables of interest were driving behavior, health, physical function, and vision function.

The subjects were part of a larger vision study at the Smith-Kettlewell Eye Research Institute.

A conclusion of the study was that older adults with early changes in spatial vision function

and depth perception appear to recognize their limitations and restrict their driving. What was

the target population? The sampled population?

36. In a pilot study conducted by Ayouba et al. (A-38), researchers studied 123 children born of HIV-1-

infected mothers in Yaound , Cameroon. Counseled and consenting pregnant women were given a

single dose of nevirapine at the onset of labor. Babies were given a syrup containing nevirapine within

the first 72 hours of life. The researchers found that 87 percent of the children were considered not

infected at 6–8 weeks of age. What is the target population? What is the sampled population?

37. Refer to Exercise 2.3.11. Construct a 95 percent confidence interval for the population mean S/R

ratio. Should you use t or z as the reliability coefficient? Why? Describe the population about

which inferences based on this study may be made.

38. Refer to Exercise 2.3.12. Construct a 90 percent confidence interval for the population mean height.

Should you use t or z as the reliability coefficient? Why? Describe the population about which

inferences based on this study may be made.

Exercises for Use with Large Data Sets Available on the Following Website:

www.wiley.com/college/daniel

1. Refer to North Carolina Birth Registry Data NCBIRTH800 with 800 observations (see Large Data

Exercise 1 in Chapter 2). Calculate 95 percent confidence intervals for the following:

(a) the percentage of male children

(b) the mean age of a mother giving birth

(c) the mean weight gained during pregnancy

(d) the percentage of mothers admitting to smoking during pregnancy

(e) the difference in the average weight gained between smoking and nonsmoking mothers

(f) the difference in the average birth weight in grams between married and nonmarried mothers

(g) the difference in the percentage of low birth weight babies between married and nonmarried

mothers

2. Refer to the serum cholesterol levels for 1000 subjects (CHOLEST). Select a simple random sample

of size 15 from this population and construct a 95 percent confidence interval for the population

mean. Compare your results with those of your classmates. What assumptions are necessary

for your estimation procedure to be valid?

3. Refer to the serum cholesterol levels for 1000 subjects (CHOLEST). Select a simple random sample

of size 50 from the population and construct a 95 percent confidence interval for the proportion

of subjects in the population who have readings greater than 225. Compare your results with those

of your classmates.

4. Refer to the weights of 1200 babies born in a community hospital (BABY WGTS). Draw a simple

random sample of size 20 from this population and construct a 95 percent confidence interval

for the population mean. Compare your results with those of your classmates. What assumptions

are necessary for your estimation procedure to be valid?

5. Refer to the weights of 1200 babies born in a community hospital (BABY WGTS). Draw a simple random

sample of size 35 from the population and construct a 95 percent confidence interval for the population

mean. Compare this interval with the one constructed in Exercise 4.

6. Refer to the heights of 1000 twelve-year-old boys (BOYHGTS). Select a simple random sample

of size 15 from this population and construct a 99 percent confidence interval for the population

mean. What assumptions are necessary for this procedure to be valid?

7. Refer to the heights of 1000 twelve-year-old boys (BOYHGTS). Select a simple random sample

of size 35 from the population and construct a 99 percent confidence interval for the population

mean. Compare this interval with the one constructed in Exercise 5.

c7