CHAPTER

5

SOME IMPORTANT

SAMPLING DISTRIBUTIONS

CHAPTER OVERVIEW

This chapter ties together the foundations of applied statistics: descriptive measures, basic probability, and inferential procedures. This chapter also includes a discussion of one of the most important theorems in statistics, the central limit theorem. Students may find it helpful to revisit this chapter from time to time as they study the remaining chapters of the book.

TO P I C S

5.1INTRODUCTION

5.2SAMPLING DISTRIBUTIONS

5.3DISTRIBUTION OF THE SAMPLE MEAN

5.4DISTRIBUTION OF THE DIFFERENCE BETWEEN TWO SAMPLE MEANS

5.5DISTRIBUTION OF THE SAMPLE PROPORTION

5.6DISTRIBUTION OF THE DIFFERENCE BETWEEN TWO SAMPLE PROPORTIONS

5.7SUMMARY

LEARNING OUTCOMES

After studying this chapter, the student will

1. be able to construct a sampling distribution of a statistic.

2. understand how to use a sampling distribution to calculate basic probabilities.

3. understand the central limit theorem and when to apply it.

4. understand the basic concepts of sampling with replacement and without

replacement.

5.1 INTRODUCTION

Before we examine the subject matter of this chapter, let us review the high points of what we have covered thus far. Chapter 1 introduces some basic and useful statistical vocabulary and discusses the basic concepts of data collection. In Chapter 2, the organization and summarization of data are emphasized. It is here that we encounter the concepts of central tendency and dispersion and learn how to compute their descriptive measures. In Chapter 3, we are introduced to the fundamental ideas of probability, and in Chapter 4 we consider the

concept of a probability distribution. These concepts are fundamental to an understanding of statistical inference, the topic that comprises the major portion of this book. This chapter serves as a bridge between the preceding material, which is essentially descriptive in nature, and most of the remaining topics, which have been selected from the area of statistical inference.

5.2 SAMPLING DISTRIBUTIONS

The topic of this chapter is sampling distributions. The importance of a clear understanding of sampling distributions cannot be overemphasized, as this concept is the very key to the understanding of statistical inference. Sampling distributions serve two purposes: (1) they allow us to answer probability questions about sample statistics, and (2) they provide the necessary theory for making statistical inference procedures valid. In this chapter we use sampling distributions to answer probability questions about sample statistics. We recall from Chapter 2 that a sample statistic is a descriptive measure, such as the mean, median, variance, or standard deviation, that is computed from the data of a sample. In the chapters that follow, we will see how sampling distributions make statistical inferences valid.

We begin with the following definition.

DEFINITION

The distribution of all possible values that can be assumed by some statistic, computed from samples of the same size randomly drawn from the same population, is called the sampling distribution of that statistic.

Sampling Distributions: Construction Sampling distributions may be constructed empirically when sampling from a discrete, finite population. To construct a sampling distribution we proceed as follows:

1. From a finite population of size N, randomly draw all possible samples of size n.

2. Compute the statistic of interest for each sample.

3. List in one column the different distinct observed values of the statistic, and in another column list the corresponding frequency of occurrence of each distinct observed value of the statistic.

The actual construction of a sampling distribution is a formidable undertaking if the population is of any appreciable size and is an impossible task if the population is infinite. In such cases, sampling distributions may be approximated by taking a large number of samples of a given size.

Sampling Distributions: Important Characteristics We usually

are interested in knowing three things about a given sampling distribution: its mean, its

variance, and its functional form (how it looks when graphed).

We can recognize the difficulty of constructing a sampling distribution according

to the steps given above when the population is large. We also run into a problem when

considering the construction of a sampling distribution when the population is infinite.

The best we can do experimentally in this case is to approximate the sampling distribu-

tion of a statistic.

Both these problems may be obviated by means of mathematics. Although the pro-

cedures involved are not compatible with the mathematical level of this text, sampling

distributions can be derived mathematically. The interested reader can consult one of

many mathematical statistics textbooks, for example, Larsen and Marx (1) or Rice (2).

In the sections that follow, some of the more frequently encountered sampling

distributions are discussed.

5.3 DISTRIBUTION OF THE SAMPLE MEAN

An important sampling distribution is the distribution of the sample mean. Let us see

how we might construct the sampling distribution by following the steps outlined in the

previous section.

EXAMPLE 5.3.1

Suppose we have a population of size N = 5, consisting of the ages of five children

who are outpatients in a community mental health center. The ages are as follows:

x 1 = 6, x 2 = 8, x 3 = 10, x 4 = 12, and x 5 = 14. The mean, m, of this population is

equal to gx i >N = 10 and the variance is

s2 =

g1x i - m22

40

=

= 8

N

5

Let us compute another measure of dispersion and designate it by capital S as

follows:

S2 =

g1x i - m22

40

=

= 10

N - 1

4

We will refer to this quantity again in the next chapter. We wish to construct the sam-

pling distribution of the sample mean, x, based on samples of size n = 2 drawn from

this population.

Solution:

Let us draw all possible samples of size n = 2 from this population. These

samples, along with their means, are shown in Table 5.3.1.

We see in this example that, when sampling is with replacement, there

are 25 possible samples. In general, when sampling is with replacement, the

number of possible samples is equal to N n.

We may construct the sampling distribution of x by listing the differ-

ent values of x in one column and their frequency of occurrence in another,

as in Table 5.3.2.

■

We see that the data of Table 5.3.2 satisfy the requirements for a probability

distribution. The individual probabilities are all greater than 0, and their sum is equal

to 1.

It was stated earlier that we are usually interested in the functional form of a sam-

pling distribution, its mean, and its variance. We now consider these characteristics for

the sampling distribution of the sample mean, x.

Sampling Distribution of x: Functional Form Let us look at the dis-

tribution of x plotted as a histogram, along with the distribution of the population, both

of which are shown in Figure 5.3.1. We note the radical difference in appearance between

the histogram of the population and the histogram of the sampling distribution of x.

Whereas the former is uniformly distributed, the latter gradually rises to a peak and then

drops off with perfect symmetry.

Sampling Distribution of x: Mean Now let us compute the mean, which

we will call mx , of our sampling distribution. To do this we add the 25 sample means

and divide by 25. Thus,

mx =

6 + 7 + 7 + 8 + Á + 14

250

gx i

=

=

= 10

n

N

25

25

We note with interest that the mean of the sampling distribution of x has the same

value as the mean of the original population.

Sampling Distribution of x: Variance Finally, we may compute the vari-

ance of x, which we call s2x as follows:

sx2 =

=

g1x i - mx22

Nn

16 - 1022 + 17 - 1022 + 17 - 1022 + Á + 114 - 1022

25

100

=

= 4

25

We note that the variance of the sampling distribution is not equal to the population vari-

ance. It is of interest to observe, however, that the variance of the sampling distribution

is equal to the population variance divided by the size of the sample used to obtain the

sampling distribution. That is,

s2x =

s2

8

= = 4

n

2

The square root of the variance of the sampling distribution, 2s2x = s> 1n is called

the standard error of the mean or, simply, the standard error.

These results are not coincidences but are examples of the characteristics of sam-

pling distributions in general, when sampling is with replacement or when sampling is

from an infinite population. To generalize, we distinguish between two situations: sam-

pling from a normally distributed population and sampling from a nonnormally distrib-

uted population.

Sampling Distribution of x: Sampling from Normally Distrib-

uted Populations When sampling is from a normally distributed population, the

distribution of the sample mean will possess the following properties:

1. The distribution of x will be normal.

2. The mean, mx , of the distribution of x will be equal to the mean of the population

from which the samples were drawn.

3. The variance, s2x of the distribution of x will be equal to the variance of the pop-

ulation divided by the sample size.

Sampling from Nonnormally Distributed Populations For the case

where sampling is from a nonnormally distributed population, we refer to an important

mathematical theorem known as the central limit theorem. The importance of this theorem

in statistical inference may be summarized in the following statement.

The Central Limit Theorem

Given a population of any nonnormal functional form with a mean m and finite

variance s 2, the sampling distribution of x, computed from samples of size n from

this population, will have mean m and variance s2>n and will be approximately

A mathematical formulation of the central limit theorem is that the distribution of

x - m

s> 1n

approaches a normal distribution with mean 0 and variance 1 as n : q . Note that the

central limit theorem allows us to sample from nonnormally distributed populations with

a guarantee of approximately the same results as would be obtained if the populations

were normally distributed provided that we take a large sample.

The importance of this will become evident later when we learn that a normally

distributed sampling distribution is a powerful tool in statistical inference. In the case of

the sample mean, we are assured of at least an approximately normally distributed sam-

pling distribution under three conditions: (1) when sampling is from a normally distrib-

uted population; (2) when sampling is from a nonnormally distributed population and

our sample is large; and (3) when sampling is from a population whose functional form

is unknown to us as long as our sample size is large.

The logical question that arises at this point is, How large does the sample have

to be in order for the central limit theorem to apply? There is no one answer, since the

size of the sample needed depends on the extent of nonnormality present in the popula-

tion. One rule of thumb states that, in most practical situations, a sample of size 30 is

satisfactory. In general, the approximation to normality of the sampling distribution of x

becomes better and better as the sample size increases.

Sampling Without Replacement The foregoing results have been given on

the assumption that sampling is either with replacement or that the samples are drawn

from infinite populations. In general, we do not sample with replacement, and in most

practical situations it is necessary to sample from a finite population; hence, we need to

become familiar with the behavior of the sampling distribution of the sample mean under

these conditions. Before making any general statements, let us again look at the data in

Table 5.3.1. The sample means that result when sampling is without replacement are

those above the principal diagonal, which are the same as those below the principal diag-

onal, if we ignore the order in which the observations were drawn. We see that there are

10 possible samples. In general, when drawing samples of size n from a finite popula-

tion of size N without replacement, and ignoring the order in which the sample values

are drawn, the number of possible samples is given by the combination of N things taken

n at a time. In our present example we have

N Cn =

N!

5!

5 # 4 # 3!

=

=

= 10 possible samples.

n!1N - n2!

2! 3!

2! 3!

The mean of the 10 sample means is

mx =

7 + 8 + 9 + Á + 13

100

g xi

=

=

= 10

10

10

N Cn

We see that once again the mean of the sampling distribution is equal to the population

mean.

The variance of this sampling distribution is found to be

s2x =

g 1x i - mx22

30

=

= 3

C

10

N n

and we note that this time the variance of the sampling distribution is not equal to the

population variance divided by the sample size, since s2x = 3 Z 8>2 = 4. There

is, however, an interesting relationship that we discover by multiplying s2>n by

1N - n2>1N - 12. That is,

s2 # N - n

8 # 5 - 2

=

= 3

n

N - 1

2

4

This result tells us that if we multiply the variance of the sampling distribution that would

be obtained if sampling were with replacement, by the factor 1N - n2>1N - 12, we

obtain the value of the variance of the sampling distribution that results when sampling

is without replacement. We may generalize these results with the following statement.

When sampling is without replacement from a finite population, the sampling distribu-

tion of x will have mean m and variance

s2x =

s2 # N - n

n

N - 1

If the sample size is large, the central limit theorem applies and the sampling

distribution of x will be approximately normally distributed.

The Finite Population Correction The factor 1N - n2>1N - 12 is called

the finite population correction and can be ignored when the sample size is small in com-

parison with the population size. When the population is much larger than the sample,

the difference between s2>n and 1s2>n231N - n2>1N - 124 will be negligible. Imagine

a population of size 10,000 and a sample from this population of size 25; the finite pop-

ulation correction would be equal to 110,000 - 252>199992 = .9976. To multiply s2>n

by .9976 is almost equivalent to multiplying it by 1. Most practicing statisticians do not

use the finite population correction unless the sample is more than 5 percent of the size

of the population. That is, the finite population correction is usually ignored when

n>N … .05.

The Sampling Distribution of x: A Summary Let us summarize the

characteristics of the sampling distribution of x under two conditions.

1. Sampling is from a normally distributed population with a known population

variance:

(a) mx = m

(b) sx = s>1n

(c) The sampling distribution of x is normal.

2. Sampling is from a nonnormally distributed population with a known population

variance:

(a) mx = m

(b) sx = s> 1n,

when

n >N … .05

N - n

sx = 1s> 1n2

,

otherwise

AN - 1

(c) The sampling distribution of x is approximately normal.

Applications As we will see in succeeding chapters, knowledge and understand-

ing of sampling distributions will be necessary for understanding the concepts of statis-

tical inference. The simplest application of our knowledge of the sampling distribution

of the sample mean is in computing the probability of obtaining a sample with a mean

of some specified magnitude. Let us illustrate with some examples.

EXAMPLE 5.3.2

Suppose it is known that in a certain large human population cranial length is approx-

imately normally distributed with a mean of 185.6 mm and a standard deviation of

12.7 mm. What is the probability that a random sample of size 10 from this popula-

tion will have a mean greater than 190?

Solution:

We know that the single sample under consideration is one of all possible

samples of size 10 that can be drawn from the population, so that the mean

that it yields is one of the x’s constituting the sampling distribution of x

that, theoretically, could be derived from this population.

When we say that the population is approximately normally distrib-

uted, we assume that the sampling distribution of x will be, for all prac-

tical purposes, normally distributed. We also know that the mean and

standard deviation of the sampling distribution are equal to 185.6 and

2112.722>10 = 12.7> 210 = 4.0161, respectively. We assume that the

population is large relative to the sample so that the finite population cor-

rection can be ignored.

We learn in Chapter 4 that whenever we have a random variable that is

normally distributed, we may very easily transform it to the standard normal

distribution. Our random variable now is x, the mean of its distribution is mx,

and its standard deviation is sx = s> 1n. By appropriately modifying the

formula given previously, we arrive at the following formula for transform-

ing the normal distribution of x to the standard normal distribution:

z =

x - mx

s> 1n

(5.3.1)

■

The probability that answers our question is represented by the area to the right of x = 190

under the curve of the sampling distribution. This area is equal to the area to the right of

z =

190 - 185.6

4.4

=

= 1.10

4.0161

4.0161

By consulting the standard normal table, we find that the area to the right of 1.10 is

.1357; hence, we say that the probability is .1357 that a sample of size 10 will have a

mean greater than 190.

Figure 5.3.2 shows the relationship between the original population, the sampling dis-

tribution of x and the standard normal distribution.

EXAMPLE 5.3.3

If the mean and standard deviation of serum iron values for healthy men are 120 and

15 micrograms per 100 ml, respectively, what is the probability that a random sample

of 50 normal men will yield a mean between 115 and 125 micrograms per 100 ml?

Solution:

The functional form of the population of serum iron values is not speci-

fied, but since we have a sample size greater than 30, we make use of the

central limit theorem and transform the resulting approximately normal

sampling distribution of x (which has a mean of 120 and a standard devi-

ation of 15> 150 = 2.12132 to the standard normal. The probability we

seek is

115 - 120

125 - 120

… z …

b

2.12

2.12

= P1-2.36 … z … 2.362

= .9909 - .0091

= .9818

P1115 … x … 1252 = Pa

■

EXERCISES

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

mated 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 m and

standard deviation s 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 sam-

pling distribution? The standard error?

5.3.2The 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 esti-

mates as the mean m and standard deviation s 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 choles-

terol level will be:

(a) Between 170 and 195

(b) Below 175

(c) Greater than 190

5.3.3If 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

5.3.4Wright 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 proba-

bility 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

5.3.5In 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 sam-

ple 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

5.3.6Given 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) P1x Ú 1002

(b) P1x … 1102

(c) P196 … x … 1082

5.3.7Given m = 50, s = 16, and n = 64, find:

(a) P145 … x … 552

(b) P1x 7 532

(c) P1x 6 472

(d) P149 … x … 562

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

tribution of x based on samples of size 2 selected without replacement. Find the mean and vari-

ance of the sampling distribution.

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

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

5.3.10Use 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.

5.4 DISTRIBUTION OF THE DIFFERENCE

BETWEEN TWO SAMPLE MEANS

Frequently the interest in an investigation is focused on two populations. Specifically, an

investigator may wish to know something about the difference between two population

means. In one investigation, for example, a researcher may wish to know if it is reason-

able to conclude that two population means are different. In another situation, the

researcher may desire knowledge about the magnitude of the difference between two

population means. A medical research team, for example, may want to know whether or

not the mean serum cholesterol level is higher in a population of sedentary office work-

ers than in a population of laborers. If the researchers are able to conclude that the pop-

ulation means are different, they may wish to know by how much they differ. A knowl-

edge of the sampling distribution of the difference between two means is useful in

investigations of this type.

Sampling from Normally Distributed Populations The following

example illustrates the construction of and the characteristics of the sampling distribu-

tion of the difference between sample means when sampling is from two normally dis-

tributed populations.

EXAMPLE 5.4.1

Suppose we have two populations of individuals—one population (population 1) has

experienced some condition thought to be associated with mental retardation, and the

other population (population 2) has not experienced the condition. The distribution of

intelligence scores in each of the two populations is believed to be approximately nor-

mally distributed with a standard deviation of 20.

Suppose, further, that we take a sample of 15 individuals from each population and

compute for each sample the mean intelligence score with the following results: x 1 = 92

and x 2 = 105. If there is no difference between the two populations, with respect to their

true mean intelligence scores, what is the probability of observing a difference this large

or larger 1x 1 - x 22 between sample means?

Solution: To answer this question we need to know the nature of the sampling distri-

bution of the relevant statistic, the difference between two sample means,

x 1 - x 2. Notice that we seek a probability associated with the difference

between two sample means rather than a single mean.

■

Sampling Distribution of x 1 - x 2: Construction Although, in prac-

tice, we would not attempt to construct the desired sampling distribution, we can concep-

tualize the manner in which it could be done when sampling is from finite populations.

We would begin by selecting from population 1 all possible samples of size 15 and com-

puting the mean for each sample. We know that there would be N1Cn1 such samples where

N1 is the population size and n 1 = 15. Similarly, we would select all possible samples of

size 15 from population 2 and compute the mean for each of these samples. We would

then take all possible pairs of sample means, one from population 1 and one from popu-

lation 2, and take the difference. Table 5.4.1 shows the results of following this procedure.

Note that the 1’s and 2’s in the last line of this table are not exponents, but indicators of

population 1 and 2, respectively.

Sampling Distribution of x 1 - x 2: Characteristics It is the distri-

bution of the differences between sample means that we seek. If we plotted the sample

differences against their frequency of occurrence, we would obtain a normal distribution

with a mean equal to m1 - m2, the difference between the two population means, and a

variance equal to 1s21>n 12 + 1s22>n 22. That is, the standard error of the difference between

sample means would be equal to 21s21>n 12 + 1s22>n 22. It should be noted that these

properties convey two important points. First, the means of two distributions can be

subtracted from one another, or summed together, using standard arithmetic operations.

Second, since the overall variance of the sampling distribution will be affected by both

contributing distributions, the variances will always be summed even if we are interested

in the difference of the means. This last fact assumes that the two distributions are inde-

pendent of one another.

For our present example we would have a normal distribution with a mean of 0

(if there is no difference between the two population means) and a variance of

312022>154 + 312022>154 = 53.3333. The graph of the sampling distribution is shown

in Figure 5.4.1.

Converting to z We know that the normal distribution described in Example

5.4.1 can be transformed to the standard normal distribution by means of a modification

of a previously learned formula. The new formula is as follows:

z =

1x 1 - x 22 - 1m1 - m22

s21

s22

+

A n1

n2

(5.4.1)

The area under the curve of x 1 - x 2 corresponding to the probability we seek is

the area to the left of x 1 - x 2 = 92 - 105 = -13. The z value corresponding to -13,

assuming that there is no difference between population means, is

z =

-13 - 0

12022

D 15

+

12022

15

=

-13

253.3

=

-13

= -1.78

7.3

By consulting Table D, we find that the area under the standard normal curve to the left

of -1.78 is equal to .0375. In answer to our original question, we say that if there is no

difference between population means, the probability of obtaining a difference between

sample means as large as or larger than 13 is .0375.

Sampling from Normal Populations The procedure we have just

followed is valid even when the sample sizes, n 1 and n 2, are different and when the

population variances, s21 and s22 have different values. The theoretical results on which

this procedure is based may be summarized as follows.

Given two normally distributed populations with means m1 and m2 and variances

s21 and s22, respectively, the sampling distribution of the difference, x 1 - x 2,

between the means of independent samples of size n 1 and n 2 drawn from these

populations is normally distributed with mean m1 - m2 and variance

21s21>n 12 + 1s22>n 22.

Sampling from Nonnormal Populations Many times a researcher is

faced with one or the other of the following problems: the necessity of (1) sampling from

nonnormally distributed populations, or (2) sampling from populations whose functional

forms are not known. A solution to these problems is to take large samples, since when

the sample sizes are large the central limit theorem applies and the distribution of the

difference between two sample means is at least approximately normally distributed with

a mean equal to m1 - m2 and a variance of 1s21>n 12 + 1s22>n 22. To find probabilities

associated with specific values of the statistic, then, our procedure would be the same as

that given when sampling is from normally distributed populations.

EXAMPLE 5.4.2

Suppose it has been established that for a certain type of client the average length of a

home visit by a public health nurse is 45 minutes with a standard deviation of 15 min-

utes, and that for a second type of client the average home visit is 30 minutes long with

a standard deviation of 20 minutes. If a nurse randomly visits 35 clients from the first

and 40 from the second population, what is the probability that the average length of

home visit will differ between the two groups by 20 or more minutes?

Solution:

No mention is made of the functional form of the two populations, so let

us assume that this characteristic is unknown, or that the populations are

not normally distributed. Since the sample sizes are large (greater than 30)

in both cases, we draw on the results of the central limit theorem to answer

the question posed. We know that the difference between sample means is

at least approximately normally distributed with the following mean and

variance:

mx1 - x2 = m1 - m2 = 45 - 30 = 15

2

s 2x1 - x2 =

2

1152

1202

s21

s22

+

=

+

= 16.4286

FIGURE 5.4.2 Sampling distribution of x 1 - x 2 and the

corresponding standard normal distribution, home visit example.

The area under the curve of x 1 - x 2 that we seek is that area to the right

of 20. The corresponding value of z in the standard normal is

z =

1x 1 - x 22 - 1m1 - m22

s21

A n1

+

s22

n2

=

20 - 15

216.4286

=

5

= 1.23

4.0532

In Table D we find that the area to the right of z = 1.23 is

1 - .8907 = .1093. We say, then, that the probability of the nurse’s ran-

dom visits resulting in a difference between the two means as great as or

greater than 20 minutes is .1093. The curve of x 1 - x 2 and the correspon-

ding standard normal curve are shown in Figure 5.4.2.

■

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 m and standard deviation s 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 x B - x A 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:

Men

Women

MeanStandard Deviation

797

660482

414

Use these estimates as the mean m and standard deviation s 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 obtain-

ing a difference between sample means of 100 mg or more?

5.4.3Given two normally distributed populations with equal means and variances of s21 = 100 and

s22 = 80, what is the probability that samples of size n 1 = 25 and n 2 = 16 will yield a value of

x 1 - x 2 greater than or equal to 8?

5.4.4Given two normally distributed populations with equal means and variances of s21 = 240 and

s22 = 350, what is the probability that samples of size n 1 = 40 and n 2 = 35 will yield a value of

x 1 - x 2 as large as or larger than 12?

5.4.5For 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 x girls - x boys will be greater

than 10?

5.5 DISTRIBUTION OF THE

SAMPLE PROPORTION