is a standard normal random variable. By transforming X into Z, we can use the table of areas for the standard normal curve to estimate probabilities associated with X.

For instance, suppose that X is a random variable that represents height. For the population of 18- to 74-year-old females in the United States, height is normally distributed with mean $\mu = 63.9$ inches and standard deviation $\sigma = 2.6$ inches [9]. This distribution is illustrated in Figure 7.16. Observe that

$$Z = \frac{X - 63.9}{2.6}$$

is a standard normal random variable.

If we randomly select a woman from this population, what is the probability that she is between 60 and 68 inches tall? For $x = 60$,

$$\begin{aligned} z &= \frac{60 - 63.9}{2.6} \\ &= -1.50, \end{aligned}$$

and for $x = 68$,

$$\begin{aligned} z &= \frac{68 - 63.9}{2.6} \\ &= 1.58. \end{aligned}$$

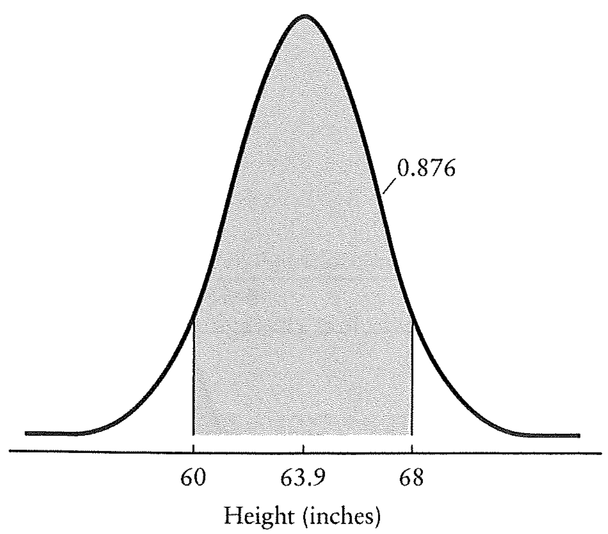

FIGURE 7.16
Distribution of height for females 18 to 74 years of age, United States, 1976–1980

As a result, the probability that x—the woman's height—lies between 60 and 68 inches is equal to the probability that z lies between -1.50 and 1.58 for the standard normal curve. The area to the left of $z = -1.50$ is 0.067, and the area to the right of $z = 1.58$ is 0.057. (Rather than Table A.3, we could use a statistical package to generate these probabilities.) Since the total area under the curve is equal to 1, the area between -1.50 and 1.58 must be

$$\begin{aligned} P(60 \le X \le 68) &= P(-1.50 \le Z \le 1.58) \\ &= 1 - [P(Z < -1.50) + P(Z > 1.58)] \\ &= 1 - [0.067 + 0.057] \\ &= 0.876. \end{aligned}$$

The probability that the woman's height is between 60 and 68 inches is 0.876.

We might also wish to know the value of height that cuts off the upper 5% of this distribution. From Table A.3, we observe that a tail area of 0.050 corresponds to $z = 1.645$. Solving for x,

$$\begin{aligned} z &= 1.645 \\ &= \frac{x - 63.9}{2.6} \end{aligned}$$

and

$$\begin{aligned} x &= 63.9 + (1.645)(2.6) \\ &= 68.2. \end{aligned}$$

Approximately 5% of the women in this population are taller than 68.2 inches.

7.6 Review Exercises

1. What is a probability distribution? What forms may a probability distribution take?
2. What are the parameters of a probability distribution?
3. What are the three properties associated with the binomial distribution?
4. What are the three properties associated with the Poisson distribution?
5. When is the binomial distribution well approximated by the Poisson?
6. What are the properties of the normal distribution?
7. Explain the importance of the standard normal distribution.
8. Let X be a discrete random variable that represents the number of diagnostic services a child receives during an office visit to a pediatric specialist; these services

include procedures such as blood tests and urinanalysis. The probability distribution for X appears below [10].

x	$P(X = x)$
0	0.671
1	0.229
2	0.053
3	0.031
4	0.010
5^+	0.006
Total	1.000

(a) Construct a graph of the probability distribution of X.
(b) What is the probability that a child receives exactly three diagnostic services during an office visit to a pediatric specialist?
(c) What is the probability that he or she receives at least one service? Four or more services?
(d) What is the probability that the child receives exactly three services given that he or she receives at least one service?

9. Suppose that you are interested in monitoring air pollution in Los Angeles, California, over a one-week period. Let X be a random variable that represents the number of days out of the seven on which the concentration of carbon monoxide surpasses a specified level. Do you believe that X has a binomial distribution? Explain.

10. Consider a group of seven individuals selected from the population of 65- to 74-year-olds in the United States. The number of persons in this sample who suffer from diabetes is a binomial random variable with parameters $n = 7$ and $p = 0.125$ [11].
(a) If you wish to make a list of the seven persons chosen, in how many ways can they be ordered?
(b) Without regard to order, in how many ways can you select four individuals from this group of seven?
(c) What is the probability that exactly two of the individuals in the sample suffer from diabetes?
(d) What is the probability that four of them have diabetes?

11. According to the National Health Survey, 9.8% of the population of 18- to 24-year-olds in the United States are left-handed [9].
(a) Suppose that you select ten individuals from this population. In how many ways can the ten persons be ordered?
(b) Without regard to order, in how many ways can you select four individuals from this group of ten?
(c) What is the probability that exactly three of the ten persons are left-handed?
(d) What is the probability that at least six of the ten persons are left-handed?
(e) What is the probability that at most two individuals are left-handed?

12. According to the Behavioral Risk Factor Surveillance System, 58% of all Americans adhere to a sedentary lifestyle [12].
(a) If you selected repeated samples of size twelve from the U.S. population, what would be the mean number of individuals per sample who do not exercise regularly? What would be the standard deviation?
(b) Suppose that you select a sample of twelve individuals and find that ten of them do not exercise regularly. Assuming that the Surveillance System is correct, what is the probability that you would have obtained results as bad as or worse than those you observed?

13. According to the Massachusetts Department of Health, 224 women who gave birth in the state of Massachusetts in 1988 tested positive for the HIV antibody. Assume that, in time, 25% of the babies born to such mothers will also become HIV-positive.
(a) If samples of size 224 were repeatedly selected from the population of children born to mothers with the HIV antibody, what would be the mean number of infected children per sample?
(b) What would be the standard deviation?
(c) Use Chebychev's inequality to describe this distribution.

14. The number of cases of tetanus reported in the United States during a single month in 1989 has a Poisson distribution with parameter $\lambda = 4.5$ [8].
(a) What is the probability that exactly one case of tetanus will be reported during a given month?
(b) What is the probability that at most two cases of tetanus will be reported?
(c) What is the probability that four or more cases will be reported?
(d) What is the mean number of cases of tetanus reported in a one-month period? What is the standard deviation?

15. In a particular county, the average number of suicides reported each month is 2.75 [13]. Assume that the number of suicides follows a Poisson distribution.
(a) What is the probability that no suicides will be reported during a given month?
(b) What is the probability that at most four suicides will be reported?
(c) What is the probability that six or more suicides will be reported?

16. Let X be a random variable that represents the number of infants in a group of 2000 who die before reaching their first birthdays. In the United States, the probability that a child dies during his or her first year of life is 0.0085 [14].
(a) What is the mean number of infants who would die in a group of this size?
(b) What is the probability that at most five infants out of 2000 die in their first year of life?
(c) What is the probability that between 15 and 20 infants die in their first year of life?

17. Consider the standard normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$.
(a) What is the probability that an outcome z is greater than 2.60?
(b) What is the probability that z is less than 1.35?
(c) What is the probability that z is between -1.70 and 3.10?
(d) What value of z cuts off the upper 15% of the standard normal distribution?
(e) What value of z marks off the lower 20% of the distribution?

18. Among females in the United States between 18 and 74 years of age, diastolic blood pressure is normally distributed with mean $\mu = 77$ mm Hg and standard deviation $\sigma = 11.6$ mm Hg [5].

(a) What is the probability that a randomly selected woman has a diastolic blood pressure less than 60 mm Hg?

(b) What is the probability that she has a diastolic blood pressure greater than 90 mm Hg?

(c) What is the probability that the woman has a diastolic blood pressure between 60 and 90 mm Hg?

19. The distribution of weights for the population of males in the United States is approximately normal with mean $\mu = 172.2$ pounds and standard deviation $\sigma = 29.8$ pounds [9].

(a) What is the probability that a randomly selected man weighs less than 130 pounds?

(b) What is the probability that he weighs more than 210 pounds?

(c) What is the probability that among five males selected at random from the population, at least one will have a weight outside the range 130 to 210 pounds?

20. In the Framingham Study, serum cholesterol levels were measured for a large number of healthy males. The population was then followed for 16 years. At the end of this time, the men were divided into two groups: those who had developed coronary heart disease and those who had not. The distributions of the initial serum cholesterol levels for each group were found to be approximately normal. Among individuals who eventually developed coronary heart disease, the mean serum cholesterol level was $\mu_d = 244$ mg/100 ml and the standard deviation was $\sigma_d = 51$ mg/100 ml; for those who did not develop the disease, the mean serum cholesterol level was $\mu_{nd} = 219$ mg/100 ml and the standard deviation was $\sigma_{nd} = 41$ mg/100 ml [15].

(a) Suppose that an initial serum cholesterol level of 260 mg/100 ml or higher is used to predict coronary heart disease. What is the probability of correctly predicting heart disease for a man who will develop it?

(b) What is the probability of predicting heart disease for a man who will not develop it?

(c) What is the probability of failing to predict heart disease for a man who will develop it?

(d) What would happen to the probabilities of false positive and false negative errors if the cutoff point for predicting heart disease is lowered to 250 mg/100 ml?

(e) In this population, does initial serum cholesterol level appear to be useful for predicting coronary heart disease? Why or why not?

Bibliography

[1] National Center for Health Statistics, "Supplements to the Monthly Vital Statistics Reports: Advance Reports, 1986," *Vital and Health Statistics*, Series 24, Number 3, March 1990.

[2] Centers for Disease Control, "The Surgeon General's 1989 Report on Reducing the Health Consequences of Smoking: 25 Years of Progress," *Morbidity and Mortality Weekly Report Supplement*, Volume 38, March 24, 1989.

[3] Ross, S. M., *Introduction to Probability Models*, Orlando, Florida: Academic Press, 1985.

[4] Wilson, R., and Crouch, E. A. C., "Risk Assessment and Comparisons: An Introduction," *Science*, Volume 236, April 17, 1987, 267–270.

[5] National Center for Health Statistics, Drizd, T., Dannenberg, A. L., and Engel, A., "Blood Pressure Levels in Persons 18–74 Years of Age in 1976–1980, and Trends in Blood Pressure From 1960 to 1980 in the United States," *Vital and Health Statistics*, Series 11, Number 234, July 1986.

[6] Castelli, W. P., and Anderson, K., "Antihypertensive Treatment and Plasma Lipoprotein Levels: The Associations in Data from a Population Study," *American Journal of Medicine Supplement*, Volume 80, February 14, 1986, 23–32.

[7] Tye, L., "Many States Tackling Issue of AIDS-Infected Health Care Workers," *The Boston Globe*, May 27, 1991, 29–30.

[8] Centers for Disease Control, "Summary of Notifiable Diseases, United States, 1989," *Morbidity and Mortality Weekly Report*, Volume 39, October 5, 1990.

[9] National Center for Health Statistics, Najjar, M. F., and Rowland, M., "Anthropometric Reference Data and Prevalence of Overweight: United States, 1976–1980," *Vital and Health Statistics*, Series 11, Number 238, October 1987.

[10] National Center for Health Statistics, Woodwell, D., "Office Visits to Pediatric Specialists, 1989," *Vital and Health Statistics*, Advance Data Report Number 208, January 17, 1992.

[11] Centers for Disease Control, "Regional Variation in Diabetes Mellitus Prevalence—United States, 1988 and 1989," *Morbidity and Mortality Weekly Report*, Volume 39, November 16, 1990.

[12] Centers for Disease Control, "Coronary Heart Disease Attributable to Sedentary Lifestyle—Selected States, 1988," *Morbidity and Mortality Weekly Report*, Volume 39, August 17, 1990.

[13] Gibbons, R. D., Clark, D. C., and Fawcett, J., "A Statistical Method for Evaluating Suicide Clusters and Implementing Cluster Surveillance," *American Journal of Epidemiology Supplement*, Volume 132, July 1990, 183–191.

[14] National Center for Health Statistics, Kochanek, K. D., and Hudson, B. L., "Advance Report of Final Mortality Statistics, 1992," *Monthly Vital Statistics Report*, Volume 43, Number 6, March 22, 1995.

[15] MacMahon, S. W., and MacDonald, G. J., "A Population at Risk: Prevalence of High Cholesterol Levels in Hypertensive Patients in the Framingham Study," *American Journal of Medicine Supplement*, Volume 80, February 14, 1986, 40–47.

According to Table A.3, the area to the right of $z = 2.27$ is 0.012 and the area to the left of $z = -2.15$ is 0.016. Therefore,

$$\begin{aligned} P(-2.15 \leq Z \leq 2.27) &= 1 - 0.012 - 0.016 \\ &= 0.972. \end{aligned}$$

About 97.2% of the samples of size 100 have a mean that lies between 70 and 78 years. If we were to select a single random sample of size 100 and find that its sample mean is $\bar{x} = 80$ years, either the sample actually came from a population with a different underlying mean—something higher than $\mu = 73.9$ years—or a rare event has occurred.

To address a different type of question, we might wish to find the upper and lower limits that enclose 80% of the means of samples of size 100. Consulting Table A.3, we find that 10% of the area under a standard normal curve lies above $z = 1.28$ and another 10% lies below $z = -1.28$. Since 80% of the area lies between -1.28 and 1.28, we are interested in values of Z for which

$$-1.28 \leq Z \leq 1.28,$$

and values of $\overline{X}$ for which

$$-1.28 \leq \frac{\overline{X} - 73.9}{1.81} \leq 1.28.$$

Multiplying all three terms of the inequality by 1.81 and adding 73.9 results in

$$73.9 + (-1.28)(1.81) \leq \overline{X} \leq 73.9 + (1.28)(1.81),$$

or, equivalently,

$$71.6 \leq \overline{X} \leq 76.2.$$

Therefore, 80% of the means of samples of size 100 lie between 71.6 years and 76.2 years.

8.5 Review Exercises

1. What is statistical inference?
2. Why is it important that a sample drawn from a population be random?
3. Why is it necessary to understand the properties of a theoretical distribution of means of samples of size n when in practice you will only select a single such sample?
4. What is the standard error of a sample mean? How does the standard error compare to the standard deviation of the population?

5. Explain the central limit theorem.
6. What happens to the amount of sampling variability among a set of sample means $\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots$ as the size of the samples increases?
7. What is consistency?
8. Among adults in the United States, the distribution of albumin levels (albumin is a type of protein) in cerebrospinal fluid is roughly symmetric with mean $\mu = 29.5$ mg/100 ml and standard deviation $\sigma = 9.25$ mg/100 ml [5]. Suppose that you select repeated samples of size 20 from this population and calculate the mean for each sample.
   - **(a)** If you were to select a large number of random samples of size 20, what would be the mean of the sample means?
   - **(b)** What would be their standard deviation? What is another name for this standard deviation of the sample means?
   - **(c)** How does the standard deviation of the sample means compare with the standard deviation of the albumin levels themselves?
   - **(d)** If you were to take all the different sample means and use them to construct a histogram, what would be the shape of their distribution?
   - **(e)** What proportion of the means of samples of size 20 are larger than 33 mg/100 ml?
   - **(f)** What proportion of the means are less than 28 mg/100 ml?
   - **(g)** What proportion of the means are between 29 and 31 mg/100 ml?
9. Consider a random variable X that has a standard normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$.
   - **(a)** What can you say about the distribution of means of samples of size 10 that are drawn from this population? List three properties.
   - **(b)** What proportion of the means of samples of size 10 are greater than 0.60?
   - **(c)** What proportion of the means are less than -0.75?
   - **(d)** What value cuts off the upper 20% of the distribution of means of samples of size 10?
   - **(e)** What value cuts off the lower 10% of the distribution of means?
10. In Denver, Colorado, the distribution of daily measures of ambient nitric acid—a corrosive liquid—is skewed to the right; it has mean $\mu = 1.81 \mu g/m^3$ and standard deviation $\sigma = 2.25 \mu g/m^3$ [6]. Describe the distribution of means of samples of size 40 selected from this population.
11. In Norway, the distribution of birth weights for infants whose gestational age is 40 weeks is approximately normal with mean $\mu = 3500$ grams and standard deviation $\sigma = 430$ grams [7].
    - **(a)** Given a newborn whose gestational age is 40 weeks, what is the probability that his or her birth weight is less than 2500 grams?
    - **(b)** What value cuts off the lower 5% of the distribution of birth weights?
    - **(c)** Describe the distribution of means of samples of size 5 drawn from this population. List three properties.
    - **(d)** What value cuts off the lower 5% of the distribution of samples of size 5?

(e) Given a sample of five newborns all with gestational age 40 weeks, what is the probability that their mean birth weight is less than 2500 grams?
(f) What is the probability that only one of the five newborns has a birth weight less than 2500 grams?

12. For the population of females between the ages of 3 and 74 who participated in the National Health Interview Survey, the distribution of hemoglobin levels has mean $\mu = 13.3$ g/100 ml and standard deviation $\sigma = 1.12$ g/100 ml [8].
(a) If repeated samples of size 15 are selected from this population, what proportion of the samples will have a mean hemoglobin level between 13.0 and 13.6 g/100 ml?
(b) If the repeated samples are of size 30, what proportion will have a mean between 13.0 and 13.6 g/100 ml?
(c) How large must the samples be for 95% of their means to lie within ± 0.2 g/100 ml of the population mean μ?
(d) How large must the samples be for 95% of their means to lie within ± 0.1 g/100 ml of the population mean?

13. In the Netherlands, healthy males between the ages of 65 and 79 have a distribution of serum uric acid levels that is approximately normal with mean $\mu = 341 \mu$mol/l and standard deviation $\sigma = 79$ μmol/l [9].
(a) What proportion of the males have a serum uric acid level between 300 and 400 μmol/l?
(b) What proportion of samples of size 5 have a mean serum uric acid level between 300 and 400 μmol/l?
(c) What proportion of samples of size 10 have a mean serum uric acid level between 300 and 400 μmol/l?
(d) Construct an interval that encloses 95% of the means of samples of size 10. Which would be shorter, a symmetric interval or an asymmetric one?

14. For the population of adult males in the United States, the distribution of weights is approximately normal with mean $\mu = 172.2$ pounds and standard deviation $\sigma = 29.8$ pounds [10].
(a) Describe the distribution of means of samples of size 25 that are drawn from this population.
(b) What is the upper bound for 90% of the mean weights of samples of size 25?
(c) What is the lower bound for 80% of the mean weights?
(d) Suppose that you select a single random sample of size 25 and find that the mean weight for the men in the sample is $\bar{x} = 190$ pounds. How likely is this result? What would you conclude?

15. At the end of Section 8.3, it was noted that for samples of serum cholesterol levels of size 25—drawn from a population with mean $\mu = 211$ mg/100 ml and standard deviation $\sigma = 46$ mg/100 ml—the probability that a sample mean $\bar{x}$ lies within the interval (193.0, 229.0) is 0.95. Furthermore, the probability that the mean lies below 226.1 mg/100 ml is 0.95, and the probability that it is above 195.9 mg/100 ml is 0.95. For all three of these events to happen simultaneously, the sample mean $\bar{x}$ would have to lie in the interval (195.9, 226.1). What is the probability that this occurs?

Bibliography

[1] Lindgren, B. W., *Statistical Theory*, New York: Macmillan, 1976.

[2] Snedecor, G. W., and Cochran, W. G., *Statistical Methods*, Ames, IA: Iowa State University Press, 1980.

[3] National Center for Health Statistics, Fulwood, R., Kalsbeek, W., Rifkind, B., Russell-Briefel, R., Muesing, R., LaRosa, J., and Lippel, K., "Total Serum Cholesterol Levels of Adults 20–74 Years of Age: United States, 1976–1980," *Vital and Health Statistics*, Series 11, Number 236, May 1986.

[4] National Center for Health Statistics, *United States Decennial Life Tables for 1979–1981*, Volume I, Number 1, August 1985.

[5] Scully, R. E., McNeely, B. U., and Mark, E. J., "Case Record of the Massachusetts General Hospital: Weekly Clinicopathological Exercises," *The New England Journal of Medicine*, Volume 314, January 2, 1986, 39–49.

[6] Ostro, B. D., Lipsett, M. J., Wiener, M. B., and Selner, J. C., "Asthmatic Responses to Airborne Acid Aerosols," *American Journal of Public Health*, Volume 81, June 1991, 694–702.

[7] Wilcox, A. J., and Skjærven, R., "Birth Weight and Perinatal Mortality: The Effects of Gestational Age," *American Journal of Public Health*, Volume 82, March 1992, 378–382.

[8] National Center for Health Statistics, Fulwood, R., Johnson, C. L., Bryner, J. D., Gunter, E. W., and McGrath, C. R., "Hematological and Nutritional Biochemistry Reference Data for Persons 6 Months–74 Years of Age: United States, 1976–1980," *Vital and Health Statistics*, Series 11, Number 232, December 1982.

[9] Loenen, H. M. J. A., Eshuis, H., Lowik, M. R. H., Schouten, E. G., Hulshof, K. F. A. M., Odink, J., and Kok, F. J., "Serum Uric Acid Correlates in Elderly Men and Women with Special Reference to Body Composition and Dietary Intake (Dutch Nutrition Surveillance System)," *Journal of Clinical Epidemiology*, Volume 43, Number 12, 1990, 1297–1303.

[10] National Center for Health Statistics, Najjar, M. F., and Rowland, M., "Anthropometric Reference Data and Prevalence of Overweight: United States, 1976–1980," *Vital and Health Statistics*, Series 11, Number 238, October 1987.

TABLE 9.1
Minitab output displaying a 95% confidence interval, standard deviation known

```
THE ASSUMED SIGMA = 6.000

           N     MEAN    STDEV    SE MEAN      95.0 PERCENT C.I.
HEIGHT    31    147.4    6.000      1.078    (145.288, 149.512)
```

TABLE 9.2
Minitab output displaying a 90% confidence interval, standard deviation known

```
THE ASSUMED SIGMA = 6.000

           N     MEAN    STDEV    SE MEAN      90.0 PERCENT C.I.
HEIGHT    31    147.4    6.000      1.078    (145.627, 149.173)
```

all 20 children. Subsequently, the children who had received methylphenidate were given the placebo, and those who had received the placebo now got the drug. (This is what is meant by a *crossover study*.) Measures of each child's attention and behavioral status, both on the drug and on the placebo, were obtained using an instrument called the Parent Rating Scale. Distributions of these scores are approximately normal with unknown means and standard deviations. In general, lower scores indicate an increase in attention. We wish to estimate the mean attention rating scores for children taking methylphenidate and for those taking the placebo.

Since the standard deviation is not known for either population, we use the t distribution to help us construct 95% confidence intervals. For a t distribution with $20 - 1 = 19$ degrees of freedom, 95% of the observations lie between -2.093 and 2.093. Therefore, before a sample of size 20 is drawn from the population, the interval

$$\left(\overline{X} - 2.093\frac{s}{\sqrt{20}},\ \overline{X} + 2.093\frac{s}{\sqrt{20}}\right)$$

has a 95% chance of covering the true mean μ.

The random sample of 20 children enrolled in the study has mean attention rating score $\bar{x}_M = 10.8$ and standard deviation $s_M = 2.9$ when taking methylphenidate and mean rating score $\bar{x}_P = 14.0$ and standard deviation $s_P = 4.8$ when taking the placebo. Therefore, a 95% confidence interval for μ_M, the mean attention rating score for children taking the drug, is

$$\left(10.8 - 2.093\frac{2.9}{\sqrt{20}},\ 10.8 + 2.093\frac{2.9}{\sqrt{20}}\right),$$

or

$$(9.44,\ 12.16),$$

TABLE 9.3
Stata output displaying 95% confidence intervals, standard deviation unknown

Variable	Obs	Mean	Std. Err.	[95% Conf.	Interval]
rating	20	10.8	.6484597	9.442758	12.15724

Variable	Obs	Mean	Std. Err.	[95% Conf.	Interval]
rating	20	14	1.073313	11.75353	16.24647

and a 95% confidence interval for μ_P, the mean rating score for children taking the placebo, is

$$\left(14.0 - 2.093\frac{4.8}{\sqrt{20}},\ 14.0 + 2.093\frac{4.8}{\sqrt{20}}\right),$$

or

$$(11.75,\ 16.25).$$

The relevant output from Stata for both of these intervals is displayed in Table 9.3. When we look at the intervals, it appears that the mean attention rating score is likely to be lower when children with attention-deficit disorder are taking methylphenidate, implying improved attention. However, there is some overlap between the two intervals.

9.5 Review Exercises

1. Explain the difference between point and interval estimation.
2. Describe the 95% confidence interval for a population mean μ. How is the interval interpreted?
3. What are the factors that affect the length of a confidence interval for a mean? Explain briefly.
4. Describe the similarities and differences between the t distribution and the standard normal distribution. If you were trying to construct a confidence interval, when would you use one rather than the other?
5. The distributions of systolic and diastolic blood pressures for female diabetics between the ages of 30 and 34 have unknown means. However, their standard deviations are $\sigma_s = 11.8$ mm Hg and $\sigma_d = 9.1$ mm Hg, respectively [8].

(a) A random sample of ten women is selected from this population. The mean systolic blood pressure for the sample is $\bar{x}_s = 130$ mm Hg. Calculate a two-sided 95% confidence interval for μ_s, the true mean systolic blood pressure.
(b) Interpret this confidence interval.
(c) The mean diastolic blood pressure for the sample of size 10 is $\bar{x}_d = 84$ mm Hg. Find a two-sided 90% confidence interval for μ_d, the true mean diastolic blood pressure of the population.
(d) Calculate a two-sided 99% confidence interval for μ_d.
(e) How does the 99% confidence interval compare to the 90% interval?

6. Consider the t distribution with 5 degrees of freedom.
(a) What proportion of the area under the curve lies to the right of $t = 2.015$?
(b) What proportion of the area lies to the left of $t = -3.365$?
(c) What proportion of the area lies between $t = -4.032$ and $t = 4.032$?
(d) What value of t cuts off the upper 2.5% of the distribution?

7. Consider the t distribution with 21 degrees of freedom.
(a) What proportion of the area under the curve lies to the left of $t = -2.518$?
(b) What proportion of the area lies to the right of $t = 1.323$?
(c) What proportion of the area lies between $t = -1.721$ and $t = 2.831$?
(d) What value of t cuts off the lower 2.5% of the distribution?

8. Before beginning a study investigating the ability of the drug heparin to prevent bronchoconstriction, baseline values of pulmonary function were measured for a sample of 12 individuals with a history of exercise-induced asthma [9]. The mean value of forced vital capacity (FVC) for the sample is $\bar{x}_1 = 4.49$ liters and the standard deviation is $s_1 = 0.83$ liters; the mean forced expiratory volume in 1 second (FEV_1) is $\bar{x}_2 = 3.71$ liters and the standard deviation is $s_2 = 0.62$ liters.
(a) Compute a two-sided 95% confidence interval for μ_1, the true population mean FVC.
(b) Rather than a 95% interval, construct a 90% confidence interval for the true mean FVC. How does the length of the interval change?
(c) Compute a 95% confidence interval for μ_2, the true population mean FEV_1.
(d) In order to construct these confidence intervals, what assumption is made about the underlying distributions of FVC and FEV_1?

9. For the population of infants subjected to fetal surgery for congenital anomalies, the distribution of gestational ages at birth is approximately normal with unknown mean μ and standard deviation σ. A random sample of 14 such infants has mean gestational age $\bar{x} = 29.6$ weeks and standard deviation $s = 3.6$ weeks [10].
(a) Construct a 95% confidence interval for the true population mean μ.
(b) What is the length of this interval?
(c) How large a sample would be required for the 95% confidence interval to have length 3 weeks? Assume that the population standard deviation σ is known and that $\sigma = 3.6$ weeks.
(d) How large a sample would be needed for the 95% confidence interval to have length 2 weeks?

10. Percentages of ideal body weight were determined for 18 randomly selected insulin-dependent diabetics and are shown below [11]. A percentage of 120 means that an individual weighs 20% more than his or her ideal body weight; a percentage of 95 means that the individual weighs 5% less than the ideal.

107	119	99	114	120	104	88	114	124	
116	101	121	152	100	125	114	95	117	(%)

(a) Compute a two-sided 95% confidence interval for the true mean percentage of ideal body weight for the population of insulin-dependent diabetics.
(b) Does this confidence interval contain the value 100%? What does the answer to this question tell you?

11. When eight persons in Massachusetts experienced an unexplained episode of vitamin D intoxication that required hospitalization, it was suggested that these unusual occurrences might be the result of excessive supplementation of dairy milk [12]. Blood levels of calcium and albumin for each individual at the time of hospital admission are shown below.

Calcium (mmol/l)	Albumin (g/l)
2.92	43
3.84	42
2.37	42
2.99	40
2.67	42
3.17	38
3.74	34
3.44	42

(a) Construct a one-sided 95% confidence interval—a lower bound—for the true mean calcium level of individuals who experience vitamin D intoxication.
(b) Compute a 95% lower confidence bound for the true mean albumin level of this group.
(c) For healthy individuals, the normal range of calcium values is 2.12 to 2.74 mmol/l and the range of albumin levels is 32 to 55 g/l. Do you believe that patients suffering from vitamin D intoxication have normal blood levels of calcium and albumin?

12. Serum zinc levels for 462 males between the ages of 15 and 17 are saved under the variable name `zinc` in the data set `serzinc` [13] (Appendix B, Table B.1). The units of measurement for serum zinc level are micrograms per deciliter.
(a) Find a two-sided 95% confidence interval for μ, the true mean serum zinc level for this population of males.
(b) Interpret this confidence interval.
(c) Calculate a 90% confidence interval for μ.
(d) How does the 90% confidence interval compare to the 95% interval?

13. The data set `lowbwt` contains information recorded for a sample of 100 low birth weight infants born in two teaching hospitals in Boston, Massachusetts [14] (Appendix B, Table B.7). Measurements of systolic blood pressure are saved under the variable name `sbp`, while indicators of gender—where 1 represents a male and 0 a female—are saved under the name `sex`.
   (a) Compute a 95% confidence interval for the true mean systolic blood pressure of male low birth weight infants.
   (b) Calculate a 95% confidence interval for the true mean systolic blood pressure of female low birth weight infants.
   (c) Do you think it is possible that males and females have the same mean systolic blood pressure? Explain briefly.

Bibliography

[1] Kaplan, N. M., "Strategies to Reduce Risk Factors in Hypertensive Patients Who Smoke," *American Heart Journal*, Volume 115, January 1988, 288–294.

[2] National Center for Health Statistics, Fulwood, R., Kalsbeek, W., Rifkind, B., Russell-Briefel, R., Muesing, R., LaRosa, J., and Lippel, K., "Total Serum Cholesterol Levels of Adults 20–74 Years of Age: United States, 1976–1980," *Vital and Health Statistics*, Series 11, Number 236, May 1986.

[3] National Center for Health Statistics, Fulwood, R., Johnson, C. L., Bryner, J. D., Gunter, E. W., and McGrath, C. R., "Hematological and Nutritional Biochemistry Reference Data for Persons 6 Months–74 Years of Age: United States, 1976–1980," *Vital and Health Statistics*, Series 11, Number 232, December 1982.

[4] Clark, M., Royal, J., and Seeler, R., "Interaction of Iron Deficiency and Lead and the Hematologic Findings in Children with Severe Lead Poisoning," *Pediatrics*, Volume 81, February 1988, 247–253.

[5] Tsou, V. M., Young, R. M., Hart, M. H., and Vanderhoof, J. A., "Elevated Plasma Aluminum Levels in Normal Infants Receiving Antacids Containing Aluminum," *Pediatrics*, Volume 87, February 1991, 148–151.

[6] Streissguth, A. P., Aase, J. M., Clarren, S. K., Randels, S. P., LaDue, R. A., and Smith, D. F., "Fetal Alcohol Syndrome in Adolescents and Adults," *Journal of the American Medical Association*, Volume 265, April 17, 1991, 1961–1967.

[7] Tirosh, E., Elhasid, R., Kamah, S. C. B., and Cohen, A., "Predictive Value of Placebo Methylphenidate," *Pediatric Neurology*, Volume 9, Number 2, 1993, 131–133.

[8] Klein, B. E. K., Klein, R., and Moss, S. E., "Blood Pressure in a Population of Diabetic Persons Diagnosed After 30 Years of Age," *American Journal of Public Health*, Volume 74, April 1984, 336–339.

[9] Ahmed, T., Garrigo, J., and Danta, I., "Preventing Bronchoconstriction in Exercise-Induced Asthma with Inhaled Heparin," *The New England Journal of Medicine*, Volume 329, July 8, 1993, 90–95.

[10] Longaker, M. T., Golbus, M. S., Filly, R. A., Rosen, M. A., Chang, S. W., and Harrison, M. R., "Maternal Outcome After Open Fetal Surgery," *Journal of the American Medical Association*, Volume 265, February 13, 1991, 737–741.

[11] Saudek, C. D., Selam, J. L., Pitt, H. A., Waxman, K., Rubio, M., Jeandidier, N., Turner, D., Fischell, R. E., and Charles, M. A., "A Preliminary Trial of the Programmable Implantable Medication System for Insulin Delivery," *The New England Journal of Medicine*, Volume 321, August 31, 1989, 574–579.

[12] Jacobus, C. H., Holick, M. F., Shao, Q., Chen, T. C., Holm, I. A., Kolodny, J. M., Fuleihan, G. E. H., and Seely, E. W., "Hypervitaminosis D Associated with Drinking Milk," *The New England Journal of Medicine*, Volume 326, April 30, 1992, 1173–1177.

[13] National Center for Health Statistics, Fulwood, R., Johnson, C. L., Bryner, J. D., Gunter, E. W., and McGrath, C. R., "Hematological and Nutritional Biochemistry Reference Data for Persons 6 Months–74 Years of Age: United States, 1976–1980," *Vital and Health Statistics*, Series 11, Number 232, December 1982.

[14] Leviton, A., Fenton, T., Kuban, K. C. K., and Pagano, M., "Labor and Delivery Characteristics and the Risk of Germinal Matrix Hemorrhage in Low Birth Weight Infants," *Journal of Child Neurology*, Volume 6, October 1991, 35–40.