In general, the upper and lower confidence limits result from replacing each $<$ in (7.6) by $=$ and solving for θ. In the insulating fluid example just considered, $2\lambda\sum x_i = 34.170$ gives $\lambda = 34.170/(2\sum x_i)$ as the upper confidence limit, and the lower limit is obtained from the other equation. Notice that the two interval limits are not equidistant from the point estimate, since the interval is not of the form $\hat{\theta} \pm c$.

Bootstrap Confidence Intervals

The bootstrap technique was introduced in Chapter 6 as a way of estimating $\sigma_{\hat{\theta}}$. It can also be applied to obtain a CI for θ. Consider again estimating the mean μ of a normal distribution when σ is known. Let's replace μ by θ and use $\hat{\theta} = \overline{X}$ as the point estimator. Notice that $1.96\sigma\sqrt{n}$ is the 97.5th percentile of the distribution of $\hat{\theta} - \theta$ [that is, $P(\overline{X} - \mu < 1.96\sigma/\sqrt{n}) = P(Z < 1.96) = .9750$]. Similarly, $-1.96\sigma/\sqrt{n}$ is the 2.5th percentile, so

$$\begin{aligned} .95 &= P(\text{2.5th percentile} < \hat{\theta} - \theta < \text{97.5th percentile}) \\ &= P(\hat{\theta} - \text{2.5th percentile} > \theta > \hat{\theta} - \text{97.5th percentile}) \end{aligned}$$

That is, with

$$\begin{aligned} l &= \hat{\theta} - \text{97.5th percentile of } \hat{\theta} - \theta \\ u &= \hat{\theta} - \text{2.5th percentile of } \hat{\theta} - \theta \end{aligned} \tag{7.7}$$

the CI for θ is (l, u). In many cases, the percentiles in (7.7) cannot be calculated, but they *can* be estimated from bootstrap samples. Suppose we obtain $B = 1000$ bootstrap samples and calculate $\hat{\theta}_1^*, \ldots, \hat{\theta}_{1000}^*$, and $\bar{\theta}^*$ followed by the 1000 differences $\hat{\theta}_1^* - \bar{\theta}^*, \ldots, \hat{\theta}_{1000}^* - \bar{\theta}^*$. The 25th largest and 25th smallest of these differences are estimates of the unknown percentiles in (7.7). Consult the Devore and Berk or Efron books cited in Chapter 6 for more information.

EXERCISES Section 7.1 (1–11)

1. Consider a normal population distribution with the value of σ known.

a. What is the confidence level for the interval $\bar{x} \pm 2.81\sigma/\sqrt{n}$?

b. What is the confidence level for the interval $\bar{x} \pm 1.44\sigma/\sqrt{n}$?

c. What value of $z_{\alpha/2}$ in the CI formula (7.5) results in a confidence level of 99.7%?

d. Answer the question posed in part (c) for a confidence level of 75%.

2. Each of the following is a confidence interval for μ = true average (i.e., population mean) resonance frequency (Hz) for all tennis rackets of a certain type:

(114.4, 115.6) (114.1, 115.9)

a. What is the value of the sample mean resonance frequency?

b. Both intervals were calculated from the same sample data. The confidence level for one of these intervals is 90% and for the other is 99%. Which of the intervals has the 90% confidence level, and why?

3. Suppose that a random sample of 50 bottles of a particular brand of cough syrup is selected and the alcohol content of each bottle is determined. Let μ denote the average alcohol content for the population of all bottles of the brand under study. Suppose that the resulting 95% confidence interval is (7.8, 9.4).

a. Would a 90% confidence interval calculated from this same sample have been narrower or wider than the given interval? Explain your reasoning.

b. Consider the following statement: There is a 95% chance that μ is between 7.8 and 9.4. Is this statement correct? Why or why not?

c. Consider the following statement: We can be highly confident that 95% of all bottles of this type of cough syrup have an alcohol content that is between 7.8 and 9.4. Is this statement correct? Why or why not?

d. Consider the following statement: If the process of selecting a sample of size 50 and then computing the corresponding 95% interval is repeated 100 times, 95 of the resulting intervals will include μ. Is this statement correct? Why or why not?

4. A CI is desired for the true average stray-load loss μ (watts) for a certain type of induction motor when the line current is held at 10 amps for a speed of 1500 rpm. Assume that stray-load loss is normally distributed with $\sigma = 3.0$.

a. Compute a 95% CI for μ when $n = 25$ and $\bar{x} = 58.3$.

b. Compute a 95% CI for μ when $n = 100$ and $\bar{x} = 58.3$.

c. Compute a 99% CI for μ when $n = 100$ and $\bar{x} = 58.3$.

d. Compute an 82% CI for μ when $n = 100$ and $\bar{x} = 58.3$.

e. How large must n be if the width of the 99% interval for μ is to be 1.0?

5. Assume that the helium porosity (in percentage) of coal samples taken from any particular seam is normally distributed with true standard deviation .75.

a. Compute a 95% CI for the true average porosity of a certain seam if the average porosity for 20 specimens from the seam was 4.85.

b. Compute a 98% CI for true average porosity of another seam based on 16 specimens with a sample average porosity of 4.56.

c. How large a sample size is necessary if the width of the 95% interval is to be .40?

d. What sample size is necessary to estimate true average porosity to within .2 with 99% confidence?

6. On the basis of extensive tests, the yield point of a particular type of mild steel-reinforcing bar is known to be normally distributed with $\sigma = 100$. The composition of bars has been slightly modified, but the modification is not believed to have affected either the normality or the value of σ.

a. Assuming this to be the case, if a sample of 25 modified bars resulted in a sample average yield point of 8439 lb, compute a 90% CI for the true average yield point of the modified bar.

b. How would you modify the interval in part (a) to obtain a confidence level of 92%?

7. By how much must the sample size n be increased if the width of the CI (7.5) is to be halved? If the sample size is increased by a factor of 25, what effect will this have on the width of the interval? Justify your assertions.

8. Let $\alpha_1 > 0$, $\alpha_2 > 0$, with $\alpha_1 + \alpha_2 = \alpha$. Then

$$P\left(-z_{\alpha_1} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha_2}\right) = 1 - \alpha$$

a. Use this equation to derive a more general expression for a $100(1 - \alpha)\%$ CI for μ of which the interval (7.5) is a special case.

b. Let $\alpha = .05$ and $\alpha_1 = \alpha/4$, $\alpha_2 = 3\alpha/4$. Does this result in a narrower or wider interval than the interval (7.5)?

9. a. Under the same conditions as those leading to the interval (7.5), $P[(\bar{X} - \mu)/(\sigma/\sqrt{n}) < 1.645] = .95$. Use this to derive a one-sided interval for μ that has infinite width and provides a lower confidence bound on μ. What is this interval for the data in Exercise 5(a)?

b. Generalize the result of part (a) to obtain a lower bound with confidence level $100(1 - \alpha)\%$.

c. What is an analogous interval to that of part (b) that provides an upper bound on μ? Compute this 99% interval for the data of Exercise 4(a).

10. A random sample of $n = 15$ heat pumps of a certain type yielded the following observations on lifetime (in years):

2.0	1.3	6.0	1.9	5.1	.4	1.0	5.3
15.7	.7	4.8	.9	12.2	5.3	.6	

a. Assume that the lifetime distribution is exponential and use an argument parallel to that of Example 7.5 to obtain a 95% CI for expected (true average) lifetime.

b. How should the interval of part (a) be altered to achieve a confidence level of 99%?

c. What is a 95% CI for the standard deviation of the lifetime distribution? [*Hint*: What is the standard deviation of an exponential random variable?]

11. Consider the next 1000 95% CIs for μ that a statistical consultant will obtain for various clients. Suppose the data sets on which the intervals are based are selected independently of one another. How many of these 1000 intervals do you expect to capture the corresponding value of μ? What is the probability that between 940 and 960 of these intervals contain the corresponding value of μ? [*Hint*: Let Y = the number among the 1000 intervals that contain μ. What kind of random variable is Y?]

7.2 Large-Sample Confidence Intervals for a Population Mean and Proportion

The CI for μ given in the previous section assumed that the population distribution is normal with the value of σ known. We now present a large-sample CI whose validity does not require these assumptions. After showing how the argument leading to this interval generalizes to yield other large-sample intervals, we focus on an interval for a population proportion p.

summary information for fracture strengths (MPa) of $n = 169$ ceramic bars fired in a particular kiln: $\bar{x} = 89.10$, $s = 3.73$.

a. Calculate a (two-sided) confidence interval for true average fracture strength using a confidence level of 95%. Does it appear that true average fracture strength has been precisely estimated?

b. Suppose the investigators had believed *a priori* that the population standard deviation was about 4 MPa. Based on this supposition, how large a sample would have been required to estimate μ to within .5 MPa with 95% confidence?

15. Determine the confidence level for each of the following large-sample one-sided confidence bounds:

a. Upper bound: $\bar{x} + .84s/\sqrt{n}$

b. Lower bound: $\bar{x} - 2.05s/\sqrt{n}$

c. Upper bound: $\bar{x} + .67s/\sqrt{n}$

16. The alternating current (AC) breakdown voltage of an insulating liquid indicates its dielectric strength. The article "Testing Practices for the AC Breakdown Voltage Testing of Insulation Liquids" (*IEEE Electrical Insulation Magazine,* 1995: 21–26) gave the accompanying sample observations on breakdown voltage (kV) of a particular circuit under certain conditions.

62	50	53	57	41	53	55	61	59	64	50	53	64	62	50	68
54	55	57	50	55	50	56	55	46	55	53	54	52	47	47	55
57	48	63	57	57	55	53	59	53	52	50	55	60	50	56	58

a. Construct a boxplot of the data and comment on interesting features.

b. Calculate and interpret a 95% CI for true average breakdown voltage μ. Does it appear that μ has been precisely estimated? Explain.

c. Suppose the investigator believes that virtually all values of breakdown voltage are between 40 and 70. What sample size would be appropriate for the 95% CI to have a width of 2 kV (so that μ is estimated to within 1 kV with 95% confidence)?

17. Exercise 1.13 gave a sample of ultimate tensile strength observations (ksi). Use the accompanying descriptive statistics output from Minitab to calculate a 99% lower confidence bound for true average ultimate tensile strength, and interpret the result.

```
 N      Mean   Median  TrMean   StDev  SE Mean
153   135.39   135.40  135.41    4.59     0.37
Minimum   Maximum       Q1       Q3
 122.20    147.70   132.95   138.25
```

18. The article "Ultimate Load Capacities of Expansion Anchor Bolts" (*J. of Energy Engr.,* 1993: 139–158) gave the following summary data on shear strength (kip) for a sample of 3/8-in. anchor bolts: $n = 78$, $\bar{x} = 4.25$, $s = 1.30$. Calculate a lower confidence bound using a confidence level of 90% for true average shear strength.

19. The article "Limited Yield Estimation for Visual Defect Sources" (*IEEE Trans. on Semiconductor Manuf.,* 1997: 17–23) reported that, in a study of a particular wafer inspection process, 356 dies were examined by an inspection probe and 201 of these passed the probe. Assuming a stable process, calculate a 95% (two-sided) confidence interval for the proportion of all dies that pass the probe.

20. The Associated Press (October 9, 2002) reported that in a survey of 4722 American youngsters aged 6 to 19, 15% were seriously overweight (a body mass index of at least 30; this index is a measure of weight relative to height). Calculate and interpret a confidence interval using a 99% confidence level for the proportion of all American youngsters who are seriously overweight.

21. In a sample of 1000 randomly selected consumers who had opportunities to send in a rebate claim form after purchasing a product, 250 of these people said they never did so ("Rebates: Get What You Deserve," *Consumer Reports,* May 2009: 7). Reasons cited for their behavior included too many steps in the process, amount too small, missed deadline, fear of being placed on a mailing list, lost receipt, and doubts about receiving the money. Calculate an upper confidence bound at the 95% confidence level for the true proportion of such consumers who never apply for a rebate. Based on this bound, is there compelling evidence that the true proportion of such consumers is smaller than 1/3? Explain your reasoning.

22. The technology underlying hip replacements has changed as these operations have become more popular (over 250,000 in the United States in 2008). Starting in 2003, highly durable ceramic hips were marketed. Unfortunately, for too many patients the increased durability has been counterbalanced by an increased incidence of squeaking. The May 11, 2008, issue of the *New York Times* reported that in one study of 143 individuals who received ceramic hips between 2003 and 2005, 10 of the hips developed squeaking.

a. Calculate a lower confidence bound at the 95% confidence level for the true proportion of such hips that develop squeaking.

b. Interpret the 95% confidence level used in (a).

23. The Pew Forum on Religion and Public Life reported on Dec. 9, 2009, that in a survey of 2003 American adults, 25% said they believed in astrology.

a. Calculate and interpret a confidence interval at the 99% confidence level for the proportion of all adult Americans who believe in astrology.

b. What sample size would be required for the width of a 99% CI to be at most .05 irrespective of the value of $\hat{p}$?

24. A sample of 56 research cotton samples resulted in a sample average percentage elongation of 8.17 and a sample standard deviation of 1.42 ("An Apparent Relation Between the Spiral Angle ϕ, the Percent Elongation E_1, and the Dimensions of the Cotton Fiber," *Textile Research J.,* 1978: 407–410). Calculate a 95% large-sample CI for the true average percentage elongation μ. What

SUPPLEMENTARY EXERCISES (47–62)

47. Example 1.11 introduced the accompanying observations on bond strength.

11.5	12.1	9.9	9.3	7.8	6.2	6.6	7.0
13.4	17.1	9.3	5.6	5.7	5.4	5.2	5.1
4.9	10.7	15.2	8.5	4.2	4.0	3.9	3.8
3.6	3.4	20.6	25.5	13.8	12.6	13.1	8.9
8.2	10.7	14.2	7.6	5.2	5.5	5.1	5.0
5.2	4.8	4.1	3.8	3.7	3.6	3.6	3.6

a. Estimate true average bond strength in a way that conveys information about precision and reliability. [*Hint*: $\Sigma x_i = 387.8$ and $\Sigma x_i^2 = 4247.08$.]

b. Calculate a 95% CI for the proportion of all such bonds whose strength values would exceed 10.

48. A triathlon consisting of swimming, cycling, and running is one of the more strenuous amateur sporting events. The article "Cardiovascular and Thermal Response of Triathlon Performance" (*Medicine and Science in Sports and Exercise,* 1988: 385–389) reports on a research study involving nine male triathletes. Maximum heart rate (beats/min) was recorded during performance of each of the three events. For swimming, the sample mean and sample standard deviation were 188.0 and 7.2, respectively. Assuming that the heart-rate distribution is (approximately) normal, construct a 98% CI for true mean heart rate of triathletes while swimming.

49. For each of 18 preserved cores from oil-wet carbonate reservoirs, the amount of residual gas saturation after a solvent injection was measured at water flood-out. Observations, in percentage of pore volume, were

23.5	31.5	34.0	46.7	45.6	32.5
41.4	37.2	42.5	46.9	51.5	36.4
44.5	35.7	33.5	39.3	22.0	51.2

(See "Relative Permeability Studies of Gas-Water Flow Following Solvent Injection in Carbonate Rocks," *Soc. of Petroleum Engineers J.,* 1976: 23–30.)

a. Construct a boxplot of this data, and comment on any interesting features.

b. Is it plausible that the sample was selected from a normal population distribution?

c. Calculate a 98% CI for the true average amount of residual gas saturation.

50. A journal article reports that a sample of size 5 was used as a basis for calculating a 95% CI for the true average natural frequency (Hz) of delaminated beams of a certain type. The resulting interval was (229.764, 233.504). You decide that a confidence level of 99% is more appropriate than the 95% level used. What are the limits of the 99% interval? [*Hint*: Use the center of the interval and its width to determine $\bar{x}$ and s.]

51. An April 2009 survey of 2253 American adults conducted by the Pew Research Center's Internet & American Life Project revealed that 1262 of the respondents had at some point used wireless means for online access.

a. Calculate and interpret a 95% CI for the proportion of all American adults who at the time of the survey had used wireless means for online access.

b. What sample size is required if the desired width of the 95% CI is to be at most .04, irrespective of the sample results?

c. Does the upper limit of the interval in (a) specify a 95% upper confidence bound for the proportion being estimated? Explain.

52. High concentration of the toxic element arsenic is all too common in groundwater. The article "Evaluation of Treatment Systems for the Removal of Arsenic from Groundwater" (*Practice Periodical of Hazardous, Toxic, and Radioactive Waste Mgmt.,* 2005: 152–157) reported that for a sample of $n = 5$ water specimens selected for treatment by coagulation, the sample mean arsenic concentration was 24.3 μg/L, and the sample standard deviation was 4.1. The authors of the cited article used t-based methods to analyze their data, so hopefully had reason to believe that the distribution of arsenic concentration was normal.

a. Calculate and interpret a 95% CI for true average arsenic concentration in all such water specimens.

b. Calculate a 90% upper confidence bound for the standard deviation of the arsenic concentration distribution.

c. Predict the arsenic concentration for a single water specimen in a way that conveys information about precision and reliability.

53. Aphid infestation of fruit trees can be controlled either by spraying with pesticide or by inundation with ladybugs. In a particular area, four different groves of fruit trees are selected for experimentation. The first three groves are sprayed with pesticides 1, 2, and 3, respectively, and the fourth is treated with ladybugs, with the following results on yield:

Treatment	n_i = Number of Trees	$\bar{x}_i$ (Bushels/Tree)	s_i
1	100	10.5	1.5
2	90	10.0	1.3
3	100	10.1	1.8
4	120	10.7	1.6