

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 $\hat{\theta} = \bar{X}$ 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, $P(\bar{X} - \mu < 1.96\sigma/\sqrt{n}) = P(Z < 1.96) = .9750$]. Similarly, $-1.96\sigma/\sqrt{n}$ is the 2.5th percentile, so

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

That is, with

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

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 $\hat{\theta}_1^*, \dots, \hat{\theta}_{1000}^*$, and $\bar{\theta}^*$ followed by the 1000 differences $\hat{\theta}_1^* - \bar{\theta}^*, \dots, \hat{\theta}_{1000}^* - \bar{\theta}^*$. 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)

- Consider a normal population distribution with the value of σ known.
 - What is the confidence level for the interval $\bar{x} \pm 2.81\sigma/\sqrt{n}$?
 - What is the confidence level for the interval $\bar{x} \pm 1.44\sigma/\sqrt{n}$?
 - What value of $z_{\alpha/2}$ in the CI formula (7.5) results in a confidence level of 99.7%?
 - Answer the question posed in part (c) for a confidence level of 75%.
- Each of the following is a confidence interval for $\mu =$ 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)

 - What is the value of the sample mean resonance frequency?
 - 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?
- 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).
 - Would a 90% confidence interval calculated from this same sample have been narrower or wider than the given interval? Explain your reasoning.
 - 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?
 - 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?
 - 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 $\sigma = 3.0$.
- Compute a 95% CI for μ when $n = 25$ and $\bar{x} = 58.3$.
 - Compute a 95% CI for μ when $n = 100$ and $\bar{x} = 58.3$.
 - Compute a 99% CI for μ when $n = 100$ and $\bar{x} = 58.3$.
 - Compute an 82% CI for μ when $n = 100$ and $\bar{x} = 58.3$.
 - 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.
- 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.
 - Compute a 98% CI for true average porosity of another seam based on 16 specimens with a sample average porosity of 4.56.
 - How large a sample size is necessary if the width of the 95% interval is to be .40?
 - 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 .
- 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.
 - 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

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

- 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.
 - 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)?
- Generalize the result of part (a) to obtain a lower bound with confidence level $100(1 - \alpha)\%$.
 - 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 | |
- 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.
 - How should the interval of part (a) be altered to achieve a confidence level of 99%?
 - 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 .

nature of the lifetime distribution? Explain your reasoning. [Note: A normal probability plot of the data exhibits a reasonably linear pattern.]

- b. Calculate and interpret a confidence interval with a 99% confidence level for true average lifetime. [Hint: $\bar{x} = 1191.6$ and $s = 506.6$.]
13. The article “**Gas Cooking, Kitchen Ventilation, and Exposure to Combustion Products**” (*Indoor Air, 2006: 65–73*) reported that for a sample of 50 kitchens with gas cooking appliances monitored during a one-week period, the sample mean CO₂ level (ppm) was 654.16, and the sample standard deviation was 164.43.
- a. Calculate and interpret a 95% (two-sided) confidence interval for true average CO₂ level in the population of all homes from which the sample was selected.
- b. Suppose the investigators had made a rough guess of 175 for the value of s before collecting data. What sample size would be necessary to obtain an interval width of 50 ppm for a confidence level of 95%?

14. The negative effects of ambient air pollution on children’s lung function has been well established, but less research is available about the impact of indoor air pollution. The authors of “**Indoor Air Pollution and Lung Function Growth Among Children in Four Chinese Cities**” (*Indoor Air, 2012: 3–11*) investigated the relationship between indoor air-pollution metrics and lung function growth among children ages 6–13 years living in four Chinese cities. For each subject in the study, the authors measured an important lung-capacity index known as FEV₁, the forced volume (in ml) of air that is exhaled in 1 second. Higher FEV₁ values are associated with greater lung capacity. Among the children in the study, 514 came from households that used coal for cooking or heating or both. Their FEV₁ mean was 1427 with a standard deviation of 325. (A complex statistical procedure was used to show that burning coal had a clear negative effect on mean FEV₁ levels.)

- a. Calculate and interpret a 95% (two-sided) confidence interval for true average FEV₁ level in the population of all children from which the sample was selected. Does it appear that the parameter of interest has been accurately estimated?
- b. Suppose the investigators had made a rough guess of 320 for the value of s before collecting data. What sample size would be necessary to obtain an interval width of 50 ml for a confidence level of 95%?

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 U.S. Army commissioned a study to assess how deeply a bullet penetrates ceramic body armor (“**Testing Body Armor Materials for Use by the U.S. Army—Phase III,**” 2012). In the standard test, a cylindrical clay model is layered under the armor vest. A projectile is then fired, causing an indentation in the clay. The deepest impression in the clay is measured as an indication of survivability of someone wearing the armor. Here is data from one testing organization under particular experimental conditions; measurements (in mm) were made using a manually controlled digital caliper:

22.4	23.6	24.0	24.9	25.5	25.6
25.8	26.1	26.4	26.7	27.4	27.6
28.3	29.0	29.1	29.6	29.7	29.8
29.9	30.0	30.4	30.5	30.7	30.7
31.0	31.0	31.4	31.6	31.7	31.9
31.9	32.0	32.1	32.4	32.5	32.5
32.6	32.9	33.1	33.3	33.5	33.5
33.5	33.5	33.6	33.6	33.8	33.9
34.1	34.2	34.6	34.6	35.0	35.2
35.2	35.4	35.4	35.4	35.5	35.7
35.8	36.0	36.0	36.0	36.1	36.1
36.2	36.4	36.6	37.0	37.4	37.5
37.5	38.0	38.7	38.8	39.8	41.0
42.0	42.1	44.6	48.3	55.0	

- a. Construct a boxplot of the data and comment on interesting features.
- b. Construct a normal probability plot. Is it plausible that impression depth is normally distributed? Is a normal distribution assumption needed in order to