

Chapter 15

Inference for One Numerical Population

15.1 Study Suggestions

The concept of a population for a numerical variable is quite complicated. I find it helpful to compare and contrast frequently the development of Chapter 15 for a numerical response with the development of Chapters 5 and 6 for a dichotomous response.

For a dichotomous response, the population is a single number p , but for a numerical response, the population is identified as a picture. This picture is either a probability histogram, for a count variable, or a probability density function, for a measurement variable.

In either case—count or measurement—the picture represents population proportions as areas. For a count response, the proportion of the population giving a response value equal to, say, 1 is the area of the rectangle centered at 1. (If there is no rectangle centered at 1, then none of the population gives 1 for the response value.) For a measurement response, the proportion of the population giving a response value between, say, 1 and 2 is the area under the pdf between the values 1 and 2.

For a dichotomous response, the sample proportion of successes, \hat{p} , is the point estimate of the population. By analogy, a density scale histogram is the estimate of the population for a numerical response. For a dichotomous response, a confidence interval for the population is obtained by taking the point estimate plus or minus the appropriate number. By analogy, for a numerical response, one would like to obtain a confidence set for the population (remember, a picture) that equals the estimate (a density scale histogram) plus or minus an appropriate picture. Unfortunately, the variety in the forms of numerical pop-

ulations make this goal too ambitious. As a result, statisticians study the simpler problems of inference for a numerical feature of the population, usually its mean or median.

The mean is equal to the center of gravity of the picture—either the pdf or the probability histogram, and, thus, is usually easy to visualize. For a pdf the median, ν , is the number that divides the pdf in half in the following sense. The area under the pdf to the left of ν equals one-half, as does the area under the pdf to right of ν . Note that if a pdf is symmetric, the mean equals the median.

Section 15.2 present methods for interval estimation and hypothesis testing for the population median. The section *More Mathematics* below provides a proof of one of the results in the text. Remember the following two items when constructing a confidence interval for ν .

- The data must be sorted from smallest to largest.
- Do not leave your answer as, say,

$$[x_{(7)}, x_{(14)}];$$

replace these symbols by their numerical values.

Section 15.3 presents interval estimation and hypothesis testing for the population mean via the t -distribution. Notice that I say “via the t -distribution” rather than “for normal populations.” As discussed in Section 15.3 in the text, my choice of phrase follows naturally from my belief that one cannot teach (or at least *should not* teach) inference for a numerical population without a serious and intellectually honest—although not necessarily mathematically cluttered—consideration of the notion of robustness.

The optional Section 15.4 provides some elementary guidance on how to decide whether a sequence of trials can be viewed as a random sample from a population. Finally, the optional Section 15.5 provides two methods for predicting a future value of a sequence of trials. I like this material very much, for the same reasons given earlier in the *Pedagogy* section for Chapter 6. Section 15.5 does not strictly depend on Section 15.4, but it is a good idea to cover Section 15.4 if you plan to cover Section 15.5.

15.2 Solutions to Odd-Numbered Exercises

Solutions for Section 15.2

7. (a) Because $n = 40$ is larger than 20, the approximate method must be used.

$$k' = \frac{41}{2} - \frac{1.96\sqrt{40}}{2} = 20.5 - 6.2 = 14.3.$$

Rounding down, $k = 14$ and the confidence interval is

$$[x_{(14)}, x_{(27)}] = [148, 174].$$

- (b) The confidence interval is

$$[x_{(14)}, x_{(27)}] = [141, 159].$$

9. (a) From Table A.7, the interval $[x_{(1)}, x_{(5)}]$ has confidence level equal to 93.8 percent. For Kymn's data, this interval equals $[489, 493]$.

- (b) The confidence interval is

$$[x_{(1)}, x_{(5)}] = [479, 488].$$

11. (a) Because $n = 40$ is larger than 20, the approximate method must be used.

$$k' = \frac{41}{2} - \frac{1.96\sqrt{40}}{2} = 20.5 - 6.2 = 14.3.$$

Rounding down, $k = 14$ and the confidence interval is

$$[x_{(14)}, x_{(27)}] = [24, 40].$$

- (b) The confidence interval is

$$[x_{(14)}, x_{(27)}] = [17, 36].$$

13. (a) Because $n = 40$ is larger than 20, the approximate method must be used.

$$k' = \frac{41}{2} - \frac{1.96\sqrt{40}}{2} = 20.5 - 6.2 = 14.3.$$

Rounding down, $k = 14$ and the confidence interval is

$$[x_{(14)}, x_{(27)}] = [107, 122].$$

- (b) The confidence interval is

$$[x_{(14)}, x_{(27)}] = [92, 108].$$

Solutions for Section 15.3

1. (a) The number of degrees of freedom is $n - 1 = 7 - 1 = 6$, yielding $t = 2.447$. Thus,

$$\bar{x} \pm ts/\sqrt{n} =$$

$$32.44 \pm 2.447(7.03)/\sqrt{7} =$$

$$32.44 \pm 6.50 = [25.94, 38.94].$$

- (b) The number of degrees of freedom is $n - 1 = 28 - 1 = 27$, yielding $t = 2.052$. Thus,

$$\bar{x} \pm ts/\sqrt{n} =$$

$$63.51 \pm 2.052(12.32)/\sqrt{28} =$$

$$63.51 \pm 4.78 = [58.73, 68.29].$$

- (c) The number of degrees of freedom is $n - 1 = 100 - 1 = 99$, yielding $t = z = 1.96$. Thus,

$$\bar{x} \pm ts/\sqrt{n} =$$

$$93.83 \pm 1.96(33.04)/\sqrt{100} =$$

$$93.83 \pm 6.48 = [87.35, 100.31].$$

5. (a) The number of degrees of freedom is $n - 1 = 10 - 1 = 9$, yielding $t = 2.262$. Thus,

$$\bar{x} \pm ts/\sqrt{n} =$$

$$333.0 \pm 2.262(8.18)/\sqrt{10} =$$

$$333.0 \pm 5.9 = [327.1, 338.9].$$

(b) The confidence interval is

$$\begin{aligned}\bar{x} \pm ts/\sqrt{n} &= \\ 319.5 \pm 2.262(7.93)/\sqrt{10} &= \\ 319.5 \pm 5.7 &= [313.8, 325.2].\end{aligned}$$

7. (a) The number of degrees of freedom is $n - 1 = 40 - 1 = 39$, yielding $t = z = 1.96$. Thus,

$$\begin{aligned}\bar{x} \pm ts/\sqrt{n} &= \\ 168.88 \pm 1.96(45.98)/\sqrt{40} &= \\ 168.88 \pm 14.25 &= [154.63, 183.13].\end{aligned}$$

(b) The confidence interval is

$$\begin{aligned}\bar{x} \pm ts/\sqrt{n} &= \\ 149.05 \pm 1.96(22.98)/\sqrt{40} &= \\ 149.05 \pm 7.12 &= [141.93, 156.17].\end{aligned}$$

9. (a) The number of degrees of freedom is $n - 1 = 5 - 1 = 4$, yielding $t = 2.776$. Thus,

$$\begin{aligned}\bar{x} \pm ts/\sqrt{n} &= \\ 491.4 \pm 2.776(1.82)/\sqrt{5} &= \\ 491.4 \pm 2.3 &= [489.1, 493.7].\end{aligned}$$

(b) The confidence interval is

$$\begin{aligned}\bar{x} \pm ts/\sqrt{n} &= \\ 484.2 \pm 2.776(3.42)/\sqrt{5} &= \\ 484.2 \pm 4.2 &= [480.0, 488.4].\end{aligned}$$

11. (a) The number of degrees of freedom is $n - 1 = 40 - 1 = 39$, yielding $t = z = 1.96$. Thus,

$$\begin{aligned}\bar{x} \pm ts/\sqrt{n} &= \\ 35.50 \pm 1.96(24.53)/\sqrt{40} &= \\ 35.50 \pm 7.60 &= [27.90, 43.10].\end{aligned}$$

(b) The confidence interval is

$$\begin{aligned}\bar{x} \pm ts/\sqrt{n} &= \\ 30.42 \pm 1.96(21.70)/\sqrt{40} &= \\ 30.42 \pm 6.72 &= [23.70, 37.14].\end{aligned}$$

13. (a) The number of degrees of freedom is $n - 1 = 40 - 1 = 39$, yielding $t = z = 1.96$. Thus,

$$\begin{aligned}\bar{x} \pm ts/\sqrt{n} &= \\ 106.87 \pm 1.96(29.87)/\sqrt{40} &= \\ 106.87 \pm 9.26 &= [97.61, 116.13].\end{aligned}$$

(b) The confidence interval is

$$\begin{aligned}\bar{x} \pm ts/\sqrt{n} &= \\ 98.18 \pm 1.96(28.33)/\sqrt{40} &= \\ 98.18 \pm 8.78 &= [89.40, 106.96].\end{aligned}$$

15. (a) The number of degrees of freedom is $n - 1 = 15 - 1 = 14$, yielding $t = 2.145$. Thus,

$$\begin{aligned}\bar{x} \pm ts/\sqrt{n} &= \\ 7.94 \pm 2.145(0.2913)/\sqrt{15} &= \\ 7.94 \pm 0.16 &= [7.78, 8.10].\end{aligned}$$

(b) The confidence interval is

$$\begin{aligned}\bar{x} \pm ts/\sqrt{n} &= \\ 7.10 \pm 2.145(0.3253)/\sqrt{15} &= \\ 7.10 \pm 0.18 &= [6.92, 7.28].\end{aligned}$$

Solutions for Section 15.4

3. (a) The number of degrees of freedom is $n - 1 = 89 - 1 = 88$, yielding $t = z = 1.96$. Thus,

$$\begin{aligned}\bar{x} \pm ts/\sqrt{n} &= \\ -0.124 \pm 1.96 \frac{3.493}{\sqrt{89}} &= \\ -0.124 \pm 0.726 &= [-0.850, 0.602].\end{aligned}$$

(b) The observed value of the test statistic is

$$t = \frac{\sqrt{89}(-0.124 - 0)}{3.493} = -0.33.$$

Because the degrees of freedom, 88, is larger than 30, the approximate P-value can be obtained from the standard normal curve. For the third alternative, the approximate P-value is twice the area under the standard normal curve to the right of 0.33. This area equals $2(0.3707) = 0.7414$.

Solutions for Section 15.5

1. (a) The sample mean equals 1,140.30, but since prices are always reported as a whole dollar amount, the point prediction is rounded off to 1,140 dollars. The number of degrees of freedom is $n - 1 = 208 - 1 = 207$, yielding $t = z = 1.96$. Thus, the 95 percent prediction interval is

$$\begin{aligned}\bar{x} \pm ts\sqrt{1 + 1/n} &= \\ 1,140.3 \pm 1.96(388.4)\sqrt{1 + 1/208} &= \\ 1,140.3 \pm 763.1 &= [377, 1,903].\end{aligned}$$

- (b) The point prediction is the sample median, 1,158.50, rounded off to either 1,158 or 1,159 dollars. By trial and error substitution into the appropriate formula, $j = 5$ gives 95.2 percent confidence and $j = 6$ gives 94.3 percent. I will use $j = 5$. The 95.2 percent prediction interval is

$$[x_{(5)}, x_{(204)}] = [459, 1,807].$$

5. (a) The number of degrees of freedom is $n - 1 = 10 - 1 = 9$, yielding $t = 2.262$. Thus, the 95 percent prediction interval is

$$\begin{aligned}\bar{x} \pm ts\sqrt{1 + 1/n} &= \\ 333.0 \pm 2.262(8.18)\sqrt{1 + 1/10} &= \\ 333.0 \pm 19.4 &= [313.6, 352.4].\end{aligned}$$

Since Brian measured his times to the nearest second, this interval should be rounded to yield [314, 352].

- (b) The widest interval is $[x_{(1)}, x_{(10)}] = [321, 347]$, but the probability that this interval is correct equals only 81.8 percent.
- (c) The prediction interval is

$$\begin{aligned}\bar{x} \pm ts\sqrt{1 + 1/n} &= \\ 319.5 \pm 2.262(7.93)\sqrt{1 + 1/10} &= \\ 319.5 \pm 18.8 &= [300.7, 338.3].\end{aligned}$$

After rounding, this interval becomes [301, 338].

- (d) The widest interval is $[x_{(1)}, x_{(10)}] = [301, 327]$, but the probability that this interval is correct equals only 81.8 percent.

7. (a) The number of degrees of freedom is $n - 1 = 40 - 1 = 39$, yielding $t = z = 1.96$. Thus, the 95 percent prediction interval is

$$\begin{aligned}\bar{x} \pm ts\sqrt{1 + 1/n} &= \\ 168.88 \pm 1.96(45.98)\sqrt{1 + 1/40} &= \\ 168.88 \pm 91.24 &= [77.64, 260.12].\end{aligned}$$

Since Jennifer measured her distances to the nearest foot, this interval should be rounded to yield [78, 260].

- (b) The widest interval is $[x_{(1)}, x_{(40)}] = [105, 350]$, and the probability this interval is correct equals 95.1 percent.
- (c) The 95 percent prediction interval is

$$\begin{aligned}\bar{x} \pm ts\sqrt{1 + 1/n} &= \\ 149.05 \pm 1.96(22.98)\sqrt{1 + 1/40} &= \\ 149.05 \pm 45.60 &= [103.45, 194.65].\end{aligned}$$

After rounding, the interval is [103, 195].

- (d) The widest interval is $[x_{(1)}, x_{(40)}] = [86, 200]$, and the probability that this interval is correct equals 95.1 percent.

9. (a) The number of degrees of freedom is $n - 1 = 4 - 1 = 3$, yielding $t = 3.182$. Thus, the 95 percent prediction interval is

$$\begin{aligned}\bar{x} \pm ts\sqrt{1 + 1/n} &= \\ 491.0 \pm 3.182(1.83)\sqrt{1 + 1/4} &= \\ 491.0 \pm 6.5 &= [484.5, 497.5].\end{aligned}$$

Since Kymn measured her times to the nearest second, this interval should be rounded to yield [484, 498]. Kymn's actual fifth time was 493; thus, the prediction interval was correct.

- (b) The 95 percent prediction interval is

$$\bar{x} \pm ts\sqrt{1 + 1/n} =$$

$$483.75 \pm 3.182(3.77)\sqrt{1 + 1/4} =$$

$$483.75 \pm 13.41 = [470.34, 497.16].$$

After rounding, this interval is [470, 497]. Kymn's actual fifth time was 486; thus, the prediction interval was correct.

11. (a) The number of degrees of freedom is $n - 1 = 40 - 1 = 39$, yielding $t = z = 1.96$. Thus, the 95 percent prediction interval is

$$\bar{x} \pm ts\sqrt{1 + 1/n} =$$

$$35.50 \pm 1.96(24.53)\sqrt{1 + 1/40} =$$

$$35.50 \pm 48.68 = [-13.18, 84.18].$$

Since a score must be a nonnegative integer, this interval should be rounded to [0, 84].

- (b) The widest interval is $[x_{(1)}, x_{(40)}] = [2, 124]$, and the probability this interval is correct equals 95.1 percent.

- (c) The 95 percent prediction interval is

$$\bar{x} \pm ts\sqrt{1 + 1/n} =$$

$$30.42 \pm 1.96(21.70)\sqrt{1 + 1/40} =$$

$$30.42 \pm 43.06 = [-12.64, 73.48].$$

After rounding, this interval becomes [0, 73].

- (d) The widest interval is $[x_{(1)}, x_{(40)}] = [0, 85]$, and the probability that this interval is correct equals 95.1 percent.

16	25	26	34	36	39	40
44	45	48	49	49	51	51
51	52	54	59	59	64	65
70	72	74	74	76	79	87
88	90	94	96	98	99	100
102	106	106	106	106	107	108
121	137	141	148	148	149	164

(Hint: $\bar{x} = 79.65$ and $s = 36.68$.) Assume that Roger's times are a random sample from a population.

- (a) Compute the point estimate of the median of the population.
- (b) Find the 90% confidence interval for the median of the population.
2. Refer to the previous question. Assume that Roger's times are a random sample from a population. Use the standard normal curve to find the approximate P-value for testing the null hypothesis that the population median equals 59.5 seconds versus the alternative that it is larger than 59.5 seconds.
3. Refer to question 1. Assume that Roger's times are a random sample from a population. Use the standard normal curve to find the approximate P-value for testing the null hypothesis that the population median equals 100.5 seconds versus the alternative that it is smaller than 100.5 seconds.
4. Roger played 49 games of Columns on his Sega Genesis system. The response is the time, in seconds, required to capture a flashing jewel. Assume that Roger's times are a random sample from a population. Given that $\bar{x} = 79.65$ and $s = 36.68$, find the 95% confidence interval for the mean of the population.
5. Refer to the previous question. Find the P-value for testing the null hypothesis that the population mean equals 60 seconds versus the alternative that it is larger than 60 seconds.
6. Cliff performs a 1000 run simulation experiment. For each run, the following three steps are followed.

15.3 Exam Questions

1. Roger played 49 games of Columns on his Sega Genesis system. The response is the time, in seconds, required to capture a flashing jewel. His times, sorted from smallest to largest, are:

- A random sample of size $n = 15$ is selected from a known pdf that has $\mu = 20$.
- The sample mean, \bar{x} , and sample standard deviation, s , are computed for the 15 numbers selected.
- The values of $\bar{x} - 1.761s/\sqrt{15}$ and $\bar{x} + 1.761s/\sqrt{15}$ are computed.

What is the point of this study and what should Cliff do next?

7. Carla selects a random sample of size $n = 25$ from a population with $\mu = 200$ and $\sigma = 8$. Which of the following is the standardized version of the sample mean \bar{X} ?

- (a) $25(\bar{X} - 200)/2.828$.
- (b) $5(\bar{X} - 200)/2.828$.
- (c) $25(\bar{X} - 200)/8$.
- (d) $5(\bar{X} - 200)/8$.
- (e) None of the above.

8. According to Table A.7 in the text, for $n = 11$ the interval $[x_{(3)}, x_{(9)}]$ has a 93.5% confidence level. If the population is not a pdf, then the actual confidence level must be _____.

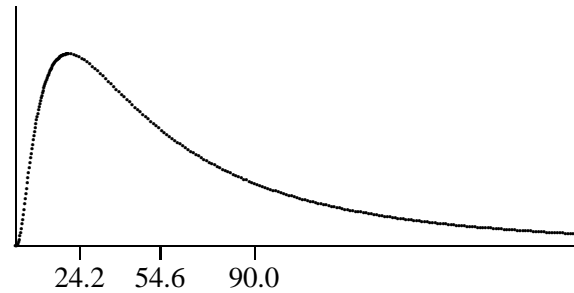
- (a) less than or equal to 93.5%
- (b) equal to 93.5%
- (c) greater than or equal to 93.5%

9. True or false? Any scientific question about a numerical population can be viewed as a question about either the population mean or the population median.

10. A random sample of size 20 is selected from a population with a numerical response. The 20 observations have a symmetric distribution with no outliers. The analyst decides to use the t-distribution to obtain a 99% confidence interval for the population mean.

True or false? The analyst can feel assured that the actual confidence level is very close to the nominal level of 99%.

11. Below is a pdf that represents a population.



Consider the numbers 24.2, 54.6 and 90.0. One of these numbers is the median of this population, and one of these numbers is the mean of this population.

- (a) What is the median of this population?
- (b) What is the mean of this population?

12. A random sample of size 60 is selected from a population described by a pdf, yielding the sorted response values listed below. (Hint: For these data, the sample mean equals 103.3, and the sample standard deviation equals 99.4.)

5	8	9	11	12
18	19	20	20	23
29	30	32	33	36
42	42	45	46	48
48	52	52	52	57
57	57	58	59	59
64	66	69	74	75
76	86	87	94	97
101	104	117	125	126
142	151	159	161	182
201	237	247	256	285
297	300	330	398	410

- (a) Compute the interquartile range of these data.
- (b) Construct the 90 percent confidence interval for the population median.

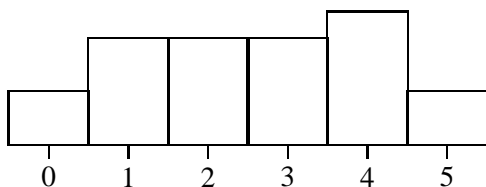
13. A random sample of size seven is selected from a population described by a pdf, yielding the response values listed below. (Hint: For these data, the sample mean equals 67.6, and the sample standard deviation equals 36.7.)

15 14 88 97 94 81 84

- Construct the dot plot of these data.
 - Compute the point estimate of the population median.
 - Use these data to construct a confidence interval for the population median. Select your confidence level, and remember to report its value.
14. A random sample of size 15 is selected from a population described by a pdf that is strongly skewed to the right, yielding the sorted response values listed below. (Hint: For these data, the sample mean equals 90.0, and the sample standard deviation equals 70.0.)

14 17 19 33 47
48 66 67 96 99
102 111 195 213 223

- Use these data to construct the 95 percent confidence interval for the population mean.
 - Briefly discuss the robustness of the confidence interval you just computed.
15. The probability histogram below is the population for a count response. Remember that such histograms represent population proportions as areas. Tom wants to determine the proportion of the population with a response between 1 and 3, inclusive. Do not attempt to compute this area for Tom, but shade the region whose area Tom wants to find.



16. Refer to the previous question. It can be shown that the center of gravity of the probability histogram equals 2.57. Tom performs a 1,000 run simulation experiment. Each run of the simulation experiment consists of the following steps:

- The computer selects a random sample of size five from Tom's population.
- The computer calculates \bar{x} and s for the five numbers just obtained.
- The computer calculates the interval

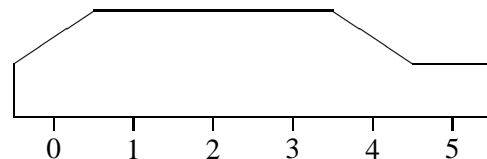
$$\bar{x} \pm 2.776\left(\frac{s}{\sqrt{5}}\right).$$

- The computer answers the question: Does the interval include the number 2.57?

This last question was answered "Yes" for 969 runs, and "No" for the remaining 31 runs.

Briefly explain what is revealed by Tom's simulation experiment. Be specific.

17. The pdf below is the population for a measurement response. Remember that pdfs represent population proportions as areas. Rachel wants to determine the proportion of the population with a response between 1 and 3, inclusive. Do not attempt to compute this area for Rachel, but shade the region whose area Rachel wants to find.



18. Refer to the previous question, Rachel's pdf. Let μ denote the mean of the population. Which one of the following statements is true?
- $\mu \leq 2$
 - $2 < \mu < 2.5$
 - $\mu \geq 2.5$

19. Jerry selects a random sample of size 15 from his favorite pdf, which he knows (because it is his favorite) has $\mu = 0$. Jerry computes the values of \bar{x} and s , plugs them into the formula

$$\bar{x} \pm 1.345\frac{s}{\sqrt{15}},$$

and obtains $[0.01, 0.21]$.

- (a) What is the nominal confidence level of this confidence interval?
- (b) Is the particular interval that Jerry computed correct?
- (c) Jerry repeats the above procedure nine additional times. (That is, nine more times he selects a random sample of size 15 and evaluates the confidence interval formula.) Jerry obtains the following nine confidence intervals for μ .

$[-0.09, 0.11]$	$[-0.21, -0.01]$
$[-0.13, 0.09]$	$[0.08, 0.24]$
$[-0.06, 0.12]$	$[-0.09, 0.10]$
$[-0.12, 0.04]$	$[-0.15, 0.01]$
$[-0.15, 0.08]$	

View Jerry's 10 confidence intervals as the outcome of a simulation experiment with 10 runs. Based on this simulation experiment, what is the simulation estimate of the actual confidence level for Jerry's favorite pdf and the formula he used?

20. Eric plays 18 holes of miniature golf on each of 25 consecutive days. Assume that Eric's scores are a random sample from a population. Test the null hypothesis that the population mean score equals 39 versus the first alternative. (Hint: $\bar{x} = 40.88$, and $s = 4.285$.)
21. George selects a random sample of size 10 from a pdf and obtains the following values:

5 2 7 9 12 3 13 19 16 25

True or false? The 89.1 percent confidence interval for the population median ν is $[7, 19]$.

22. Twenty researchers select random samples from the same population, and each one computes a confidence interval for the mean. The 20 lower bounds of the confidence intervals, after sorting, are:

22.58	22.85	22.88	22.94	22.97
23.22	23.25	23.57	23.67	23.71
23.84	23.85	23.96	24.27	24.40
24.64	24.83	24.99	25.30	25.47

The 20 upper bounds of the confidence intervals, after sorting, are:

24.32	24.52	24.87	24.98	25.08
25.29	25.46	25.56	25.58	25.59
25.72	25.89	26.10	26.44	26.50
26.52	26.59	27.02	27.13	27.45

Unfortunately, I do not know how to match the lower and upper bounds! Given that $\mu = 25$, how many of the 20 confidence intervals are correct?

15.4 Solutions to Exam Questions

1. (a) $\tilde{x} = 74$.
(b)

$$k' = \frac{50}{2} - \frac{1.645\sqrt{49}}{2} = 25 - 5.8 = 19.2.$$

After rounding down, $k = 19$ and the 90% confidence interval for the median of the population is

$$[x_{(19)}, x_{(31)}] = [59, 94].$$

2. There are $Y = 30$ observations larger than 59.5 seconds. Thus,

$$z = \frac{2(30) - 49}{\sqrt{49}} = \frac{11}{7} = 1.57,$$

and the approximate P-value is 0.0582.

3. There are $Y = 14$ observations larger than 100.5 seconds. Thus,

$$z = \frac{2(14) - 49}{\sqrt{49}} = \frac{-21}{7} = -3.00,$$

and the approximate P-value is 0.0013.

4. The 95 percent confidence interval for the mean is below.

$$79.65 \pm 1.96\left(\frac{36.68}{\sqrt{49}}\right) = 79.65 \pm 10.27 = [69.38, 89.92].$$

5. The test statistic is

$$t = \frac{\sqrt{49}(79.65 - 60)}{36.68} = 3.75,$$

with 48 degrees of freedom. Thus, the P-value is smaller than 0.0002.

6. Cliff is trying to estimate the actual confidence level of the nominal 90 percent confidence interval.

For each simulated interval, Cliff needs to determine whether the interval includes the population mean 20. If it does, the trial is a success; otherwise, it is a failure. The number of successes divided by 1000 will be Cliff's estimate of the actual confidence level.

7. (d).

8. (c).

9. False.

10. False.

11. The population median (ν) equals 54.6, and the population mean (μ) equals 90.0. (My reasoning follows. Because of the severe skewness to the right, the mean is larger than the median. Thus, the median cannot be 90.0. In addition, there is much more area to the right of 24.2 than to its left; thus, 24.2 is not the median. As a result, 54.6 must be the median.)

12. (a) $Q_1 = (36+42)/2 = 39$, and $Q_3 = (126+142)/2 = 134$. Thus, $IQR = 134 - 39 = 95$.

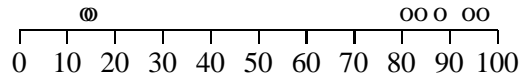
(b) A 90 percent confidence level gives $z = 1.645$; thus,

$$k' = \frac{60 + 1}{2} - \frac{1.645\sqrt{60}}{2} = 30.5 - 6.4 = 24.1.$$

Thus, $k = 24$, and $n + 1 - k = 37$. The confidence interval is

$$[x_{(24)}, x_{(37)}] = [52, 86].$$

13. (a) The dot plot is below.



(b) The sample median is the point estimate of the population median. The sample median equals 84.

(c) The sample size is 20 or smaller, so the exact confidence level can be obtained from Table A.7 of the textbook. There are two acceptable answers. First,

$$[x_{(1)}, x_{(7)}] = [14, 97],$$

is a 98.4 percent confidence interval for the median. Second,

$$[x_{(2)}, x_{(6)}] = [15, 94],$$

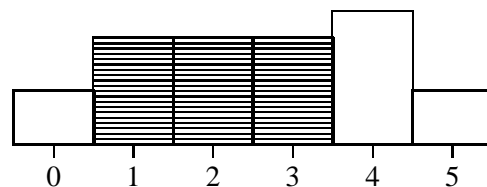
is an 87.5 percent confidence interval for the median.

14. (a) The 95 percent confidence interval is

$$90.0 \pm 2.145 \frac{70.0}{\sqrt{15}} = 90.0 \pm 38.8 = [51.2, 128.8].$$

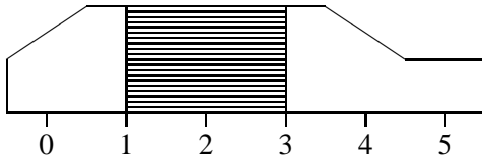
(b) You are told that the population is strongly skewed, perhaps like the lognormal example in Chapter 15 of the text. Based on the findings of the simulation study in the text, I suspect that the actual confidence level for the current population might be substantially smaller than 95 percent.

15. The desired area is shaded below.



16. Each run of the simulation experiment constructed the t-distribution 95 percent confidence interval for the population mean. Of the simulated intervals, 96.9 percent were correct (contained the population mean of 2.57). This is very close to the nominal level of 95 percent.

17. The desired area is shaded below.



18. (b). If a fulcrum is placed at 2, clearly the histogram will tip right-side down; if a fulcrum is placed at 2.5, clearly the histogram will tip left-side down. Thus, the center of gravity is between 2 and 2.5.
19. (a) 80 percent.
 (b) No.
 (c) 70 percent.
20. The test statistic is
- $$t = \frac{\sqrt{25}(40.88 - 39)}{4.285} = 2.19$$
- with 24 degrees of freedom. Thus, the P-value is between 0.01 and 0.025.
21. False. (The data need to be sorted.)
22. It is easier to count incorrect intervals. An interval is incorrect if its lower bound is larger than μ or if its upper bound is smaller than μ . Two intervals have lower bounds larger than 25 and four intervals have upper bounds smaller than 25; thus $2 + 4 = 6$ intervals are incorrect, and 14 of the 20 intervals are correct.

15.5 More Mathematics

The approximate confidence interval for the median given in Section 15.2 of the text is nonstandard for an introductory course, but its derivation is quite simple.

The goal is to find an integer k such that

$$P(X_{(k)} \leq \nu \leq X_{(n+1-k)})$$

is approximately equal to the desired confidence level. Because the population is assumed to be a pdf, the probability is zero that any of the observations is

exactly equal to ν , and we will condition on the event that no observations equal ν .

Define Y to be the number of observations that are smaller than ν . Given that no observations equal ν ,

$$P(X_{(k)} \leq \nu \leq X_{(n+1-k)}) = P(k \leq Y \leq n - k).$$

The sampling distribution of Y is the binomial with parameters n and $p = 0.5$. Using a continuity correction,

$$P(k \leq Y \leq n - k) = P(k - 0.5 \leq Y \leq n + 0.5 - k).$$

Standardize this last expression to obtain

$$P\left(\frac{k - 0.5 - 0.5(n)}{0.5\sqrt{n}} \leq Z \leq \frac{n + 0.5 - k - 0.5(n)}{0.5\sqrt{n}}\right),$$

where Z is the standardized version of Y . Finally, set

$$\frac{k - 0.5 - 0.5(n)}{0.5\sqrt{n}} = -z,$$

where z is the appropriate number (determined by the desired confidence level) from the standard normal curve. Solving for k gives the result in the book, namely,

$$k' = \frac{n + 1}{2} - \frac{z\sqrt{n}}{2}.$$

Because k' typically is not an integer, I advocate rounding down to make the interval wider which will, one hopes, make the actual confidence level at least as large as the desired level.