This is “Small Sample Estimation of a Population Mean”, section 7.2 from the book Beginning Statistics (v. 1.0). For details on it (including licensing), click here.

For more information on the source of this book, or why it is available for free, please see the project's home page. You can browse or download additional books there.

Has this book helped you? Consider passing it on:

Help Creative Commons

Creative Commons supports free culture from music to education. Their licenses helped make this book available to you.

Help a Public School

DonorsChoose.org helps people like you help teachers fund their classroom projects, from art supplies to books to calculators.

Previous Section | Table of Contents | Next Section

7.2 Small Sample Estimation of a Population Mean

LEARNING OBJECTIVES

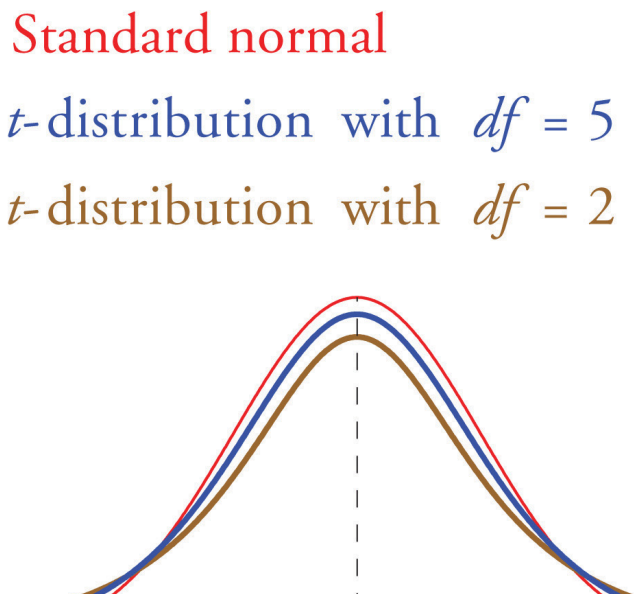
1. To become familiar with Student’s *t*-distribution.
2. To understand how to apply additional formulas for a confidence interval for a population mean.

The confidence interval formulas in the previous section are based on the Central Limit Theorem, the statement that for large samples $\overline{x}$ is normally distributed with mean μ and standard deviation $\sigma / \sqrt{n}$. When the population mean μ is estimated with a small sample ($n < 30$), the Central Limit Theorem does not apply. In order to proceed we assume that the numerical population

from which the sample is taken has a normal distribution to begin with. If this condition is satisfied then when the population standard deviation σ is known the old formula $\overline{x} \pm z_{\alpha/2}(\sigma / \sqrt{n})$ can still be used to construct a $100(1-\alpha)\%$ confidence interval for μ.

If the population standard deviation is unknown and the sample size n is small then when we substitute the sample standard deviation s for σ the normal approximation is no longer valid. The solution is to use a different distribution, called **Student's *t*-distribution with** $n-1$ **degrees of freedom**. Student's *t*-distribution is very much like the standard normal distribution in that it is centered at 0 and has the same qualitative bell shape, but it has heavier tails than the standard normal distribution does, as indicated by Figure 7.5 "Student's ", in which the curve (in brown) that meets the dashed vertical line at the lowest point is the *t*-distribution with two degrees of freedom, the next curve (in blue) is the *t*-distribution with five degrees of freedom, and the thin curve (in red) is the standard normal distribution. As also indicated by the figure, as the sample size n increases, Student's *t*-distribution ever more closely resembles the standard normal distribution. Although there is a different *t*-distribution for every value of n, once the sample size is 30 or more it is typically acceptable to use the standard normal distribution instead, as we will always do in this text.

Figure 7.5 *Student's t-Distribution*



Just as the symbol z_c stands for the value that cuts off a right tail of area c in the standard normal distribution, so the symbol t_c stands for the value that cuts off a right tail of area c in the standard normal distribution. This gives us the following confidence interval formulas.

Small Sample $100(1-\alpha)\%$ Confidence Interval for a Population Mean

If σ is known: $\overline{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

If σ is unknown: $\overline{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$ (degrees of freedom $df = n-1$)

The population must be normally distributed.

A sample is considered small when $n < 30$.

To use the new formula we use the line in Figure 12.3 "Critical Values of " that corresponds to the relevant sample size.

EXAMPLE 5

A sample of size 15 drawn from a normally distributed population has sample mean 35 and sample standard deviation 14. Construct a 95% confidence interval for the population mean, and interpret its meaning.

Solution:

Since the population is normally distributed, the sample is small, and the population standard deviation is unknown, the formula that applies is

$$\overline{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Confidence level 95% means that $\alpha = 1 - 0.95 = 0.05$ so $\alpha / 2 = 0.025$. Since the sample size is *n* = 15, there are $n-1 = 14$ degrees of freedom. By Figure 12.3 "Critical Values of " $t_{0.025} = 2.145$. Thus

$$\overline{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} = 35 \pm 2.145 \frac{14}{\sqrt{15}} = 35 \pm 7.8$$

One may be 95% confident that the true value of *μ* is contained in the interval $(35 - 7.8, 35 + 7.8) = (27.2, 42.8)$.

EXAMPLE 6

A random sample of 12 students from a large university yields mean GPA 2.71 with sample standard deviation 0.51. Construct a 90% confidence interval for the mean GPA of all students at the university. Assume that the numerical population of GPAs from which the sample is taken has a normal distribution.

Solution:

Since the population is normally distributed, the sample is small, and the population standard deviation is unknown, the formula that applies is

$$\overline{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Confidence level 90% means that $\alpha = 1 - 0.90 = 0.10$ so $\alpha / 2 = 0.05$. Since the sample size is *n* = 12, there are $n-1 = 11$ degrees of freedom. By Figure 12.3 "Critical Values of " $t_{0.05} = 1.796$. Thus

$$\overline{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} = 2.71 \pm 1.796 \frac{0.51}{\sqrt{12}} = 2.71 \pm 0.26$$

One may be 90% confident that the true average GPA of all students at the university is contained in the interval $(2.71 - 0.26, 2.71 + 0.26) = (2.45, 2.97)$.

Compare Note 7.9 "Example 4" in Section 7.1 "Large Sample Estimation of a Population Mean" and Note 7.16 "Example 6". The summary statistics in the two samples are the same, but the 90% confidence interval for the average GPA of all students at the university in Note 7.9 "Example 4" in Section 7.1 "Large Sample Estimation of a Population Mean", $(2.63, 2.79)$, is shorter than the 90% confidence interval $(2.45, 2.97)$, in Note 7.16 "Example 6". This is partly because in Note 7.9 "Example 4" the sample size is larger; there is more information pertaining to the true value of μ in the large data set than in the small one.

KEY TAKEAWAYS

- In selecting the correct formula for construction of a confidence interval for a population mean ask two questions: is the population standard deviation σ known or unknown, and is the sample large or small?
- We can construct confidence intervals with small samples only if the population is normal.

EXERCISES

BASIC

1. A random sample is drawn from a normally distributed population of known standard deviation 5. Construct a 99.8% confidence interval for the population mean based on the information given (not all of the information given need be used).

a. n = 16, $\overline{x} = 98$, s = 5.6

b. n = 9, $\overline{x} = 98$, s = 5.6

2. A random sample is drawn from a normally distributed population of known standard deviation 10.7. Construct a 95% confidence interval for the population mean based on the information given (not all of the information given need be used).

   a. n = 25, $\overline{x} = 103.3$, s = 11.0

   b. n = 4, $\overline{x} = 103.3$, s = 11.0

3. A random sample is drawn from a normally distributed population of unknown standard deviation. Construct a 99% confidence interval for the population mean based on the information given.

   a. n = 18, $\overline{x} = 386$, s = 24

   b. n = 7, $\overline{x} = 386$, s = 24

4. A random sample is drawn from a normally distributed population of unknown standard deviation. Construct a 98% confidence interval for the population mean based on the information given.

   a. n = 8, $\overline{x} = 58.3$, s = 4.1

   b. n = 27, $\overline{x} = 58.3$, s = 4.1

5. A random sample of size 14 is drawn from a normal population. The summary statistics are $\overline{x} = 933$ and s = 18.

   a. Construct an 80% confidence interval for the population mean μ.

   b. Construct a 90% confidence interval for the population mean μ.

   c. Comment on why one interval is longer than the other.

6. A random sample of size 28 is drawn from a normal population. The summary statistics are $\overline{x} = 68.6$ and s = 1.28.

   a. Construct a 95% confidence interval for the population mean μ.

   b. Construct a 99.5% confidence interval for the population mean μ.

   c. Comment on why one interval is longer than the other.

APPLICATIONS

7. City planners wish to estimate the mean lifetime of the most commonly planted trees in urban settings. A sample of 16 recently felled trees yielded mean age 32.7 years with standard deviation 3.1 years. Assuming the lifetimes of all such trees are normally distributed, construct a 99.8% confidence interval for the mean lifetime of all such trees.

8. To estimate the number of calories in a cup of diced chicken breast meat, the number of calories in a sample of four separate cups of meat is measured. The sample mean is 211.8 calories with sample standard deviation 0.9 calorie. Assuming the caloric content of all such chicken meat is normally distributed, construct a 95% confidence interval for the mean number of calories in one cup of meat.

9. A college athletic program wishes to estimate the average increase in the total weight an athlete can lift in three different lifts after following a particular training program for six weeks. Twenty-five randomly selected athletes when placed on the program exhibited a mean gain of 47.3 lb with standard deviation 6.4 lb. Construct a 90% confidence interval for the mean increase in lifting capacity all athletes would experience if placed on the training program. Assume increases among all athletes are normally distributed.

10. To test a new tread design with respect to stopping distance, a tire manufacturer manufactures a set of prototype tires and measures the stopping distance from 70 mph on a standard test car. A sample of 25 stopping distances yielded a sample mean 173 feet with sample standard deviation 8 feet. Construct a 98% confidence interval for the mean stopping distance for these tires. Assume a normal distribution of stopping distances.

11. A manufacturer of chokes for shotguns tests a choke by shooting 15 patterns at targets 40 yards away with a specified load of shot. The mean number of shot in a 30-inch circle is 53.5 with standard deviation 1.6. Construct an 80% confidence interval for the mean number of shot in a 30-inch circle at 40 yards for this choke with the specified load. Assume a normal distribution of the number of shot in a 30-inch circle at 40 yards for this choke.

12. In order to estimate the speaking vocabulary of three-year-old children in a particular socioeconomic class, a sociologist studies the speech of four children. The mean and

standard deviation of the sample are $\overline{x} = 1120$ and s = 215 words. Assuming that speaking vocabularies are normally distributed, construct an 80% confidence interval for the mean speaking vocabulary of all three-year-old children in this socioeconomic group.

13. A thread manufacturer tests a sample of eight lengths of a certain type of thread made of blended materials and obtains a mean tensile strength of 8.2 lb with standard deviation 0.06 lb. Assuming tensile strengths are normally distributed, construct a 90% confidence interval for the mean tensile strength of this thread.

14. An airline wishes to estimate the weight of the paint on a fully painted aircraft of the type it flies. In a sample of four repaintings the average weight of the paint applied was 239 pounds, with sample standard deviation 8 pounds. Assuming that weights of paint on aircraft are normally distributed, construct a 99.8% confidence interval for the mean weight of paint on all such aircraft.

15. In a study of dummy foal syndrome, the average time between birth and onset of noticeable symptoms in a sample of six foals was 18.6 hours, with standard deviation 1.7 hours. Assuming that the time to onset of symptoms in all foals is normally distributed, construct a 90% confidence interval for the mean time between birth and onset of noticeable symptoms.

16. A sample of 26 women's size 6 dresses had mean waist measurement 25.25 inches with sample standard deviation 0.375 inch. Construct a 95% confidence interval for the mean waist measurement of all size 6 women's dresses. Assume waist measurements are normally distributed.

ADDITIONAL EXERCISES

17. Botanists studying attrition among saplings in new growth areas of forests diligently counted stems in six plots in five-year-old new growth areas, obtaining the following counts of stems per acre:

$$\begin{array}{ccc} 9{,}432 & 11{,}026 & 10{,}539 \\ 8{,}773 & 9{,}868 & 10{,}247 \end{array}$$

Construct an 80% confidence interval for the mean number of stems per acre in all five-

year-old new growth areas of forests. Assume that the number of stems per acre is normally distributed.

18. Nutritionists are investigating the efficacy of a diet plan designed to increase the caloric intake of elderly people. The increase in daily caloric intake in 12 individuals who are put on the plan is (a minus sign signifies that calories consumed went down):

$$121 \quad 284 \quad -94 \quad 295 \quad 183 \quad 312$$
$$188 \quad -102 \quad 259 \quad 226 \quad 152 \quad 167$$

Construct a 99.8% confidence interval for the mean increase in caloric intake for all people who are put on this diet. Assume that population of differences in intake is normally distributed.

19. A machine for making precision cuts in dimension lumber produces studs with lengths that vary with standard deviation 0.003 inch. Five trial cuts are made to check the machine's calibration. The mean length of the studs produced is 104.998 inches with sample standard deviation 0.004 inch. Construct a 99.5% confidence interval for the mean lengths of all studs cut by this machine. Assume lengths are normally distributed. Hint: Not all the numbers given in the problem are used.

20. The variation in time for a baked good to go through a conveyor oven at a large scale bakery has standard deviation 0.017 minute at every time setting. To check the bake time of the oven periodically four batches of goods are carefully timed. The recent check gave a mean of 27.2 minutes with sample standard deviation 0.012 minute. Construct a 99.8% confidence interval for the mean bake time of all batches baked in this oven. Assume bake times are normally distributed. Hint: Not all the numbers given in the problem are used.

21. Wildlife researchers tranquilized and weighed three adult male polar bears. The data (in pounds) are: 926, 742, 1,109. Assume the weights of all bears are normally distributed.

    a. Construct an 80% confidence interval for the mean weight of all adult male polar bears using these data.
    b. Convert the three weights in pounds to weights in kilograms using the conversion 1 lb = 0.453 kg (so the first datum changes to $(926)(0.453) = 419$). Use the converted data to construct an 80% confidence interval for the mean weight of all adult male polar bears expressed in kilograms.

c. Convert your answer in part (a) into kilograms directly and compare it to your answer in (b). This illustrates that if you construct a confidence interval in one system of units you can convert it directly into another system of units without having to convert all the data to the new units.

22. Wildlife researchers trapped and measured six adult male collared lemmings. The data (in millimeters) are: 104, 99, 112, 115, 96, 109. Assume the lengths of all lemmings are normally distributed.

    a. Construct a 90% confidence interval for the mean length of all adult male collared lemmings using these data.
    b. Convert the six lengths in millimeters to lengths in inches using the conversion 1 mm = 0.039 in (so the first datum changes to (104)(0.039) = 4.06). Use the converted data to construct a 90% confidence interval for the mean length of all adult male collared lemmings expressed in inches.
    c. Convert your answer in part (a) into inches directly and compare it to your answer in (b). This illustrates that if you construct a confidence interval in one system of units you can convert it directly into another system of units without having to convert all the data to the new units.

ANSWERS

1. a. 98 ± 3.9
   b. 98 ± 5.2

3. a. 386 ± 16.4
   b. 386 ± 33.6

5. a. 933 ± 6.5
   b. 933 ± 8.5
   c. Asking for greater confidence requires a longer interval.

7. 32.7 ± 2.9

9. 47.3 ± 2.19

11. 53.5 ± 0.56

13. 8.2 ± 0.04

15. 18.6 ± 1.4

17. 9981 ± 486

19. 104.998 ± 0.004

21. a. 926 ± 200

    b. 419 ± 90

    c. 419 ± 91