ROUND-OFF RULE FOR SAMPLE SIZE *n*

If the computed sample size n is not a whole number, round the value of n up to the next larger whole number.

Example 4: A researcher wants to estimate the mean grade point average of all current college students in the United States. She has developed a procedure to standardize scores from colleges using something other than a scale from 0 and 4. How many grade point averages must be obtained so that the sample mean is within 0.1 of the population mean. Assume that a 90% confidence level is desired. Also assume that a pilot study showed that the population standard deviation is estimated to be 0.88.

$1-\alpha = 90\%$

$\alpha = 0.1$

$\alpha/2 = 0.05$

$z_{0.05} = 1.645$

$\sigma = 0.88$

$E = 0.1$

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

$$n = \left(\frac{1.645 \cdot 0.88}{0.1}\right)^2 \rightarrow \boxed{n \approx 210}$$

7.4 ESTIMATING A POPULATION MEAN: SIGMA NOT KNOWN

Key Concept…

In this section, we present methods for estimating a population mean when the population standard deviation σ is not known. With σ unknown, we use the student t distribution instead of a normal distribution, assuming the relevant requirements are satisfied. The student t distribution was developed by William Gosset (1876-1937). William Gosset was a Guinness Brewery employee. He needed a distribution that could be used with small samples. The brewery where he worked did not allow the publication of research results so he published under the pseudonym "Student". In real circumstances,

σ is typically unknown, which makes the methods of this section realistic and practical.

POINT ESTIMATE

The sample mean $\bar{x}$ is an unbiased estimator of the population mean μ.

STUDENT *t* DISTRIBUTION

If a population has a normal distribution, then the distribution is a student t distribution for all samples of size n. A student t distribution is referred to as a t distribution. Because we do not know the value of the population standard deviation σ, we estimate it with the value of the sample standard deviation s, but this introduces another source of unreliability, especially with small samples. In order to maintain a desired confidence level, we compensate for this additional unreliability by making the confidence interval wider: we use critical values $t_{\alpha/2}$ that are larger than the critical values of $z_{\alpha/2}$ from the normal distribution. A critical value of $t_{\alpha/2}$ can be found using technology or Table A-3.

DEFINITION

The number of **degrees of freedom** for a collection of Sample data is the number of Sample values that can vary after certain restrictions have been imposed on all data values. The number of degrees of freedom is often abbreviated as d.f..

For example: If 10 students have quiz scores with a mean of 80, we can freely assign values to the first 9 scores, but the 10th score is then determined. The Sum of the 10 scores must be 800 so the 10th score must be 800 minus the sum of the first 9 scores.

Because the first 9 scores can be freely selected to any values, we say there are 9 degrees of freedom available. For the applications of this section, the number of degrees of freedom is simply the Sample size minus 1.

Example 1: A sample size of 21 is a simple random sample selected from a normally distributed population. Find the critical value $t_{\alpha/2}$ corresponding to a 95% confidence level.

$n = 21$

$d.f. = 21 - 1 = 20$

$1 - \alpha = 0.95$

$\alpha = 0.05$

area in 2 tails or 0.025 area in 1 tail

$t_{d.f., \alpha/2} = t_{20, 0.025}$

$= \boxed{2.086}$

PROCEDURE FOR CONSTRUCTING A CONFIDENCE INTERVAL FOR μ WITH UNKNOWN σ.

1. Verify that the requirements are satisfied.
2. Using n-1 degrees of freedom, refer to table A3 or use technology to find the critical value $t_{d.f.,\alpha/2}$ that corresponds to the desired confidence level. For the confidence level, refer to the "area in one tail". $t_{df,\alpha/2}$
3. Evaluate the margin of error $E = t_{d.f.,\alpha/2} \cdot \frac{s}{\sqrt{n}}$.
4. Using the value of the calculated margin of error E and the value of the sample mean $\bar{x}$, find the values of the confidence interval limits: $\bar{x} - E$ and $\bar{x} + E$. Substitute those values in the general format for the confidence interval.
5. Round the resulting values by using the following round-off rule.

ROUND-OFF RULE FOR CONFIDENCE INTERVALS USED TO ESTIMATE μ

1. When using the original set of data to construct a confidence interval, round the confidence interval limits to one more decimal place than is used for the original set of data.
2. When the original set of data is unknown and only the summary statistics ($\bar{x}, s, n$) are used, round the confidence interval limits to the same number of digits as the sample mean.

Example 2: In a study designed to test the effectiveness of acupuncture for treating migraine, 142 subjects were treated with acupuncture and 80 subjects were given a sham treatment. The numbers of migraine attacks for the acupuncture treatment group had a mean of 1.8 and a standard deviation of 1.4. The numbers of migraine attacks for the sham treatment group had a mean of 1.6 and a standard deviation of 1.2.

a. Construct a 95% confidence interval estimate of the mean number of migraine attacks for those treated with acupuncture.

$n = 142,\ d.f. = 141$

$\bar{x} = 1.8$

$s = 1.4$

$1-\alpha = 0.95$

$\alpha = 0.05$

$\alpha/2 = 0.025$

$t_{141,\,0.025} = t_{100,\,0.025}$

$= 1.984$

$E = 1.984 \cdot \frac{1.4}{\sqrt{142}} \approx 0.233$

$\bar{x} - E < \mu < \bar{x} + E$

$1.8 - 0.233 < \mu < 1.8 + 0.233$

$\boxed{1.6 < \mu < 2.0}$

b. Construct a 95% confidence interval estimate of the mean number of migraine attacks for those given a sham treatment.

$n = 80,\ d.f. = 79$

$\bar{x} = 1.6$

$s = 1.2$

$1-\alpha = 0.95$

$\alpha = 0.05$

$\alpha/2 = 0.025$

$t_{79,\,0.025} = t_{80,\,0.025}$

$= 1.990$

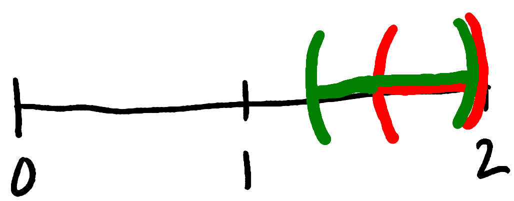
$E = 1.990 \cdot \frac{1.2}{\sqrt{80}} \approx 0.267$

$\bar{x} - E < \mu < \bar{x} + E$

$1.6 - 0.267 < \mu < 1.6 + 0.267$

$\boxed{1.3 < \mu < 1.9}$

c. Compare the two confidence intervals. What do the results suggest about the effectiveness of acupuncture?



red = acupuncture

green = sham

Acupuncture does not seem to be an effective treatment.

IMPORTANT PROPERTIES OF THE STUDENT *t* DISTRIBUTION

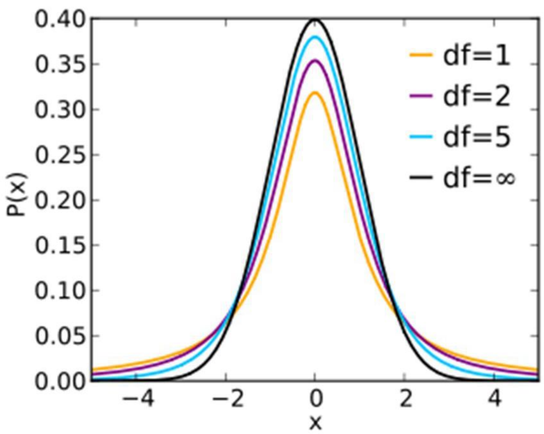
1. The Student *t* distribution is different for different samples sizes.

2. The Student *t* distribution has the same general symmetric bell-shape as the standard normal distribution, but it reflects the greater variability (with wider distributions) that is expected of small samples.

3. The Student *t* distribution has a mean of 0 (just as the standard normal distribution has a mean of 0).

4. The standard deviation of the Student *t* distribution varies with the sample size, but is greater than 1 (unlike the standard normal distribution, which has $\sigma = 1$).

5. As the sample size increases, the Student *t* distribution gets closer to the standard normal distribution.



CHOOSING THE APPROPRIATE DISTRIBUTION

It is sometimes difficult to decide whether to use the standard normal distribution or the student t distribution.

METHOD	CONDITIONS
Use normal (z) distribution	σ known and normally distributed population or σ known and $n > 30$
Use *t* distribution	σ not known and normally distributed population or σ not known and $n > 30$
Use a nonparametric method or bootstrapping	Population is not normally distributed and $n \leq 30$

Example 3: Choosing distributions. You plan to construct a confidence interval for the population mean μ. Use the given data to determine whether the margin of error *E* should be calculated using a critical value of $z_{\sigma/2}$ from the normal distribution, $t_{\sigma/2}$ from a *t* distribution, or neither (methods of this chapter cannot be used).

a. $n = 7,\ \bar{x} = 80,\ s = 8$, and the population has a very skewed distribution

Neither

b. $n = 150,\ \bar{x} = 23.5,\ \sigma = 0.2$, and the population has a skewed distribution

$z_{\alpha/2}$

c. $n = 10,\ \bar{x} = 65,\ s = 12$, and the population has a normal distribution

$t_{\alpha/2}$

d. $n = 13,\ \bar{x} = 5,\ \sigma = 3$, and the population has a normal distribution

$z_{\alpha/2}$

e. $n = 92,\ \bar{x} = 20.7,\ s = 2.5$, and the population has a skewed distribution

$t_{\alpha/2}$

FINDING A POINT ESTIMATE AND *E* FROM A CONFIDENCE INTERVAL

The __sample__ __mean__ __$\bar{x}$__ is the value __midway__ between the __confidence__ __interval__ __limits__.

The __margin__ of __error__ __E__ is __one__ __half__ the __difference__ between those __limits__.

Point estimate of μ:

$$\bar{x} = \frac{\text{upper CI limit} + \text{lower CI limit}}{2}$$

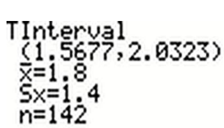
Margin of error:

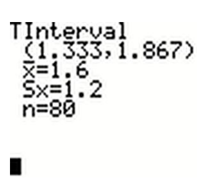
$$E = \frac{\text{upper CI limit} - \text{lower CI limit}}{2}$$

USING CONFIDENCE INTERVALS TO DESCRIBE, EXPLORE, OR COMPARE DATA

In some cases, we might use a __confidence__ __interval__ to achieve an ultimate goal of __estimating__ the __value__ of a __population__ __parameter__. In other cases, __confidence__ __intervals__ might be among the different __tools__ used to __describe__, __explore__, or __compare__ data sets. When two or more data sets have __overlapping__ confidence intervals, one could __possibly__ conclude that there does not appear to be a significant difference between the estimated __means__.

TI-83/84 PLUS





Example 2 using calculator

Example 4: In a sample of seven cars, each car was tested for nitrogen-oxide emissions (in grams per mile) and the following results were obtained: 0.06, 0.11, 0.16, 0.15, 0.14, 0.08, 0.15 (based on data from the EPA).

a. Assuming that this sample is representative of the cars in use, construct a 98% confidence interval estimate of the mean amount of nitrogen-oxide emissions for all cars.

loosely normal

$n = 7$, $d.f. = 6$

$\alpha = 0.02$, $\alpha/2 = 0.01$

$t_{6,0.01} = 3.143$

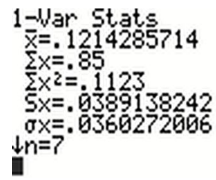
$\bar{x} = 0.121$, $s = 0.039$

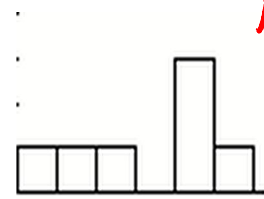
$E = 3.143 \cdot \frac{0.039}{\sqrt{7}}$

$E \approx 0.0463$

$\bar{x} - E < \mu < \bar{x} + E$

$\boxed{0.075 < \mu < 0.167}$





b. If the EPA requires that nitrogen-oxide emissions be less than 0.165 g/mi, can we safely conclude that this requirement is being met?

Example 5: Listed below are 12 lengths (in minutes) of randomly selected movies from Data Set 9 in Appendix B.

110	96	125	94	132	120	136	154	149	94	119	132

a. Construct a 99% confidence interval estimate of the mean length of all movies.

b. Assuming that it takes 30 minutes to empty a theater after a movie, clean it, allow time for the next audience to enter, and show previews, what is the minimum time that a theater manager should plan between start times of movies, assuming that this time will be sufficient for typical movies?