Author: Brenda Gunderson, Ph.D., 2012

Mind on Statistics
Utts/Heckard, 4th Edition, Cengage L, 2012
Text Only: ISBN 9781285135984
Bundled version: ISBN 9780538733489

Module 3: One Sample Confidence Intervals

Objective: This module will help you understand the ideas involved in confidence interval estimation. SPSS allows us to construct one-sample *t* confidence intervals (for a population mean), and to check that the conditions necessary for the interval are valid. We will also examine confidence levels using an applet that simulates taking many random samples (from the same population) and using each sample to construct a confidence interval for the population mean. While the applet addresses confidence intervals and confidence levels for means, remember that both can be created for other parameters, such as proportions.

Overview: Since, generally, a population parameter is an unknown number, we are interested in seeing how close a sample statistic (computed based on a random sample) will be to its corresponding parameter value. Hence, it is important to provide both the sample statistic (the estimate) in addition to a statement that describes the precision of the estimation process. **Confidence intervals** provide a method of stating both how close the value of a statistic is likely to be to the value of a parameter and the accuracy of it being that close. The basic structure for any confidence interval is: **estimate ± (multiplier x standard error)**. The "multiplier x standard error" portion is also called the **margin of error** (or error margin). The multiplier used will depend on the confidence level and the parameter of interest (mean versus proportion). In contrast, the **confidence level** is the **proportion of times the method will produce an interval that contains the true parameter in repeated random sampling**.

Confidence Interval (CI) Summary

1. **Principles for using Confidence Intervals to Guide Decision Making:**

   Principle 1: A value in a CI is an "acceptable" or "reasonable" possibility for the value of a population parameter. A value not in a CI can be rejected as a likely value of the population parameter.

   Principle 2: When the CIs for parameters for two different populations do not overlap, it is reasonable to conclude that the parameters for the two populations are different.

2. The **probability** that the true parameter is a particular, already computed confidence interval is either 0 or 1. Both the interval and the parameter are fixed entities, so either the parameter is in that particular interval, or it is not.

3. **Interpreting a 95% Confidence Interval:** We are 95% confident that the true parameter value lies in side the CI. The interval provides a range of reasonable values for the population parameter.

4. **Interpreting a 95% Confidence Level:** If the procedure were repeated many times (that is, if we repeatedly took a random sample of the same size, and computed a 95% CI based on each sample), we would expect 95% of the resulting CIs to contain the true population parameter.

The above interpretations can be used for any problem by adjusting the confidence level and parameter under study, given the context of the problem. Additionally, phrases such as "true parameter value" and "population" should be stated in context of the problem.

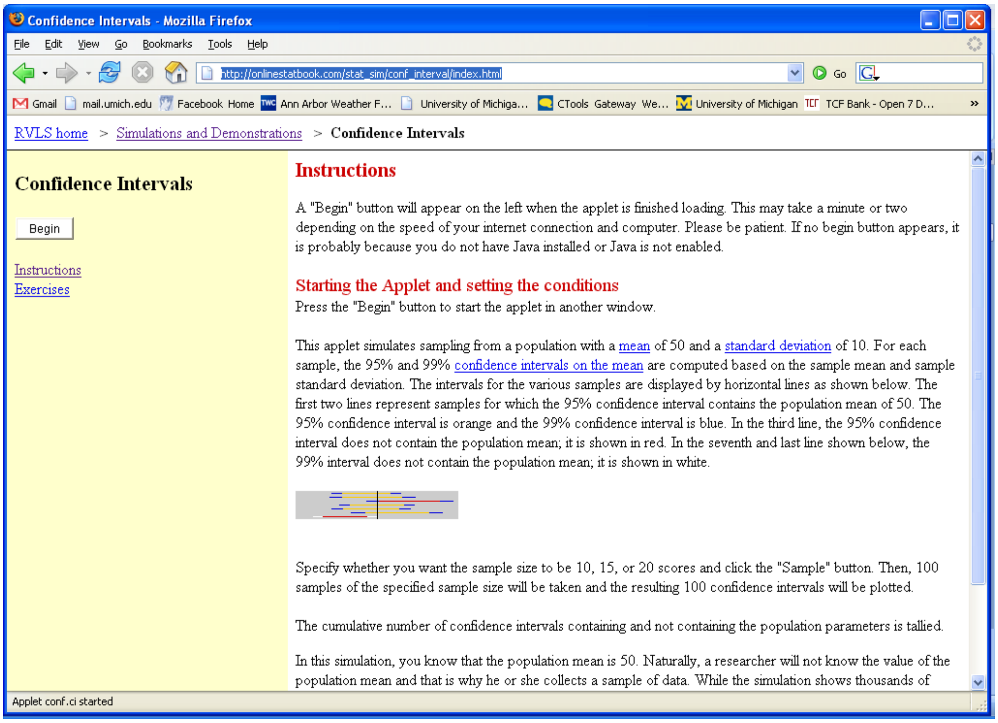
Activity 1: What Does the 95% Confidence Level Really Mean?

In this activity, you will see what the 95% confidence level means.

Open the **confidence interval applet** from the applet link in the "Links to Applets for Modules" folder on the Stat 250 CTools site (in the "Lab Info" folder, which is in the "Resources" folder). Alternatively, the original applet can be found at

http://onlinestatbook.com/stat_sim/conf_interval/index.html

This web site contains a Java applet that will help you understand the meaning of a confidence level. Follow the instructions on the main page of the applet (also described below).

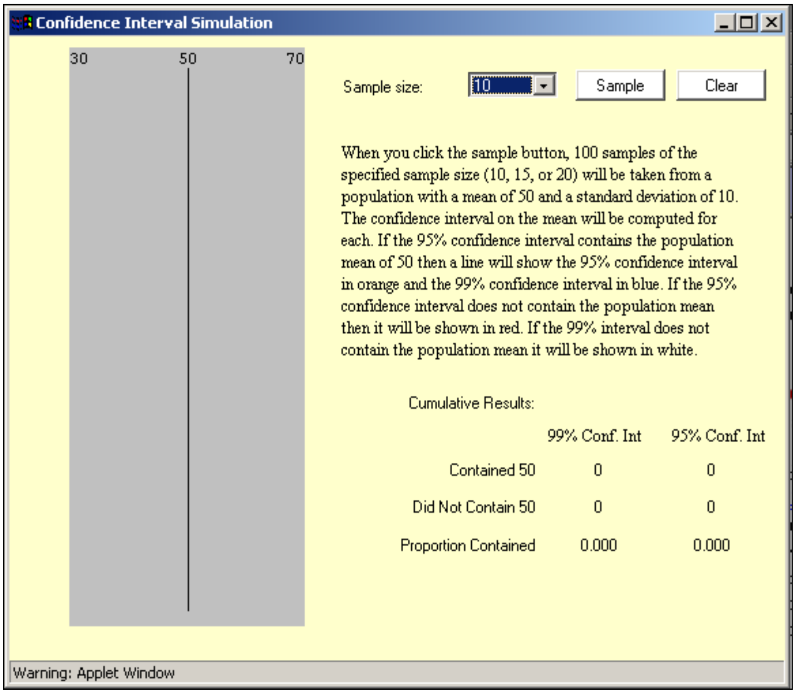


PD-SELF

This applet simulates sampling from a population with a mean of 50 and a standard deviation of 10. For each sample, the 95% and 99% confidence intervals for the mean are computed based on the sample mean and sample standard deviation. The intervals for the various samples are displayed by horizontal lines, with the population mean represented by a solid vertical black line. Each line contains two colors – the part that is either orange or red represents the 95% confidence interval, and the part that is either blue or white represents the 99% confidence interval. Notice that 99% confidence intervals are larger than 95% confidence intervals.

If the 95% confidence interval contains the population mean, it appears orange, and if it does not contain the population mean, it is red. If the 99% confidence interval contains the population mean, it appears blue, and if it does not contain the population mean, it is white. Notice that if the 95% confidence interval contains the population mean, so does the 99% interval.

After you have read the instructions, click on the **Begin** button and you will see the applet window.



Specify the sample size to be 20 and click the **Sample** button one time to have one hundred samples of size n = 20 generated from the population and to display the resulting one hundred 95% confidence intervals and 100 99% confidence intervals. The (cumulative) numbers of confidence intervals that do contain and do not contain the population mean of 50 are tallied at the bottom.

- ❑ Did the first 95% interval you generated contain the population mean of 50? **Yes** **No**
- ❑ Did the second 95% interval you generated contain the population mean of 50? **Yes** **No**
- ❑ Did all of your 95% intervals contain the population mean of 50? **Yes** **No**

- ❑ What proportion of your 95% intervals contained the population mean of 50? _________

- ❑ Did the first 99% interval you generated contain the population mean of 50? **Yes** **No**

- ❑ Did the second 99% interval you generated contain the population mean of 50? **Yes** **No**

- ❑ Did all of the 99% intervals contain the population mean of 50? **Yes** **No**

- ❑ What proportion of the 99% intervals contained the population mean of 50? _________

- ❑ How do the 95% intervals compare to the 99% intervals in terms of width? _________

- ❑ Which of these answers will be the same for every person?

Task: Take additional sets of 100 samples of varying sizes, and look at the cumulative number of 95% and 99% confidence intervals that do contain the population mean of 50.

- After many samples have been drawn for ***n*** **= 10**, the cumulative proportion of 95% confidence intervals that do contain the population mean of 50 is _______________ and the cumulative proportion of 99% confidence intervals that do contain the population mean of 50 is _______________.

- After many samples have been drawn for ***n*** **= 15**, the cumulative proportion of 95% confidence intervals that do contain the population mean of 50 is _______________ and the cumulative proportion of 99% confidence intervals that do contain the population mean of 50 is _______________.

- After many samples have been drawn for ***n*** **= 20**, the cumulative proportion of 95% confidence intervals that do contain the population mean of 50 is _______________ and the cumulative proportion of 99% confidence intervals that do contain the population mean of 50 is _______________.

- Comment on the differences between the confidence intervals for n = 10, n = 15, and n = 20

Activity 2: Interpreting Confidence Levels

In this activity, you will learn how (and how not) to interpret confidence levels. You will be given several examples of interpretations and asked to decide if each one is correct or incorrect.

Background: A study was conducted to learn about eating habits for American families. A random sample of 200 families was selected, and an adult head of household was asked to complete a survey. One question asked was, "Did your family eat dinner together last Sunday night – yes, no?" Based on the results, a 95% (conservative) confidence interval for the proportion of all such families that ate dinner together last Sunday night is given by (0.56, 0.70).

Task: Here are possible interpretations of the confidence *level*. Some are correct and others are not correct. Can you tell the difference?

Interpretation of the 95% Confidence Level	**Correct or Not Correct?**
❑ If we take a large number of independent samples from the population of all American families, 95% of the intervals we calculate based on these samples will contain the true population proportion of families that dinner together last Sunday night.	**Correct** **Not Correct**
❑ If this survey was repeated many times, we are confident the true population proportion of families that ate dinner together last Sunday night is represented in the data 95% of the time.	**Correct** **Not Correct**
❑ The data has a 95% confidence level that the true proportion of families that ate dinner together last Sunday night is within the calculated confidence interval of (0.56, 0.70).	**Correct** **Not Correct**
❑ If the procedure is repeated many times, about 95% of the confidence intervals will contain the true population proportion of families that ate dinner together last Sunday night.	**Correct** **Not Correct**
❑ If the study were repeated many times, we could say with 95% confidence that the proportion of families that ate dinner together last Sunday night was about 0.63.	**Correct** **Not Correct**
❑ The 95% confidence level means that with this method, and for similar samples, we would construct many confidence intervals, and of these, 95% would contain the true population proportion of families that ate dinner together last Sunday night.	**Correct** **Not Correct**

Check Your Understanding:
What is the difference between interpreting a confidence level and interpreting a confidence interval?

Activity 3: Using a One Sample t Confidence Interval to Estimate the Mean Parental Attention Time

In this activity you will see how to use SPSS to compute a 95% confidence interval for a population mean μ based on a random sample from that population. Recall the formula for constructing such an interval for the population mean is given by: $\bar{x} \pm t^*\text{s.e.}(\bar{x})$.

Background: A fact long known but little understood is that twins, in their early years, tend to have lower intelligence quotients and pick up language more slowly than single birth children. Recently, psychologists have speculated that the slower intellectual growth of twins may result from parental neglect (parents having to divide their dedicated time between the two children). A random sample of n = 50 sets of 2½-year-old twin boys was selected. For reasons related to the psychological effect on very young children of any type of external supervision, the questionnaires were self-administered by the parents. Respondents were asked to report the total number of hours dedicated to the attention of the children by the parents during a one week time period.

Task: We will assume that the parental attention time to 2½-year-old twin boys follows a $N(\mu, \sigma)$ distribution. We wish to estimate the mean attention time, μ, given to all 2½-year-old twin boys by their parents, using a 95% confidence interval. The data (in hours) for a random sample of n = 50 families are given in the dataset **attention.sav**.

1. Construct the 95% confidence interval based on this data. You may use **Analyze> Compare Means> One-Sample T Test** (leave test value as 0). What is the resulting interval?

2. Calculate the sample mean using the interval.

3. Interpret the standard error of the sample mean that was used to compute the interval.

4. Interpret the 95% confidence interval using the context of the problem.

- Can you say whether the population mean, μ, is in your interval that you just computed?

  Circle one: ***Yes*** ***No***
  Explain.

6. Can you say there is a 95% probability that the interval you just computed contains the population mean μ?
   Circle one: ***Yes*** ***No***
   Explain.

7. Interpret the 95% confidence level using the context of the problem.

8. Construct the 99% confidence interval and report it.

9. How does the 99% confidence interval compare to the 95% confidence interval?

Check Your Understanding:

A similar study with 50 girl twins was performed and the 95% CI was found to be (19.6, 22.1).

What was the sample mean for the study with girl twins?

Can you conclude girl twins get more parental attention time on average than boy twins using the 95% confidence intervals? ***Yes*** ***No*** because…

What is the ONLY possible reason for the difference in width between the girl and boy twin 95% intervals? (You may need to refer back to the formula used to compute the intervals.)

Think About It:

What assumptions are required for the confidence interval for a population mean to be valid?

Recall in Module 2 that a Q-Q plot was used to assess if a variable appears to follow a normal model for the population. If the attention times were obtained over time, a sequence plot might also be reasonable to assess what aspect of the sample?

Additional Problems for Confidence Level Interpretations

Consider once again the survey of American families, and the question of whether they ate dinner together last Sunday night.

Here are a few more possible interpretations of the 95% confidence interval for practice. Can you tell which are correct and which are not?

Interpretation of the 95% Confidence Level	Correct or Not Correct?
c. In data collected from a normal distribution, 95% of the observed values will lie within two standard deviations of the mean.	**Correct** **Not Correct**
d. 95% of all confidence intervals for different samples taken will contain the true population proportion of families that ate dinner together last Sunday night.	**Correct** **Not Correct**
e. In the long run, the population proportion of families that ate dinner together last Sunday night will be found in the interval 0.56 to 0.70, 95% of the time.	**Correct** **Not Correct**
f. If the procedure is repeated many times, about 95% of the confidence intervals will contain the true population proportion of families that ate dinner together last Sunday night.	**Correct** **Not Correct**

Example Exam Questions on Confidence Intervals for a Population Proportion

1. **Point Estimate and Standard Error**
   After once again losing a basketball game to their archrival, a college's alumni association conducted a survey to see if alumni were in favor of firing the coach. A random sample of 100 alumni was taken and 70 were in favor of firing the coach.

   a. Report the sample proportion of alumni in favor of firing the coach and the corresponding standard error.

b. Which of the following is/are correct interpretation(s) of this standard error?
   1. If repeated samples of 100 alumni were obtained, we would estimate the resulting sample proportions to be about 0.046 from the true population proportion *p* on average.
   2. If repeated samples of 100 alumni were obtained, we would estimate the sample proportion of 0.70 to differ from the true population proportion *p* by about 0.046.
   3. We would estimate that our sample proportion of 0.70 is about 0.046 away from the true population proportion *p* on average.

2. Construct a CI for population proportion *p* and Find the Sample Size.

- Give a (general) 95% confidence interval estimate for the population proportion of alumni in favor of firing the coach.

- Interpret the 95% confidence **interval**.

- Interpret the 95% confidence **level**.

Continued on next page...

Continued from previous page…

d. Give a **conservative** 95% confidence interval estimate for the population proportion of alumni in favor of firing the coach.

e. Is the conservative interval in part (d) wider or narrower than the interval in part (a)?

f. What sample size would be needed for a 95% (conservative) confidence interval with an error margin of 0.05?

3. Point Estimate and Interpretation of the Confidence Level

The 99% confidence interval for the population proportion of members who participated in the fundraiser is given by (0.30, 0.40).

a. The above interval was constructed from the results of a random sample of n = 200 members.

a. What was the proportion of members in the sample that participated in the fundraiser?

b. How many members in the sample participated in the fundraiser?

b. Does the 99% confidence level imply that P(0.30 ≤ p ≤ 0.40) = 0.99?
Yes ***No*** Explain.

Example Exam Question on Confidence Intervals for a Population Mean

A jar of peanuts is supposed to contain 16 ounces of peanuts. The filling machine inevitably experiences some fluctuations in filling, so a quality-control manager randomly samples 50 jars of peanuts from the day's production and measures the ounces of peanuts in each jar. The data sampled from yesterday's production resulted in a 95% confidence interval for μ of (15.4, 16.8).

a. What is the value of the sample mean?
Can you give the value of the population mean?

b. Determine if each statement is true or false. ***Clearly circle your answer.***

About 95% of jars produced yesterday contain between 15.4 and 16.8 ounces of peanuts in them.	**True**	**False**
There is a 95% probability that the sample mean lies between 15.4 and 16.8 ounces.	**True**	**False**
There is a 95% probability that the population mean lies between 15.4 and 16.8 ounces.	**True**	**False**
If repeated samples of 50 jars were obtained, we would expect 95% of the resulting samples to have a sample mean between 15.4 and 16.8 ounces.	**True**	**False**
If repeated samples of 50 jars were obtained, we would expect 95% of the resulting intervals to contain the population mean.	**True**	**False**
If repeated samples of 50 jars were obtained, we would expect 95% of the resulting sample means to lie between 15.4 and 16.8 ounces.	**True**	**False**

c. Suppose we wish to test the claim that the mean number of ounces of peanuts in a jar is 16 using our confidence interval. Is 16 a reasonable value for the mean? ***Yes*** ***No*** ***Can't Tell***
Because …

d. Several plots of the content data were constructed to help verify some of the data conditions. A Q-Q plot is provided for checking the assumption that the population of responses is normally distributed. However, this plot shows some departure from a straight line with a positive slope. Is this a cause for concern that inference based on our confidence interval would not be valid?
Circle one: ***Yes*** ***No***
Because….