Lecture 16c: SPSS output for Confidence Interval Estimates of the Mean

The purpose of this lecture is to illustrate the SPSS output to perform a confidence interval estimate of the mean. So we will estimate the population mean with a spread of values and a certain level of confidence.

Since we are computing means, you need to use an interval or ratio level variable from your data set.

**In your study you will choose one interval/ratio level variable to use**, but here I will provide a few examples using three variables from the study I did on drunk drivers in Honolulu. I will use the variables “Blood Alcohol Content (BAC),” “age” (in years), and “number or prior DUI convictions.” **Note that all three variables are ratio level variables.**

When we are done with the 95% confidence interval of means we will be able to conclude the following:

For BAC: “We are 95% confident that the population mean BAC of people arrested for drunk driving in Honolulu is between __ and ___.”

For Age: “We are 95% confident that the population mean age of people arrested for drunk driving in Honolulu is between ___ years and ___ years.”

For Number of Prior DUI Convictions: “We are 95% confident that the mean number of prior DUI convictions for the population of those arrested for drunk driving in Honolulu is between ___ convictions and ___ convictions.”

**95% confidence interval estimate of mean DUI Blood Alcohol Content (BAC)**

Below are some actual data from my DUI study in Honolulu. The data is a random sample of all people arrested for drunk driving in calendar year 2001. The population would be “all people arrested for drunk driving in Honolulu in 2001.”

The data below is the sample mean Blood Alcohol Content (BAC) from random sample of 475 people arrested for DUI. One may be arrested for drunk driving with a BAC above 0.08, so we would expect the mean of the population to be above .08, but how far?

Well, doing this statistical test will allow us to provide and estimate of the population mean BAC with 95% confidence. In plain English we will be able to say, “We are 95% confident that the population mean BAC of people arrested for drunk driving in Honolulu is between ___ and ___.”
Doing the test on SPSS

To have SPSS perform this operation from the menus choose:

Analyze
   Compare Means >
   One Sample T-Test...

A window will pop up and you highlight and move your *interval/ratio level variable* over to the “Test Variables:” box with the arrow. Then below there is a box that says “Test Value.”

If you want SPSS to automatically compute a 95% confidence interval for you put “0” in the “Test Value” box!

Then press OK. Below is the output you will see
Plugging the numbers from SPSS output into the generic confidence interval formula

Below I show you how to use SPSS to do this sort of calculation in the real world, but for the test I want you to know that SPSS uses the formula we've learned to actually come up with the answer. There are six formulas. They are basically the same thing, but differ based upon the information or data you actually have. The first four are used when your sample size \((n)\) is greater than 30 – that is why they use the \(z\) (table). The last two are used when your sample size \((n)\) is less than 30 – that is why they use the \(t\) (table).

HINT: most student projects have a sample size \((n)\) is less than 30!!! This means you will use formula #5 or #6.

An infinite population is the same thing as an “unknowable population.” If there is NO way to accurately count the members of the population it is unknowable or infinite. If you cannot count the number of people in the population or you do not know it – you do NOT know \(N\). Recall this picture about “big \(N\).”

HINT: most student projects do not know "Big \(N\)" or the size of their population. Thus for must student projects your population is infinite.
Generic Confidence Interval Formulas

Formula #1

Infinite population when \( \sigma \) is known

\[
\bar{x} - z \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z \frac{\sigma}{\sqrt{n}} \quad \text{where} \quad \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}}
\]

Formula #2

Finite population when \( \sigma \) is known

\[
\bar{x} - z \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z \frac{\sigma}{\sqrt{n}} \quad \text{where} \quad \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}} \quad \text{and} \quad N = \text{population size} \quad n = \text{sample size}.
\]

Generic Formula #3

Infinite population when \( \sigma \) is unknown and \( n > 30 \)

\[
\bar{x} - t_{a/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + t_{a/2} \frac{\sigma}{\sqrt{n}} \quad \text{where} \quad \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}}
\]

Formula #4

Finite population when \( \sigma \) is unknown and \( n > 30 \)

\[
\bar{x} - t_{a/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + t_{a/2} \frac{\sigma}{\sqrt{n}} \quad \text{where} \quad \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}} \quad \text{and} \quad N = \text{population size} \quad n = \text{sample size}.
\]

Generic Formula #5

Infinite population when \( \sigma \) is unknown and \( n \leq 30 \) (assume population is normally distributed!)

\[
\bar{x} - t_{a/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + t_{a/2} \frac{\sigma}{\sqrt{n}} \quad \text{where} \quad \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}}
\]

Formula #6

Finite population when \( \sigma \) is unknown and \( n \leq 30 \) (assume population is normally distributed!)

\[
\bar{x} - t_{a/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + t_{a/2} \frac{\sigma}{\sqrt{n}} \quad \text{where} \quad \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}} \quad \text{and} \quad N = \text{population size} \quad n = \text{sample size}.
\]
Plugging the SPSS numbers into the proper generic formula

In this problem my n is greater than 30 but my population is unknown or infinite, so I use formula #2 and the z table. [If your n is less than 30 you will NOT use this formula!].

HINT: If your n is less than 30, the only difference between this example and your test problem is you will use a t-value from a t-table rather than a z value from the z table. Thus the formula you would use will be #5 or #6. An example using the t table is included below.

Since infinite populations are by far the most common, SPSS assumes an infinite population and uses

\[ x - z \frac{\sigma}{\sqrt{n}} < \mu < x + z \frac{\sigma}{\sqrt{n}} \]

where \( \frac{\sigma}{\sqrt{n}} = \frac{s}{\sqrt{n}} \).

SPSS uses the standard deviation and the sample size from the top box to compute “std. error mean.” [By the way, SPSS is lame so it uses “big N” instead of the proper “small n.”]

Recall the “top box” of SPSS output from above.

### One-Sample Statistics

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>real bac refused made into system missing</td>
<td>475</td>
<td>.13502</td>
<td>.062602</td>
<td>.002872</td>
</tr>
</tbody>
</table>

\[ \frac{s}{\sqrt{n}} = \frac{.062602}{\sqrt{475}} = \frac{.062602}{21.79} = .0028723 \]

So now we have the mean \( \bar{x} \), and the standard error of the mean \( \frac{\sigma}{\sqrt{n}} \) and all we need is the z from the z table. So if we are doing a 95% confidence interval, then there is 5% error. The error is split up so 2.5% goes in each tail, so we need the z score for an area that corresponds to .4750 or 47.5%. Looking in the body of the z table for .4750 we find it! \( Z = 1.96 \).
Now we finally have all the information we need to plug the numbers from SPSS into the generic formula. We have

\[ \bar{x} = .13502, \quad \text{standard error of the mean (} \hat{\sigma}_x \text{)} = .0028723, \quad \text{and the } z \text{ from the } z \text{ table} = 1.96 \]

Now we just plug them into the formula:

\[ \bar{x} - z \hat{\sigma}_x < \mu < \bar{x} + z \hat{\sigma}_x \]

\[ .13502 - 1.96(.0028723) < \mu < .13502 + 1.96(.0028723) \]

Note that for the take home test you would stop here. Do not show me the rest of the math “by hand” as it may not match SPSS do to rounding error. If your answers don’t match SPSS on the test you will lose points. I show you the rest of the math below so you can see “where” SPSS gets the answers.

\[ .13502 - .0056 < \mu < .13502 + .0056 \]

\[ 0.12939 < \mu < 0.14064 \]

See how this answer matches the “bottom box” of spss output?

### One-Sample Statistics

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>real bac refused mad...</td>
<td>475</td>
<td>.13502</td>
<td>.062602</td>
<td>.0028723</td>
</tr>
</tbody>
</table>

### One-Sample Test

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
<th>Mean Difference</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>real bac refused mad...</td>
<td>47.005</td>
<td>474</td>
<td>.000</td>
<td>.135017</td>
<td>[0.12937, 0.14066]</td>
</tr>
</tbody>
</table>

**Plain English:** We can be 95% confident that the mean BAC [or blood alcohol content] of the population of those arrested for DUI on Oahu is between 0.129 and 0.141. (It is okay to round on the test.)
An example using the t table

Pretend I have a project that uses UHWO students as my population. We have a sample size of \( n=8 \) and a ratio variable that measures the number of plate lunches bought in the past month. Our population is knowable, but since I don’t actually know how many UHWO students there are (\( N \)) I will assume it is infinite. Thus we use formula #5.

We need the top box of spss output to plug the numbers into the formula and the bottom box to check our answer.

<table>
<thead>
<tr>
<th>One-Sample Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
</tr>
<tr>
<td># of plate lunches bought</td>
</tr>
</tbody>
</table>

We need t from t table. Df=\( n-1 \). 8-1=7. So we look in that row of the t table.

Since we are doing a 95% confidence interval we have 5% error. Half of that error goes in each tail of the bell curve. 2.5% = .025 so we look in that column of the t table.

So the t value for df= 7 and .025 = 2.365

\[
\begin{align*}
\bar{x} - t\frac{s}{\sqrt{n}} &< \mu < \bar{x} + t\frac{s}{\sqrt{n}} \\
5.1250 - 2.365(.71807) &< \mu < 5.1250 + 2.365(.71807)
\end{align*}
\]

Note that for the take home test you would stop here. Do not show me the rest of the math “by hand” as it may not match SPSS do to rounding error. If your answers don’t match SPSS on the test you will lose points. I show you the rest of the math below so you can see “where” SPSS gets the answers.

5.1250 - 1.698 < \( \mu \) < 5.1250 + 1.698

3.427 < \( \mu \) < 6.823

See how that matches the bottom box of the SPSS output? See the far lower right boxes.
Plain English: We can be 95% confident that the mean number of plate lunches bought last month for the population of UHWO students is somewhere between 3.4 and 6.8. (It is okay to round on the test.)

Using SPSS in the real world (but not the test)

Above I showed you how SPSS used the actual generic formula to come up with the answer and how you could compute it by hand using the formula. But in the real world (but not the test!) you can do it the fast way. For the 95% confidence interval you add the “Test Value” to the “Lower” and “Upper” numbers under “95% Confidence Interval of the Difference.” All you need is the “bottom box” of SPSS output:

You may NOT use this simple formula for the test but here’s how you do it in the real world

Test Value + “Lower” number $< \mu <$ Test Value + “Upper” number

0+ .1293 7 < $\mu <$ 0+ .14066

.12937 < $\mu <$ .14066

We can be 95% confident that the mean BAC of people arrested in the City and County of Honolulu for DUI in 2001 is somewhere between 0.12937 and 0.14066.
99% confidence interval estimate of mean DUI Blood Alcohol Content (BAC)

SPSS only computes a 95% confidence interval, so if you want to do another level of confidence you have to compute it by hand.

If you wanted to compute a 99% confidence level (or any other level of confidence for that matter) all you have to do is change the z-score

<table>
<thead>
<tr>
<th>x-bar</th>
<th>n</th>
<th>s</th>
<th>sq. root of n</th>
<th>s/sqroot n</th>
<th>z score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.13502</td>
<td>475</td>
<td>0.062602</td>
<td>21.79449472</td>
<td>0.002872377</td>
<td>2.575</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>x-bar - z*SE</th>
<th>&lt;pop mean&lt;</th>
<th>xbar + z*SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.13502</td>
<td>0.00739637</td>
<td>0.13502</td>
</tr>
<tr>
<td>0.12762363</td>
<td>0.14241637</td>
<td></td>
</tr>
</tbody>
</table>
95% confidence interval estimate of mean age of DUI arrestees

Below is the SPSS output. Computer a 95% confidence interval estimate using the SPSS output then do 99% confidence level interval estimate by hand.

### One-Sample Statistics

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Mean Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>501</td>
<td>32.91</td>
<td>12.590</td>
<td>.562</td>
</tr>
</tbody>
</table>

### One-Sample Test

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
<th>Mean Difference</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>58.504</td>
<td>500</td>
<td>.000</td>
<td>32.908</td>
<td>31.80 - 34.01</td>
</tr>
</tbody>
</table>

**95% confidence interval using SPSS**

For the 95% confidence interval you add the “Test Value” to the “Lower” and “Upper” numbers under “95% Confidence Interval of the Difference”

0+ 31.80 < µ < 0+ 34.01

31.80 < µ < .34.01 So you can be 95% confident that the mean age of people arrested in the City and County of Honolulu for DUI in 2001 is somewhere between 31.8 and 34.01 years.

**99% confidence interval by hand**

see below

<table>
<thead>
<tr>
<th>x-bar</th>
<th>n</th>
<th>s</th>
<th>sq.root of n</th>
<th>s/sqroot n</th>
<th>z score</th>
</tr>
</thead>
<tbody>
<tr>
<td>32.91</td>
<td>501</td>
<td>12.59</td>
<td>22.38302929</td>
<td>0.562479718</td>
<td>2.575</td>
</tr>
</tbody>
</table>

x-bar - z*SE  

<pop mean< xbar + z*SE

<table>
<thead>
<tr>
<th>x-bar</th>
<th>z*SE</th>
<th>&lt;pop mean&lt; xbar + z*SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>32.91</td>
<td>1.448385274</td>
<td>32.91 + 1.448385274</td>
</tr>
<tr>
<td>31.46161473</td>
<td>34.35838527</td>
<td></td>
</tr>
</tbody>
</table>

So you can be 99% confident that the mean age of people arrested in the City and County of Honolulu for DUI in 2001 is somewhere between 31.46 and 34.56 years.
95% confidence interval estimate of mean number of prior DUI convictions of people arrested for DUI in the City and County of Honolulu in 2001

Below is the SPSS output. Compute a 95% confidence interval estimate using the SPSS output then do 99% confidence level interval estimate by hand.

### One-Sample Statistics

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of Pre-dui study dui convictions ...</td>
<td>461</td>
<td>.63</td>
<td>1.632</td>
<td>.076</td>
</tr>
</tbody>
</table>

### One-Sample Test

<table>
<thead>
<tr>
<th></th>
<th>Test Value = 0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t</td>
</tr>
<tr>
<td>Total number of Pre-dui study dui convictions ...</td>
<td>8.245</td>
</tr>
</tbody>
</table>

### 95% confidence interval using SPSS

For the 95% confidence interval you add the “Test Value” to the “Lower” and “Upper” numbers under “95% Confidence Interval of the Difference”

\[
0^+ .48 < \mu < 0^+ .78
\]

\[
.48 < \mu < .78 \quad \text{So you can be 95% confident that the mean number of prior DUI convictions of the population of people arrested in the City and County of Honolulu for DUI in 2001 is somewhere between .48 and .78 convictions}
\]

### 99% confidence interval by hand

see below
<table>
<thead>
<tr>
<th>x-bar</th>
<th>n</th>
<th>s</th>
<th>sq. root of n</th>
<th>s/sqroot n</th>
<th>z score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.63</td>
<td>461</td>
<td>1.632</td>
<td>21.47091055</td>
<td>0.076009818</td>
<td>2.575</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>x-bar -</th>
<th>z*SE</th>
<th>&lt;pop mean&gt;</th>
<th>xbar +</th>
<th>z*SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.63</td>
<td>0.195725281 &lt;pop mean&gt;</td>
<td>0.63</td>
<td>0.195725281</td>
<td></td>
</tr>
<tr>
<td>0.434274719 &lt;pop mean&gt;</td>
<td>0.825725281</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

So you could be 99% confident that the mean number of prior DUI convictions of the population of people arrested in the City and County of Honolulu for DUI in 2001 is somewhere between .43 and .86 convictions.