Confidence Intervals / Estimation with the Population Standard Deviation

THIS PRESENTATION IS BASED ON MATERIAL AND GRAPHS FROM OPEN STAX AND IS COPYRIGHTED BY OPEN STAX AND GEORGIA HIGHLANDS COLLEGE.
Confidence intervals

We use sample data to make generalizations about an unknown population. This part of statistics is called **inferential statistics**.

The sample data help us to make an estimate of a **population parameter**. We realize that the point estimate is most likely not the exact value of the population parameter, but close to it. After calculating point estimates, we construct interval estimates, called confidence intervals.
Point estimate

Means:
You use $\bar{x}$ (sample mean) to estimate the population mean, $\mu$.
The sample mean $\bar{x}$ is called the point estimate for the population mean, $\mu$.

Standard Deviation
You use $s$ (sample standard deviation) to estimate the population standard deviation.
The sample standard deviation $s$ is called point estimate for the population standard deviation, $\sigma$. 
Confidence intervals explained

A **confidence interval** is another type of estimate but, instead of being just one number, it is an interval of numbers. It provides a range of reasonable values in which we expect the population parameter to fall. There is no guarantee that a given confidence interval does capture the parameter, but there is a predictable probability of success.
Example of confidence interval

If you worked in the marketing department of an entertainment company, you might be interested in the mean number of songs a consumer downloads a month from iTunes. If so, you could conduct a survey and calculate the sample mean, $\bar{x}$, and the sample standard deviation, $s$. 
Example of c.i. continued

Suppose, for the iTunes example,
• the population mean $\mu$ is unknown
• population standard deviation is $\sigma = 1$
• our sample size is 100.

The standard deviation for the sample mean is $\frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{100}} = 0.1$
Using the Empirical rule

The **empirical rule**, which applies to bell-shaped distributions, says that in approximately 95% of the samples, the sample mean, $\bar{x}$, will be within two standard deviations of the population mean $\mu$.

For our iTunes example, two standard deviations is $(2)(0.1) = 0.2$. The sample mean, $\bar{x}$, is likely to be within 0.2 units of $\mu$. 
Empirical rule

Because \( \bar{x} \) is within 0.2 units of \( \mu \), which is unknown, then \( \mu \) is likely to be within 0.2 units of \( \bar{x} \) in 95% of the samples. The population mean \( \mu \) is contained in an interval whose lower number is calculated by taking the sample mean and subtracting two standard deviations \((2)(0.1)\) and whose upper number is calculated by taking the sample mean and adding two standard deviations.

In other words, \( \mu \) is between \( \bar{x} - 0.2 \) and \( \bar{x} + 0.2 \) in 95% of all the samples.
Example of confidence interval cont.

For the iTunes example, suppose that a sample produced a sample mean $\bar{x}$ is 2.

Then the unknown population mean $\mu$ is between

$\bar{x} - 0.2 = \bar{x} + 0.2 =$

$\begin{align*}
2 - 0.2 &= 1.8 \\
2 + 0.2 &= 2.2
\end{align*}$

We say that we are 95% confident that the unknown population mean number of songs downloaded from iTunes per month is between 1.8 and 2.2. The 95% confidence interval is $(1.8, 2.2)$. 
Confidence interval

The 95% confidence interval implies two possibilities.

1. Either the interval (1.8, 2.2) contains the true mean $\mu$ or our sample produced an $\bar{x}$ that is not within 0.2 units of the true mean $\mu$.

2. The second possibility happens for only 5% of all the samples (95–100%).

Remember that a confidence interval is created for an unknown population parameter like the population mean, $\mu$. Confidence intervals for some parameters have the form: (point estimate – margin of error, point estimate + margin of error)
Calculating the Confidence Interval with a known population standard deviation

Always check to see if the following two statements are true

(It will be noted in the problems).

• There is a simple random sample
• The sample size is $\geq 30$ or the population is normally distributed

The problem will give you:
  • the population mean
  • the population standard deviation
  • the sample size.
Margin of error

Remember that a confidence interval is created for an unknown population parameter like the population mean, $\mu$. Confidence intervals for some parameters have the form:

$$(\text{point estimate} - \text{margin of error}, \text{point estimate} + \text{margin of error})$$

The margin of error depends on the confidence level or percentage of confidence and the standard error of the mean.
A Single Population Mean using the Normal Distribution

A confidence interval for a population mean with a known standard deviation is based on the fact that the sample means follow an approximately normal distribution.

Suppose that our sample has a mean of $\bar{x} = 10$ and we have constructed the 90% confidence interval (5, 15) where $EBM = 5$. 
A single population mean using the normal distribution (where the population standard deviation is known)

To construct a confidence interval for a single unknown population mean \( \mu \), where the population standard deviation is known, we need \( \bar{x} \) as an estimate for \( \mu \) and we need the margin of error.

Here, the margin of error (\( EBM \)) is called the error bound for a population mean. The sample mean \( \bar{x} \) is the point estimate of the unknown population mean \( \mu \).
Confidence interval estimate

The confidence interval estimate will have the form:

(point estimate - error bound, point estimate + error bound)
or, in symbols, (\(\bar{x} - EBM, \bar{x} + EBM\))
Example of a 95% confidence interval

Suppose we have collected data from a sample. We know the sample mean but we do not know the mean for the entire population. The sample mean is seven, and the error bound for the mean is 2.5.

\( \bar{x} = 7 \) and \( EBM = 2.5 \)
The confidence interval is \((7 - 2.5, 7 + 2.5)\), calculating the values gives \((4.5, 9.5)\).

If the confidence level \( CL \) is 95%, then we say that, "We estimate with 95% confidence that the true value of the population mean is between 4.5 and 9.5."
There is another probability called alpha (\(\alpha\)). \(\alpha\) is related to the confidence level, \(CL\). \(\alpha\) is the probability that the interval does not contain the unknown population parameter.

Mathematically, \(\alpha + CL = 1\).

If your confidence level is 90%, then the \(\alpha\) is 10% or 0.10.

If your confidence level is 95%, then the \(\alpha\) is 5% or 0.05.
Two tails

A confidence interval for a population mean with a known standard deviation is based on the fact that the sample means follow an approximately normal distribution.
Two tails for a 90% confidence interval

Suppose that our sample has a mean of $\bar{x} = 10$, and we have constructed the 90% confidence interval $(5, 15)$ where $EBM = 5$.

To get a 90% confidence interval, we must include the central 90% of the probability of the normal distribution.

If we include the central 90%, we leave out a total of $\alpha = 10\%$ in both tails, or 5% in each tail, of the normal distribution.
What it looks like

\[ \bar{x} = 10 \]
\[ EBM = 5 \]
\[ \bar{x} - EBM = 5 \]
\[ \bar{x} + EBM = 15 \]

Confidence Level (CL) = 0.90
Using normal distribution

In summary, as a result of the central limit theorem:

• $\bar{X}$ is normally distributed, that is, $\bar{X} \sim N(\mu_x, \frac{\sigma_x}{\sqrt{n}})$

When the population standard deviation $\sigma$ is known, we use a normal distribution to calculate the error bound.
Examples of confidence levels and z-scores

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Z-scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.70</td>
<td>1.04</td>
</tr>
<tr>
<td>0.75</td>
<td>1.15</td>
</tr>
<tr>
<td>0.80</td>
<td>1.28</td>
</tr>
<tr>
<td>0.85</td>
<td>1.44</td>
</tr>
<tr>
<td>0.90</td>
<td>1.645</td>
</tr>
<tr>
<td>0.92</td>
<td>1.75</td>
</tr>
<tr>
<td>0.95</td>
<td>1.96</td>
</tr>
<tr>
<td>0.96</td>
<td>2.05</td>
</tr>
<tr>
<td>0.98</td>
<td>2.33</td>
</tr>
<tr>
<td>0.99</td>
<td>2.58</td>
</tr>
</tbody>
</table>
Error bound of population mean (EBM)

The error bound formula for an unknown population mean $\mu$ when the population standard deviation $\sigma$ is known is

$$EBM = \left(\frac{z_{\alpha/2}}{2}\right)\left(\frac{\sigma}{\sqrt{n}}\right)$$

• $z_{\alpha/2}$ comes from the table on the slide before

• If you want to be 95% confident, you would use: 1.96
• If you want to be 90% confident, you would use: 1.645
Writing the Interpretation

The interpretation should clearly state the confidence level (CL), explain what population parameter is being estimated (here, a population mean), and state the confidence interval (both endpoints).

"We estimate with ___% confidence that the true population mean (include the context of the problem) is between ___ and ___ (include appropriate units)."
Example of C.I. with population standard deviation

Suppose scores on exams in statistics are normally distributed with an unknown population mean and a population standard deviation of three points. A random sample of 36 scores is taken and gives a sample mean (sample mean score) of 69. Find a confidence interval estimate for the population mean exam score (the mean score on all exams).

Find a 95% confidence interval for the true (population) mean of statistics exam scores.

\[ \bar{x} = 69, \quad \sigma = 3 \]

\[ n = 36, \quad CL = 95\% \]
Example of C.I. with population standard deviation continued

Find a 95% confidence interval for the true (population) mean of statistics exam scores. Using the chart, the z-value for 95% is 1.96.

To get EBM, 

\[ (z_{\alpha/2}) \left( \frac{\sigma}{\sqrt{n}} \right) \]

EBM = \( (1.96) \left( \frac{3}{\sqrt{36}} \right) = 0.98 \)

\[ \bar{x} - EBM = 69 - 0.98 = 68.02 \]
\[ \bar{x} + EBM = 69 + 0.98 = 69.98 \]

The 95% confidence interval is \( (68.02, 69.98) \).
Confidence Z-interval in Calculator

If no data is given,

Enter STAT, Arrow over to TESTS, chose ZInterval

1. The next screen will show the following for entering information:
2. Inpt: Stats (choose if you are not entering data)
3. \( \sigma \): Population standard deviation
4. \( \bar{x} \): sample mean
5. \( n \): sample size
6. C-Level: confidence level (what interval)
7. Scroll to Calculate and it will go to a new screen

The output screen is:

\[ Z\text{interval} \]

(upper bound, lower bound) of confidence interval

\( \bar{x} \): sample mean

\( n \): sample size
Input into calculator
Output from calculator
WEBSITE INSTRUCTIONS
(If only sample mean is given)

WWW.SOCSCISTATISTICS.com
Instructions

Go to Calculators, A Single-Sample Confidence Interval Calculator Using the Z Statistic (It is near the bottom of the page)

Z-test If no data is given,

Enter the following:
   Sample mean ($M$)
   Sample size ($n$)
   Standard Deviation ($s$): (enter population standard deviation)
   Confidence Level: pick the correct one

You will be given the confidence interval as well as the math behind it.
The Calculation

Please enter your data into the fields below, select a confidence level (the calculator defaults to 95%), and then hit Calculate. Your result will appear at the bottom of the page.

Sample Mean $(M)$: 69
Sample Size $(n)$: 36
Standard Deviation $(s)$: 3
Confidence Level: 95% ▼

Please enter your values above, and then hit the calculate button.

Calculate
Result

$M = 69$, 95% CI $[68.02, 69.98]$. You can be 95% confident that the population mean ($\mu$) falls between 68.02 and 69.98.

Calculation

$M = 69$

$t = 1.96$

$s_M = \sqrt{3^2/36} = 0.5$

$\mu = M \pm Z(s_M)$

$\mu = 69 \pm 1.96 \times 0.5$

$\mu = 69 \pm 0.98$

Output from the website
Working Backwards to Find the Error Bound

When we calculate a confidence interval, we find the sample mean, calculate the error bound, and use them to calculate the confidence interval. However, sometimes when we read statistical studies, the study may state the confidence interval only. If we know the confidence interval, we can work backwards to find both the error bound and the sample mean.

Finding the Error Bound

From the upper value for the interval, subtract the sample mean,

OR, from the upper value for the interval, subtract the lower value. Then divide the difference by two.
Calculate the error bound

Suppose we know that a confidence interval is *(67.18, 68.82)* and we want to find the error bound. We may know that the sample mean is 68, or perhaps our source only gave the confidence interval and did not tell us the value of the sample mean.

**Calculate the Error Bound:**

If we know that the sample mean is 68: \( EBM = 68.82 - 68 = 0.82 \).

OR

If we don't know the sample mean: \( EBM = \frac{(68.82 - 67.18)}{2} = 0.82 \)
Working Backwards to Find the Sample Mean

**Finding the Sample Mean**

Subtract the error bound from the upper value of the confidence interval,

OR, average the upper and lower endpoints of the confidence interval.
Calculate the sample mean

Suppose we know that a confidence interval is \((67.18, 68.82)\) and we want to find the error bound. We may know that the sample mean is 68, or perhaps our source only gave the confidence interval and did not tell us the value of the sample mean.

**Calculate the Sample Mean:**

If we know the error bound: \(\bar{x} = 68.82 - 0.82 = 68\)

OR

If we don't know the error bound: \(\bar{x} = \frac{(68.82+67.18)}{2} = 68\)
Example to find EBM and sample mean

Suppose we know that a confidence interval is (42.12, 47.88). Find the error bound and the sample mean.

Since we don't know the sample mean: \( EBM = \frac{(47.88 - 42.12)}{2} = 2.88 \)

Since we don't know the error bound: \( \bar{x} = \frac{(47.88 + 42.12)}{2} = 45 \)
Example 1

The standard deviation of the weights of elephants is known to be approximately 15 pounds. We wish to construct a 95% confidence interval for the mean weight of newborn elephant calves. Fifty newborn elephants are weighed. The sample mean is 244 pounds. The sample standard deviation is 11 pounds.

Find the following information:

a. Sample mean
b. Population standard deviation
c. Sample size
d. Confidence interval
Example 1 (a)

The standard deviation of the weights of elephants is known to be approximately 15 pounds. We wish to construct a 95% confidence interval for the mean weight of newborn elephant calves. Fifty newborn elephants are weighed. The sample mean is 244 pounds. The sample standard deviation is 11 pounds.

Find the following information:

a. Sample mean = 244 pounds
b. Population standard deviation = 15 pounds
c. Sample size = 50 elephants
d. Confidence interval = 95%
Example 1 (b)

The standard deviation of the weights of elephants is known to be approximately 15 pounds. We wish to construct a 95% confidence interval for the mean weight of newborn elephant calves. Fifty newborn elephants are weighed. The sample mean is 244 pounds. The sample standard deviation is 11 pounds.

Using the calculator or website, I get the following:
The lower bound is 239.84
The upper bound is 248.16

We estimate with 95% confidence that the true population mean of newborn elephant weights is between 239.84 and 248.16 pounds.
Example 1 (c)

The standard deviation of the weights of elephants is known to be approximately 15 pounds. We wish to construct a 95% confidence interval for the mean weight of newborn elephant calves. Fifty newborn elephants are weighed. The sample mean is 244 pounds. The sample standard deviation is 11 pounds.

Find the error bound margin:
The lower bound is 239.84
The upper bound is 248.16

To find the error bound margin, 244 - 239.84 = 4.16
or 248.16 – 244 = 4.16
The error bound margin is 4.16