Confidence Intervals about a Population Mean
MATH 130, Elements of Statistics I

J. Robert Buchanan
Department of Mathematics

Fall 2017
Motivation

**Goal:** to estimate a population mean $\mu$ based on data collected in a sample.

**Assumption:** the population standard deviation $\sigma$ is known. This is not strictly required, but simplifies the steps involved.

Remark: a fair question to ask is “How often will we know $\sigma$, but not $\mu$?” It is more likely that we will know $x$ and $s$ but not know $\mu$ or $\sigma$. 
Motivation

**Goal:** to estimate a population mean $\mu$ based on data collected in a sample.

**Assumption:** the population standard deviation $\sigma$ is known. This is not strictly required, but simplifies the steps involved.

**Remark:** a fair question to ask is “How often will we know $\sigma$, but not $\mu$?” It is more likely that we will know $\bar{x}$ and $s$ but not know $\mu$ or $\sigma$. 
A point estimate is the value of a statistic (taken from a sample) that estimates the value of a parameter (for a population).
Point Estimates

Definition
A point estimate is the value of a statistic (taken from a sample) that estimates the value of a parameter (for a population).

Example

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Point Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>$\bar{X}$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>$S$</td>
</tr>
</tbody>
</table>
Example

The mean length of machine parts manufactured by a certain factory is to be estimated by taking a sample of ten machine parts. The lengths in millimeters are as follows.

\[
75.3 \quad 76.0 \quad 75.0 \quad 77.0 \quad 75.4 \\
76.3 \quad 77.0 \quad 74.9 \quad 76.5 \quad 75.8
\]

The sample mean is $\bar{x} = 75.92$. Thus the point estimate of $\mu$ is 75.92.
Relationship Between the Point Estimate and $\mu$

If the point estimate is close to $\mu$, we are interested in understanding how close. This brings up the notions of

- margin of error,
- confidence interval, and
- level of confidence.
Experiment

Suppose the mean age of a sample of college students is 20.1 years.

1. Into what interval would you be willing to place the population mean age of college students with 90% confidence?

   $20.1 \pm \ ?. $

2. Into what interval would you be willing to place the population mean age of college students with 95% confidence?

3. Into what interval would you be willing to place the population mean age of college students with 99% confidence?
Experiment

Suppose the mean age of a sample of college students is 20.1 years.

1. Into what interval would you be willing to place the population mean age of college students with 90% confidence?

   \[ 20.1 \pm ? \]

2. Into what interval would you be willing to place the population mean age of college students with 95% confidence?

   \[ 20.1 \pm ? \]
Experiment

Suppose the mean age of a sample of college students is 20.1 years.

1. Into what interval would you be willing to place the population mean age of college students with 90% confidence?
   
   \[ 20.1 \pm ? \]

2. Into what interval would you be willing to place the population mean age of college students with 95% confidence?
   
   \[ 20.1 \pm ? \]

3. Into what interval would you be willing to place the population mean age of college students with 99% confidence?
   
   \[ 20.1 \pm ? \]
Confidence Intervals and Level of Confidence

Definition

- A **confidence interval** for an unknown parameter consists of an interval of numbers.
- The **level of confidence** represents the expected proportion of intervals that will contain the parameter if a large number of different samples is obtained. The level of confidence is denoted

\[(1 - \alpha) \cdot 100\%.
\]
Margin of Error

We will express confidence intervals in the form:

point estimate ± margin of error

where the margin of error depends on

1. the level of confidence (as the level of confidence increases so does the margin of error),
2. the sample size (as the sample size increases the margin of error decreases),
3. the population standard deviation (the greater the spread in the population characteristic, the larger the margin of error).
Margin of Error

We will express confidence intervals in the form:

point estimate ± margin of error

where the margin of error depends on

1. the level of confidence (as the level of confidence increases so does the margin of error),
2. the sample size (as the sample size increases the margin of error decreases),
3. the population standard deviation (the greater the spread in the population characteristic, the larger the margin of error).

\[
\text{margin of error } = E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}
\]
Example

Suppose the mean age of a sample of 35 college students is 20.1 years. The population standard deviation is 4.0 years.

1. Construct the 90% confidence interval (90% CI) for the population mean age of college students.

2. Construct the 95% confidence interval (95% CI) for the population mean age of college students.
90% Confidence Interval

\[
90\% \text{ CI} = (1 - 0.10) \cdot 100\% \text{ CI} \implies \alpha = 0.10
\]
\[
\alpha/2 = 0.05
\]
\[
z_{\alpha/2} = 1.645
\]

Thus the 90% confidence interval estimate of \( \mu \) (the population mean age of college students) is
\[
x \pm E \iff 20.1 \pm 1.1 \iff (19.0, 21.2)
\]
90% Confidence Interval

90% CI = (1 - 0.10) \cdot 100\% \text{ CI} \implies \alpha = 0.10

\alpha/2 = 0.05

z_{\alpha/2} = 1.645

E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1.645 \cdot \frac{4.0}{\sqrt{35}} = 1.1

Thus the 90% confidence interval estimate of \( \mu \) (the population mean age of college students) is \( x \pm E \Rightarrow 20.1 \pm 1.1 \Rightarrow (19.0, 21.2) \).
90% Confidence Interval

\[ 90\% \text{ CI} = (1 - 0.10) \cdot 100\% \text{ CI} \implies \alpha = 0.10 \]
\[ \alpha/2 = 0.05 \]
\[ z_{\alpha/2} = 1.645 \]
\[ E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1.645 \cdot \frac{4.0}{\sqrt{35}} = 1.1 \]

Thus the 90\% confidence interval estimate of \( \mu \) (the population mean age of college students) is

\[ \bar{x} \pm E \iff 20.1 \pm 1.1 \iff (19.0, 21.2). \]
95% Confidence Interval

\[ 95\% \text{ CI} = (1 - 0.05) \cdot 100\% \text{ CI} \implies \alpha = 0.05 \]

\[ \frac{\alpha}{2} = 0.025 \]

\[ z_{\alpha/2} = 1.96 \]

Thus the 95% confidence interval estimate of \( \mu \) (the population mean age of college students) is \( x \pm E \iff 20 \pm 1.3 \iff (18.8, 21.4) \).

Remark: with greater level of confidence we must use a wider confidence interval.
95% Confidence Interval

\[
95\% \text{ CI} = (1 - 0.05) \cdot 100\% \text{ CI} \Rightarrow \alpha = 0.05
\]

\[
\alpha/2 = 0.025
\]

\[
z_{\alpha/2} = 1.96
\]

\[
E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{4.0}{\sqrt{35}} = 1.3
\]

Thus the 95% confidence interval estimate of \( \mu \) (the population mean age of college students) is 

\[
x \pm E \Rightarrow 20.1 \pm 1.3 \Rightarrow (18.8, 21.4)
\]

Remark: with greater level of confidence we must use a wider confidence interval.
95% Confidence Interval

\[
95\% \text{ CI} = (1 - 0.05) \cdot 100\% \text{ CI} \implies \alpha = 0.05
\]

\[
\alpha/2 = 0.025
\]

\[
z_{\alpha/2} = 1.96
\]

\[
E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{4.0}{\sqrt{35}} = 1.3
\]

Thus the 95% confidence interval estimate of \(\mu\) (the population mean age of college students) is

\[
\bar{x} \pm E \iff 20.1 \pm 1.3 \iff (18.8, 21.4).
\]

Remark: with greater level of confidence we must use a wider confidence interval.
95% Confidence Interval

\[
\text{95\% CI} = (1 - 0.05) \cdot 100\% \text{ CI} \implies \alpha = 0.05
\]
\[
\alpha/2 = 0.025
\]
\[
z_{\alpha/2} = 1.96
\]
\[
E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{4.0}{\sqrt{35}} = 1.3
\]

Thus the 95\% confidence interval estimate of \( \mu \) (the population mean age of college students) is

\[
\bar{x} \pm E \iff 20.1 \pm 1.3 \iff (18.8, 21.4).
\]

**Remark:** with greater level of confidence we must use a wider confidence interval.
Interpretation of a Confidence Interval

If the 95% confidence interval estimate for a population mean is \( \bar{x} \pm E \)

it does not mean there is a 95% chance that \( \bar{x} - E < \mu < \bar{x} + E \).

It does mean that if we took a large number of samples (all of the same size) and constructed a 95% confidence interval from each sample, then \( \mu \) would lie in 95% of the confidence intervals.
Simulation

100 samples of size 35.
Example

The Third International Mathematics and Science Study (TIMSS) in 1999 examined eighth-graders’ proficiency in math and science. The mean geometry score for a sample of 25 eighth-grade students was 47.3 with a standard deviation of 4.4. Construct the 95% confidence interval for the mean geometry score for all eighth-grade students in the United States.
95% Confidence Interval

95% CI = (1 - 0.05) \cdot 100\% CI \implies \alpha = 0.05

\alpha/2 = 0.025

z_{\alpha/2} = 1.96

Thus the 95\% confidence interval estimate of \( \mu \) (the population mean geometry score of eight-grade students) is \( x \pm E \iff 47.3 \pm 1.7 \iff (45.6, 49.0) \).
95% Confidence Interval

\[
95\% \text{ CI} \quad = \quad (1 - 0.05) \cdot 100\% \text{ CI} \quad \implies \quad \alpha = 0.05
\]
\[
\alpha/2 \quad = \quad 0.025
\]
\[
z_{\alpha/2} \quad = \quad 1.96
\]
\[
E \quad = \quad z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad = \quad 1.96 \cdot \frac{4.4}{\sqrt{25}} \quad = \quad 1.7
\]

Thus the 95% confidence interval estimate of \( \mu \) (the population mean geometry score of eight-grade students) is \( x \pm E \iff 47.3 \pm 1.7 \iff (45.6, 49.0) \).
95% Confidence Interval

95% CI = (1 − 0.05) · 100% CI \implies \alpha = 0.05
\alpha/2 = 0.025
z_{\alpha/2} = 1.96

E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{4.4}{\sqrt{25}} = 1.7

Thus the 95% confidence interval estimate of \mu (the population mean geometry score of eight-grade students) is

\bar{x} \pm E \iff 47.3 \pm 1.7 \iff (45.6, 49.0).
Determining Sample Size

If the confidence level, margin of error, and population standard deviation are known, we can estimate the sample size necessary to create the confidence interval.

\[
E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}
\]

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

If \( n \) is not a whole number, we must round up to the nearest whole number.
Example

The weights of second graders is normally distributed with a standard deviation of 3 pounds.

1. Determine the sample size needed to estimate the population mean weight of all second graders within a margin of error of 1 pound at the 90% confidence level.

2. Determine the sample size needed to estimate the population mean weight of all second graders within a margin of error of 1 pound at the 99% confidence level.
Solution

1. 90% CI implies $z_{\alpha/2} = 1.645$, thus

$$n = \left( \frac{Z_{\alpha/2} \cdot \sigma}{E} \right)^2 = \left( \frac{(1.645)(3)}{1} \right)^2 = 24.354225 \approx 25.$$
1. 90% CI implies \( z_{\alpha/2} = 1.645 \), thus

\[
n = \left( \frac{Z_{\alpha/2} \cdot \sigma}{E} \right)^2 = \left( \frac{(1.645)(3)}{1} \right)^2 = 24.354225 \approx 25.
\]

2. 99% CI implies \( z_{\alpha/2} = 2.575 \), thus

\[
n = \left( \frac{Z_{\alpha/2} \cdot \sigma}{E} \right)^2 = \left( \frac{(2.575)(3)}{1} \right)^2 = 59.675625 \approx 60.
\]
Solution

1. 90% CI implies $z_{\alpha/2} = 1.645$, thus

$$n = \left( \frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2 = \left( \frac{(1.645)(3)}{1} \right)^2 = 24.354225 \approx 25.$$

2. 99% CI implies $z_{\alpha/2} = 2.575$, thus

$$n = \left( \frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2 = \left( \frac{(2.575)(3)}{1} \right)^2 = 59.675625 \approx 60.$$

Remark: keeping the margin of error the same but increasing the level of confidence requires a larger sample.
Homework

- Read Section 9.1.
- Exercises: 13–21 odd, 25, 29, 33, 37, 41