Statistical Inference Procedures

- Confidence Intervals
- Hypothesis Tests
Statistical inference produces answers to specific questions about the population of interest based on the information in a sample. Inference procedures must include a statement of how confident we can be that the answer is correct.
This presentation considers only confidence intervals for $\mu$.

The basis for this important topic is the sampling distribution of $\bar{X}$. 
Confidence Interval

- A confidence interval is a formula that tells us how to use sample data to estimate a population parameter.

- The confidence level indicates the percentage of intervals constructed in this manner which contain the value of the parameter.
Confidence Interval for $\mu$

- Since $\bar{x}$ is an **unbiased estimator** of the population mean, $\mu$, we use $\bar{x}$ to estimate $\mu$.

- The sampling distribution of $\bar{x}$ provides the tools we need to build a confidence interval for $\mu$. 
A $100(1-\alpha)\%$ Confidence Interval for $\mu$

- $(1-\alpha)$ is the confidence coefficient. It is the probability that the interval estimator encloses the population parameter.

- If the confidence coefficient is $(1-\alpha)$, then $100(1-\alpha)\%$ is the confidence level. The confidence level is the percentage of the intervals constructed by the formula will contain the true value of $\mu$. 
A 100(1−α)% Confidence Interval for μ

- Assume the population standard deviation $\sigma$ is known.

- Assume that $X$ is normally distributed with mean, $\mu$, and standard deviation, $\sigma$.

- For samples of size $n$, $\bar{X}$ is normally distributed with mean, $\mu$, and standard deviation, $\frac{\sigma}{\sqrt{n}}$. 
Let’s Use What We Know

\[ P \left( \mu - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = P \left( -z_{\alpha/2} < z < z_{\alpha/2} \right) \]

What is \( z_{\alpha/2} \)?

\( z_{\alpha/2} \) is the value of \( z \) that answers the question what is the value of \( z \) such that 100\((1-\alpha)\)% of the values of \( z \) lie between \(-z_{\alpha/2}\) and \( z_{\alpha/2} \).

If 100\((1-\alpha) = 95\), then \( P(- z_{\alpha/2} < z < z_{\alpha/2} ) = .95. \)

Therefore, \( z_{\alpha/2} = 1.96. \)
What is the formula for a 100(1-\(\alpha\))% confidence interval for \(\mu\)?

Using algebra which you are not responsible for doing, I can take the expression below and change it from a probability statement about \(\bar{X}\) to one about \(\mu\).

\[
P\left( \mu - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) =
\]

\[
P\left( \bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)
\]

In this expression we are talking about the variable \(\bar{X}\), not an observed value of the variable.
The formula for a C\% confidence interval for \( \mu \) is

\[
\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}
\]

When we substitute the value of the sample mean for a particular sample, we can no longer talk about the probability. Instead we say that we are 100(1−\( \alpha \))\% sure that the true mean \( \mu \) lies between the two values we obtained by using the sample mean.
An Example

We want to estimate the mean time required to perform a task with 95% confidence. We select a random sample of 25 subjects and find that $\bar{x} = 43$ seconds. Assume that $\sigma = 2.5$ seconds. We must also assume that the time required to perform this task is normally distributed.
Give a 95% Confidence Interval for \( \mu \)

\[
\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}
\]

\[
43 - 1.96 \frac{2.5}{\sqrt{25}} < \mu < 43 + 1.96 \frac{2.5}{\sqrt{25}}
\]

\[
42.02 < \mu < 43.98
\]

We are 95% confident that the mean time required to perform this task is between 42.02 and 43.98 seconds.
This Formula is usually written

\[ \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \]

This formula is in the form we use for many confidence intervals, estimate \( \pm \) margin of error.
Give a 95% CI for $\mu$

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$43 \pm 1.96 \frac{2.5}{\sqrt{25}}$$

$$43 \pm .98$$

$$42.02 < \mu < 43.98$$

This is the way I want you to show your work.
Interpret the Confidence Interval

We are 95% confident that the mean time to perform this task is between 42.02 and 43.98 seconds.
Properties of the Margin of Error

- All other things being equal, the margin of error of a confidence interval decreases as
  - the confidence level $100(1-\alpha)$ decreases.
  - the sample size $n$ increases.
  - the population standard deviation $\sigma$ decreases.
What happens when we do not know the population’s standard deviation?

- We use the sample standard deviation as a substitute for \( \sigma \).
- In this case the formula to be used depends on the sample size, \( n \).
- If \( n < 30 \), we need to know the properties of a new sample statistic called \( t \).

\[
t = \frac{\bar{x} - \mu}{s} \cdot \frac{1}{\sqrt{n}}
\]
Compare the z-score for $\overline{x}$ to the t-score for $\overline{x}$.

\[
z = \frac{\overline{x} - \mu}{\sigma \sqrt{n}} \quad \text{and} \quad t = \frac{\overline{x} - \mu}{s \sqrt{n}}
\]
The t Distributions

- The t distributions are a family of distributions one for each number of degrees of freedom. We use df to denote the number of degrees of freedom.
- Each t distribution is symmetric about 0.
- A table giving values of $t_\alpha$ for various probabilities and numbers of degrees of freedom is given in the front cover of your book.
- The table entries are $t_\alpha$ such that $P(t > t_\alpha) = \alpha$.
- Example: If $\alpha = .025$ and df = 14, $t_\alpha = 2.145$. 
The formula for a $100(1-\alpha)\%$ confidence interval for $\mu$ when $\sigma$ is not known and $n < 30$ is

$$
\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}}
$$

$$
df = n - 1
$$
This Formula is usually written

\[ \bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}, \quad df = n - 1 \]

This formula is in the form we use for many confidence intervals,

estimate \( \pm \) margin of error.
An Example using t

We want to estimate the mean time required to perform a task with 95% confidence. We select an SRS of 25 subjects and find that $\bar{X} = 43$ seconds and $s = 2.5$ seconds. We must assume that the time required to perform this task is normally distributed.
Give a 95% CI for $\mu$

$$\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2} \frac{t}{\sqrt{n}}$$

$$43 - 2.064 \frac{2.5}{\sqrt{25}} < \mu < 43 + 2.064 \frac{2.5}{\sqrt{25}}$$

$$41.968 < \mu < 44.032$$

We are 95% confident that the mean time required to perform this task is between 41.968 and 44.032 seconds.
Give a 95% CI for $\mu$

$$
\bar{x} \pm t_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \quad df = n - 1 = 24
$$

$$
43 \pm 2.064 \frac{2.5}{\sqrt{25}}
$$

$$
43 \pm 1.032
$$

$$
41.968 < \mu < 44.032
$$

This is the way I want you to show your work.
What happens when X is not normally distributed?
The Central Limit Theorem provides a guide.

- If \( n \) is large enough the sampling distribution of \( \bar{x} \) is approximately normally distributed.
- Remember, we claim that \( n \) is large enough if \( n \) is at least 30.
- If we knew \( \sigma \) then we could use the formula

\[
\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}.
\]
What if we do not know $\sigma$ and $n$ is at least 30?

- We simply use the formula

$$
\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}
$$

and substitute $s$ for $\sigma$. 
Summary

• There are 2 formulas used to create a confidence interval for $\mu$.
• How do we choose which formula to use?
• We ask the following questions:
  – Is $X$ normally distributed?
  – Is $\sigma$ known?
  – What is $n$?
If $X$ is normally distributed and $\sigma$ is known, then for every $n$ use

$$
\overline{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}
$$
If $X$ is normally distributed and $\sigma$ is not known, then for $n < 30$

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}, \quad df = n - 1$$
If $X$ is normally distributed and $\sigma$ is not known, then for $n \geq 30$ use

$$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$
If $X$ is not normally distributed and $\sigma$ is not known, then for $n$ greater than or equal to 30 use

$$
\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}
$$
If \( X \) is not normally distributed and \( \sigma \) is not known, then for \( n \) less than 30 neither of these formulas can be used.