Quiz scores are on eCommons
- email me if your score is not there.

Quizzes returned in section.

Midterm next Tuesday
- 1 page of notes (double sided)
- bring a calculator

Practice midterm on course website
(not eCommons)

this is a guide to the style of the midterm.

---

Today's topic: Confidence Intervals.
Confidence Intervals.

Population

unknown $\mu$
unknown $\sigma$

Sample

\[ \bar{x} \]
\[ s \]
\[ n \]

use these to estimate a range of plausible values for $\mu$.

How confident are we in our estimate?

Best estimate of $\mu$ is $\bar{x}$ - Point estimate.

Quantify the uncertainty in the estimate.
Before we measure any data:

\[ P(-1.96 \leq z \leq 1.96) = 0.95 \]

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

\[ z = \frac{\bar{x} - \mu_x}{\sigma_{\bar{x}}} \]

\[ P\left(-1.96 \leq \frac{\bar{x} - \mu_x}{\sigma_{\bar{x}}} \leq 1.96\right) = 0.95 \]

\[ P\left(-1.96 \sigma_{\bar{x}} \leq \bar{x} - \mu_x \leq 1.96 \sigma_{\bar{x}}\right) = 0.95 \]

\[ P\left(-1.96 \frac{\sigma_{\bar{x}}}{\sqrt{n}} \leq \bar{x} - \mu_x \leq 1.96 \frac{\sigma_{\bar{x}}}{\sqrt{n}}\right) = 0.95 \]

\[ P\left(1.96 \frac{\sigma_{\bar{x}}}{\sqrt{n}} \leq \mu_x - \bar{x} \leq -1.96 \frac{\sigma_{\bar{x}}}{\sqrt{n}}\right) = 0.95 \]

\[ P\left(\frac{\bar{x} + 1.96 \sigma_{\bar{x}}}{\sqrt{n}} \leq \mu_x \leq \frac{\bar{x} - 1.96 \sigma_{\bar{x}}}{\sqrt{n}}\right) = 0.95 \]

\[ P\left(\frac{\bar{x} - 1.96 \sigma_{\bar{x}}}{\sqrt{n}} \leq \mu_x \leq \frac{\bar{x} + 1.96 \sigma_{\bar{x}}}{\sqrt{n}}\right) = 0.95 \]
We have defined a random interval
\[
\bar{x} - 1.96 \frac{s}{\sqrt{n}} \text{ to } \bar{x} + 1.96 \frac{s}{\sqrt{n}}
\]
\[95\% \text{ confidence interval}\]

This interval will contain the true population mean with probability 0.95.

95\% of intervals constructed this way will contain the true population mean.

Measure \( \bar{x} \) from a sample.

Construct the corresponding interval \( \bar{x} \pm z_{\alpha/2} \).

What's the probability that the population mean, \( \mu \), lies in this interval?

\[\text{Unknown parameter} - \text{takes a fixed value.}\]

This value is either within the interval or it is not, but we don't know which.
What does the 95% mean?

If we generate lots of samples, and construct lots of confidence intervals (one for each sample), then 95% of the confidence intervals will cover the true population mean.

Example. Sample (bag) of 32 cookies.

\[
\bar{x} = 10.9 \text{ g} \\
\sigma = 0.5 \text{ g} \quad \text{population standard deviation.}
\]

95% CI for the mean weight of all cookies.

\[
\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} \quad \text{to} \quad \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}
\]

\[
10.9 - 1.96 \times 0.5 \times \frac{1}{\sqrt{32}} \quad \text{to} \quad 10.9 + 1.96 \times 0.5 \times \frac{1}{\sqrt{32}}
\]

10.73 \quad \text{to} \quad 11.07.

(10.73, 11.07) is a 95% confidence interval for \( \mu \).
"We are 95% confident that \( \mu \) is in this interval."

- On average, 95% of the intervals constructed this way will contain \( \mu \). ["covers" \( \mu \)]

- We don't know if a particular interval covers or not.

- CI is a range of plausible values for \( \mu \).

---

\[
(1 - \alpha) \text{ CI.} \quad \bar{x} \pm E = \frac{Z_{\alpha/2}}{\sqrt{n}} \times s
\]

---

In general:

\[
Z_{\alpha/2}
\]

---

\[
\frac{\alpha}{2} \quad \frac{\alpha}{2}
\]

---

\[
\text{P}(Z \leq -Z_{\alpha/2}) = \frac{\alpha}{2}
\]
Example.

A manufacturer claims that their bottles of soda contain 20 oz. The standard deviation is known to be 0.03 oz. Is the right amount of soda going into each bottle?

A sample of 34 bottles was taken, and had a mean content of 19.98 oz.

Find a 90% CI for the population mean.

\[ 1 - \alpha = 0.9 \quad \alpha = 0.1 \]

\[ P\left( \frac{z}{\bar{X}} \right) = 0.05 \]

\[ P\left( z < \frac{-1.645}{\bar{X}} \right) = 0.05 \]

\[ E = \frac{z_\alpha}{\bar{X}} \times \frac{\sigma}{\sqrt{n}} = \frac{1.645 \times 0.03}{\sqrt{34}} = 0.0085 \]

90% CI is \[ 19.98 \pm 0.0085 \]

(19.9715, 19.9885)

This CI does not cover the claimed value of 20 oz.
How do we determine the sample size needed for a desired margin of error?

Example:

Suppose we want to estimate soda contents with a margin of error, $E$, of 0.001.

$$0.001 = E = \frac{Z_{\alpha/2} \cdot \sigma}{\sqrt{n}}$$

For 90% confidence level, $Z_{\alpha/2} = 1.645$

$$0.001 = 1.645 \times \frac{0.03}{\sqrt{n}}$$

$$\sqrt{n} = \frac{1.645 \times 0.03}{0.001} = 49.35$$

$$n = (49.35)^2 = 2435.4$$

need to round up, so that the margin of error is no larger than specified, $n = 2436$.

Note: the sample size increases rapidly as the margin of error reduces.

In general $n = \left( \frac{Z_{\alpha/2} \cdot \sigma}{E} \right)^2$. 
CIs for proportions.

- Based on Normal approximation to the Binomial distribution.

Requirements: Simple random sample.

- Data are Binomial. (4 conditions).

  CLT applies.

  - At least 5 successes: \( np \geq 5 \)
  - At least 5 failures: \( np(1-p) \geq 5 \)

\[ P = \frac{x}{n} \]

\[ \hat{P} = \frac{x}{n} \]

\[ \bar{x} = \frac{x}{n} \]

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

\[ \sigma = \sqrt{np(1-p)} \]

\[ \frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sqrt{p(1-p)} \]

\[ (1-\alpha) \text{ CI} \]

\[ \hat{p} = \alpha + E \]

Approximate the population proportion \( p \) by the sample proportion \( \hat{p} \)
Example.

In an opinion poll of 1207 people, 53% of the respondents disapproved of healthcare reform.

Is 53% commensurate with an even split?

\[
\begin{align*}
\text{Population} & : 0:50 \\
\text{Sample} & : 53:47 \\
\text{\( n \)} & = 1207
\end{align*}
\]

is this likely to happen based just on random variation?

→ construct a confidence interval for \( p \) (population proportion) based on the data in the sample.

→ does that CI cover \( p = 0.5 \)?

\[
\hat{p} = 0.53 \\
\text{\( n \)} = 1207
\]

95\% CI \( \alpha = 0.05 \)

\[
E = \frac{Z_{0.025}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \\
= 1.96 \sqrt{\frac{0.53 \times (1-0.53)}{1207}} \\
= 0.028. \quad \text{(Margin of error)}
\]
95% CI is \( \hat{p} \pm E \)

\[
0.53 \pm 0.025 \quad (0.502, 0.558) \quad \text{does not cover} \quad \hat{p} = 0.5
\]

95% confident that more than 50% of the population disapproved of healthcare reform.

Determining the sample size for a given margin of error for a population proportion.

\[
n = \left( \frac{Z_{\alpha/2}}{E} \right)^2 \hat{p} (1 - \hat{p})
\]

\( \hat{p} \) is the estimated sample proportion

If \( \hat{p} \) is unknown, use \( \hat{p} = 0.5 \) - this will give the largest value of \( n \) for any value of \( \hat{p} \), so we can be sure we have \( n \) large enough to give the desired margin of error.

\[
n = \left( \frac{Z_{\alpha/2}}{E} \right)^2 \times 0.5^2
\]
In the example, we had $E = 0.028 \ (\approx 2.8\%)$.

If we want $E = 0.01 \ (\approx 1\%)$,

Assuming $\hat{p}$ unknown,

$$n = \left( \frac{Z_{\alpha/2}}{E} \right)^2 \times 0.5^2$$

$$= \left( \frac{1.96}{0.01} \right)^2 \times 0.5^2 = 9604$$