Confidence Intervals for the Mean & the Proportion

Lecture 20

03/06/2013

Turn on your clickers!

The confidence intervals are for the population parameters (i.e. μ or p) or for the sample statistics (i.e. $\bar{x}$, $\hat{p}$)?

(a) population parameters (i.e. μ or p)

(b) sample statistics (i.e. $\bar{x}$, $\hat{p}$)

Central Limit Theorem

- $X_1, \ldots, X_n$: *simple random sample*
- Large sample size ($n > 30$)

$$\bar{x} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Significance Level: α

- Each confidence level corresponds to a significance level

$$\alpha = 1 \ - \ \text{confidence level}$$

- *Example*: A 68% CI corresponds to significance level equal to

$$\alpha = 1 - 0.68 = 0.32$$

CI for significance level α

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},$$

where α is the significance level and $z_{\alpha/2}$ is the z-score that corresponds to a probability equal to $\alpha/2$.

Margin of Error

CI for significance level α

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},$$

where α is the significance level and $z_{\alpha/2}$ is the z-score that corresponds to a probability equal to $\alpha/2$.

- The Margin of Error is

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

Example: *Vehicle's age*

In a survey of 30 adults, the *average age of a person's primary vehicle* is 5.6 years.

Assuming that the population standard deviation is 0.8 years, find the *best point estimate of the population mean and a 90% confidence interval for the population mea*n.

Example: *Vehicle's age*

- The best point estimate of the population mean is $\bar{x} = 5.6$ years.
- Since we care about a 90% confidence interval

$$\alpha = 1 - 0.9 = 0.1$$

- The following the z-score:

$$z_{\alpha/2} = z_{0.1/2} = z_{0.05} = -1.65$$

$$\begin{aligned} 5.6 - 1.65\frac{0.8}{\sqrt{30}} < \quad & \mu \quad < 5.6 + 1.65\frac{0.8}{\sqrt{30}} \\ 5.6 - 0.24 < \quad & \mu \quad 5.6 + 0.24 \\ 5.36 < \quad & \mu \quad < 5.84 \end{aligned}$$

Steps to Compute a Confidence Interval for μ with known σ

- Step 1: Compute (if necessary) the sample mean of the sample. Sometimes this might be given to you.
- Step 2: Find $\alpha/2$. If you are asked to find a 95% confidence interval, then $\alpha = 1 - 0.95 = 0.05$ and $\alpha/2 = 0.025$.
- Step 3: Find $z_{\alpha/2}$, that is the corresponding z-score from the normal distribution table.
- Step 4: Substitute in the formula

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}.$$

Sample Size Calculation

The minimum sample size needed for an interval estimate with significance level α of the population mean is

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

Example: *What is the sample size?*

The president of a college asks the statistics teacher to estimate the average age of the students in the college.

Assuming that the population standard deviation is 3 years, *how large a sample is necessary in order to obtain a 99% confidence interval for the average age of the students with a margin of error $E = 1$?*

Example: *What is the sample size?*

- $\alpha = 1 - 0.99 = 0.01$
- $z_{\alpha/2} = 2.58$.
- We already know that $E = 1$ and $\sigma = 3$, so substituting in the formula

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2 = \left(\frac{2.58 \cdot 3}{1}\right)^2 = 59.9$$

Confidence Intervals for the population average μ: Unknown σ

- Large sample size (n>30), IID Population
- Small sample size (n<30), IID Normal Population

Large sample size, IID Population

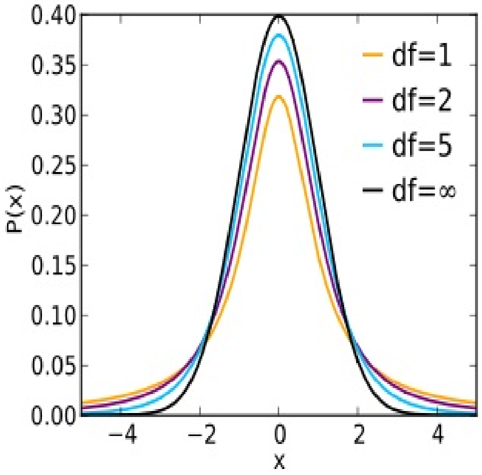
$$\bar{x} - z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}.$$

Remark!

Same formula as before with the difference that you replace σ (the population standard deviation) with s (the sample standard deviation).

Small sample size, IID Normal Population

t-distribution



Small sample size, IID Normal Population

- The degrees of freedom (*df*) is the number of values that are free to vary after a sample statistic has been computed and tell the researcher which specific curve to use.

$$df = n - 1,$$

where n is the sample size.

Small sample size, IID Normal Population

Confidence Interval

$$\left(\bar{x} - t_{\alpha/2}\,\frac{s}{\sqrt{n}}\,,\ \bar{x} + t_{\alpha/2}\,\frac{s}{\sqrt{n}}\right),$$

with degrees of freedom $n - 1$.

The values $t_{\alpha/2}$ can be found in the corresponding table in your book.

Example

Find the $t_{\alpha/2}$ value for a 95% confidence interval when the sample size is 22.

Since the sample size is 22, then

$$df = 22 - 1 = 21$$

So, we are looking for 21 in the left column of the t-distribution and we find the value under the column labeled 95%.

Degrees of Freedom	90%	95%	98%	99%
20	1.725	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819

Steps to Compute a Confidence Interval for μ with unknown σ and $n \leq 30$

- Step 1: Compute the sample mean ($\bar{x}$) and the sample standard deviation (s).
- Step 2: Find $\alpha/2$.
- Step 3: Find $t_{\alpha/2}$ in the distribution table.
- Step 4: Substitute in the formula

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

Example: *College wrestlers*

A **sample** of 6 college wrestlers had an **average weight of 276 pounds** with a **sample standard deviation 12** pounds.
Find a 90% confidence interval of the average weight of all college wrestlers.

Example: *College wrestlers*

- $n = 6, \;\; \bar{x} = 276, \;\; s = 12$
- $\alpha = 1 - 0.9 = 0.1$
- The corresponding value of $t_{\alpha/2}$ found from the table for $df = 6 - 1 = 5$ is

$$t_{\alpha/2} = t_{0.05} = 2.015$$

- Plug-in ...

$$\begin{aligned}
\bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} < \quad \mu \quad &< \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \\
276 - 2.015 \cdot \frac{12}{\sqrt{6}} < \quad \mu \quad &< 276 + 2.015 \cdot \frac{12}{\sqrt{6}} \\
266.13 < \quad \mu \quad &< 285.87.
\end{aligned}$$

To summarize...

Confidence Intervals for the proportion

Elections' Example

How to estimate the proportion of people voting for the Republican candidate?

- Population:

  p = proportion of people *in the population* voting for the Republican candidate

- Sample:

  $\hat{p}$ = proportion of people *in the sample* voting for the Republican candidate

Turn on your clickers!

- p is a parameter or a statistic?

(a) Parameter

(b) Statistic

Turn on your clickers!

- $\hat{p}$ is a parameter or a statistic?

(a) Parameter

(b) Statistic

Elections' Example

How to estimate the proportion of people voting for the Republican candidate?

- Population:

  p = proportion of people *in the population* voting for the Republican candidate

  p = usually unknown parameter (parameter)

- Sample:

  $\hat{p}$ = proportion of people *in the sample* voting for the Republican candidate

  $\hat{p}$ = computed based on the values in the sample (statistic)

Approximate distribution of $\hat{p}$

- $X_1, \ldots, X_n$ are independent and identically distributed (SRS).
- *If the sample size n is large enough* from the **Central Limit Theorem**

$$\hat{p} \sim \mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right).$$

68%-95% Confidence Intervals for p

When the sample size n is large enough we have that

$$\hat{p} - \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \ < \ p \ < \ \hat{p} + \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

is an (approximate) *a 68% confidence interval for p*,

$$\hat{p} - 2 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \ < \ p \ < \ \hat{p} + 2 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

is an (approximate) *a 95% confidence interval for p*.

Confidence Intervals for p

- Generally,

$$\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- Margin of Error

$$E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

The margin of error is the maximum likely difference between the point estimate of a parameter and the actual value of the parameter.

Example: *Elections*

Suppose that in a simple random sample of n=100 Americans we observe that the proportion of these voting for the Republican candidate is 30%. Compute a 95% confidence interval for the population proportion.

The sample proportion in this example is $\hat{p} = 0.3$
and a 95% confidence interval will be:

$$\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.3 - 1.96 \cdot 0.045 \ < p < \ 0.3 + 1.96 \cdot 0.045$$

$$0.21 < p < 0.39.$$

Example: *Freshman college students*

- In recent years over 70% of freshman college students responding to a national survey have identifies “being well-off financially” as an important personal goal.
- A liberal arts college finds that in an **SRS** of 200 of its freshman, 132 consider this goal important.
- Give a 95% confidence interval for the proportion of all freshman at the college who would identify being well-off as an important personal goal.

Example: *Freshman college students*

- Sample of size $n = 200$.
- *Sample* proportion of freshman that identify "being well-off financially" as an important personal goal is

$$\hat{p} = \frac{132}{200} = 0.66$$

- A 95% confidence interval is given by

$$\hat{p} - 1.96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + 1.96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.66 - 1.96 \cdot \sqrt{\frac{0.66(1-0.66)}{200}} < p < \hat{p} + 1.96 \cdot \sqrt{\frac{0.66(1-0.66)}{200}}$$

$$0.66 - 0.03 < p < 0.66 + 0.03$$

$$0.63 < p < 0.69$$

Sample Size Calculation

The minimum sample size required for a study in order to compute a confidence interval with a specified margin of error is equal to

$$n = \left(\frac{z}{E}\right)^2 p \cdot (1 - p).$$