

## Chapter 9

Confidence interval for means and proportions

Sample size calculations

\* Homework 7: 3, 6, 12 only

Popper Ch. 9 at the end of the script

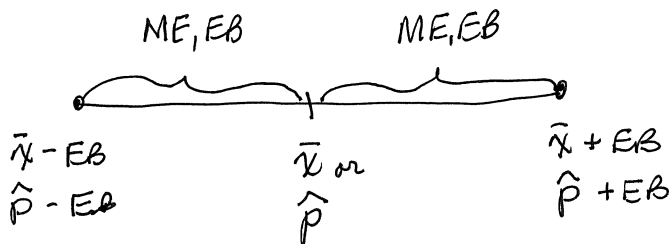
No essay

### Confidence Interval Overview

A confidence interval is a horizontal interval centered at a sample statistic that may contain the associated population parameter.

Let's look at an example:

$\mu \neq \rho?$



Confidence Level (aka Level of Significance) is a percent and comes from the process of finding the error bound (aka margin of error). It tells us how often the true parameter is actually contained in the confidence interval.

90% confidence

$\frac{90}{100}$  true mean or  $p$   
is somewhere in interval

$\frac{10}{100}$  not!

Let's look at the steps to create a CI:

1. take a random sample from a population *every sample is equally likely*
2. find the descriptive statistic (for us mean or proportion)  $\bar{x}$   $\hat{p}$
3. Decide on a confidence level (do a sample size calculation or follow the convention for your area of study)
4. Find the relevant z-score.
5. Compute the ME (EB) and report your interval along with the Level of Significance.

Steps 1 and 2 – we've already talked about that.

Step 3 Decide on a confidence level

Sample size costs and practicalities

Traditional levels in your field *sociology 80% medical 99*

90%, 95%, 99% *standard ones*

Step 4 Use the chart to find the relevant z-score by finding the confidence level in the chart body and move left and up to get the z-score you'll use

*← .95 ↑ like percentile work*

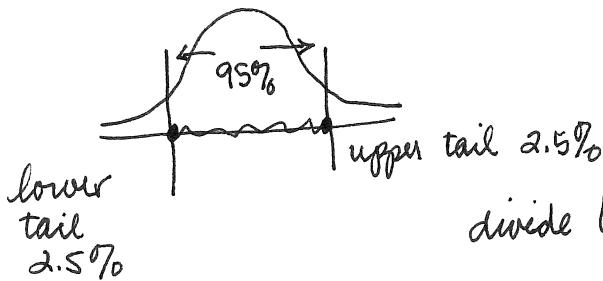
Step 5 Compute the ME. Let's look closely at this one.

Suppose your ME is 90%. We use this in the context of a Standard Normal Distribution:

We want our CI to be CENTERED on the horizontal axis under the Normal Distribution.

$\alpha = 95\%$  level of confidence

$\bar{x} \neq \hat{p}$  are normally dist



divide level of sig by 2 for tails

$$\frac{\alpha}{2}$$

So we will have TWO areas off of our interval. A lower tail and an upper tail.

Suppose

Now our level of confidence is 90% and we have 10% left to deal with under the curve. That's 5% under the lower tail and 5% under the upper tail -- we divide our level of confidence by 2! So our interval is centered.

The relevant z-score is called the Critical Value or z-crit. It's 1.645 because it's found by averaging the z-scores for the area values .949 and .951 in the chart.

The total amount of area to distribute is called alpha. And when you distribute the area under each tail you call it alpha/2.

$\alpha = 10\%$  if level of sig is 90%  
 $\alpha/2 = 5\%$  under each tail

Let's look at the alphas associated with the usual levels of significance:

$$90\% \quad \alpha = 10\% \quad \alpha/2 = 5\%$$

$$95\% \quad \alpha = 5\% \quad \alpha/2 = 2.5\%$$

$$99\% \quad \alpha = 1\% \quad \alpha/2 = .5\%$$

Step 5 continued

The formula for computing the ME is

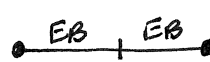
$$\bar{x} \pm z_{crit} \cdot \frac{s}{n}$$

kbh

$$\hat{p} \pm z_{crit} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Note that it uses the sample stat standard deviation and is a +/-.

Step 6 draw the interval and include the level of significance or the alpha in the body of your report or in a caption.



90% LS.

Note that if our alpha is 10% and our level of confidence is 90% we are saying that 90 times out of 100 if we do this process on this population the population parameter will be in the CI, but 10% not. Do we have a 90%er or a 10%er?

WE DON'T KNOW!

Now an example on the next page:

A large hospital wants a confidence interval for average length of stay. They randomly sample 100 recent, completed patient records.

$n > 30!$   
good.

Next page: so the whole calculation will be on one page

here's the facts

Sample mean: 7.84      Variance: 88.85      n=100

Confidence level: 95%      alpha: 5%      half of alpha: 2.5%

z-crit: ~~±1.645~~ 1.96

$$\bar{x} \pm z_{crit} \cdot \frac{s}{\sqrt{n}}$$
$$7.84 \pm \frac{1.645}{1.96} \cdot \sqrt{\frac{88.85}{100}} \qquad 7.84 \pm 1.85$$



Report: 95% confident that  $\mu$  is in  $[5.99, 9.69]$

where? don't know

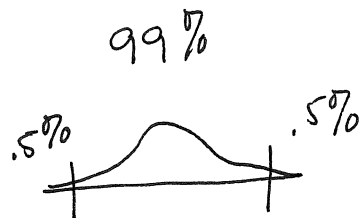
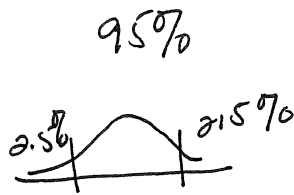
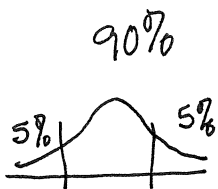
Summary of z-crits:

90% 1.65

95% 1.96

99% 2.58

Lengths of CI for each confidence level:



Another example. This time proportions!

We toss a flat cardboard up in the air vigorously. It can land on the red side (success) or the blue side (failure). What is a 90% CI for landing on the red side?

P-hat red: .58 Q-hat blue: .42 n = 100 alpha = 10%

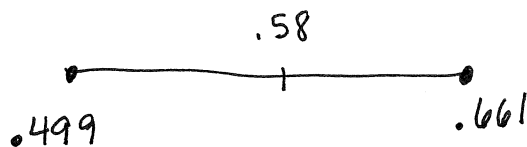
z-crit 1.645

Formula: using the standard deviation formula for proportions!

$$\hat{p} \pm z_{crit} \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad .58 \pm 1.645 \sqrt{\frac{.58 \times .42}{100}}$$

$$.58 \pm 0.081$$

Picture:



Findings:

$\frac{90}{100}$  times yep  $\frac{10}{100}$  nope which do we have? don't know!

Sample sizes:

For practical reasons it is often a good idea to figure out the appropriate sample size BEFORE you do the random sample. Sometimes you are on a budget and have a limit on what size sample you can do, for example. Sometimes you need a specific sized alpha to be consistent with other research papers in your field (pharmacy...)

Let's look at both formulas. We will start with the EB and derive the sample size formulas:

mean

$$EB_{\bar{x}} = z_{crit} \times \frac{A}{\sqrt{n}}$$

known

$$E^{\sigma} = z_c \frac{\Delta^{\sigma}}{n}$$

$$n E^{\sigma} = z_c \Delta^{\sigma}$$

$$n = \left( \frac{z_c \Delta^{\sigma}}{E} \right)^2$$

SQ b sides

SFn

$$EB_{\hat{p}} = z_{crit} \times \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

known

$$E^{\sigma} = z_c \times \frac{\hat{p}\hat{q}}{n}$$

$$n E^{\sigma} = z_c \hat{p}\hat{q}$$

$$n = \frac{z_c^2 \hat{p}\hat{q}}{E^{\sigma}}$$

Example: using the proportion example above let's fill in the proportion sample size formula and see what size n we need to hit 90%.

Using the red side/blue side problem

$\hat{p} = .58$     $\hat{q} = .42$     $n = 100$    change CI to 99%  
 $\alpha = 1\%$     $\alpha/2 = .5\%$   $\downarrow$   
 $z_{crit} = 2.576!!$    desired EB for @ side

$$n = \frac{(z_{crit})^2 \hat{p} \hat{q}}{E^2} \Rightarrow \frac{(2.576)^2 (.58)(.42)}{.5\% = (.005)^2} =$$

$\uparrow$  for side

646.59  $\rightarrow$  647

that's a lot!

lower  $E^2$ ,  $z_{crit}$



You can also use "n" in the formula and solve for one EB!



Popper 9

1. The number is the middle of a confidence interval is

a  $\mu$  or  $\rho$   
b  $\bar{x}$  or  $\hat{p}$

2. You can do a nonstandard CI of 80% level of significance by following the steps. What amount of area would be under the upper tail?

a 10%  
b 15%  
c 20%  
d 25%

3. A 99% CI is longer than an 90% CI.

a T  
b F

4. An 85% CI has the population parameter in the interval 15% of the times it is taken.

a T  
b F