Packet 5: Confidence Interval for the Population Proportion Textbook pages: 428 – 442

After completing this material, you should be able to:

- construct a confidence interval for the proportion using the appropriate format.
- state when it is valid to use this procedure.
- explain what "confidence" means.
- calculate the sample size needed to estimate the proportion within a specified margin of error for a desired degree of confidence.

NKU and Murray State University are two regional state schools in Kentucky. While both of these schools draw from the surrounding states to fill its student body, most believe that the majority of students are from the state of Kentucky. The Council of Postsecondary Education would like to sample students from both schools to estimate the proportion of students from KY. A sample of 562 NKU students found that 400 were from KY. Similarly, a random sample of 485 Murray State students found that 354 were from KY.

Given only this information, what would your estimate be for the proportion of **all** NKU students that are from KY? What would the estimate be for **all** Murray State students that are from KY?

NKU: $\hat{p} = \frac{400}{562} = .7117$ Murray: $\hat{p} = \frac{354}{485} = .7299$

Do you have any confidence that these two estimates are *exactly* correct? What could be done to improve your estimate?

I have almost zero confidence that these are exactly correct, but I'm fairly confident that they're awfully close. Pretty big samples.

What is a confidence interval? A confidence interval is an interval that estimates a number - a parameter. The confidence we demand (degree of certainty) reflects the likelihood that the true value of the parameter is actually inside that interval.

We want to develop a confidence interval which will be used to estimate the population proportion. What will the formula for this interval be?

point estimate ± margin of error

$$\hat{p} \pm z\sqrt{\frac{p(1-p)}{n}}$$

← we don't know p! So use $\hat{p}$...

$$\approx \boxed{\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

z crit related to the confidence required

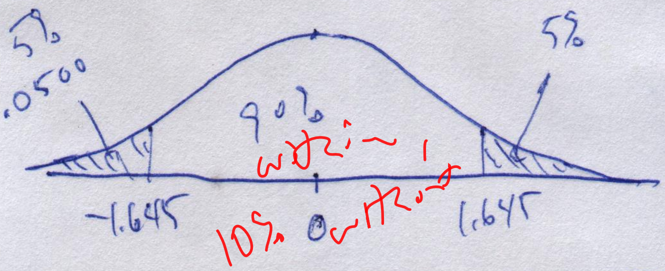
When calculating the confidence interval, the margin of error requires a critical point from the normal distribution. Let's fill in the table below with critical values associated with common levels of confidence:

The higher the / The bigger the

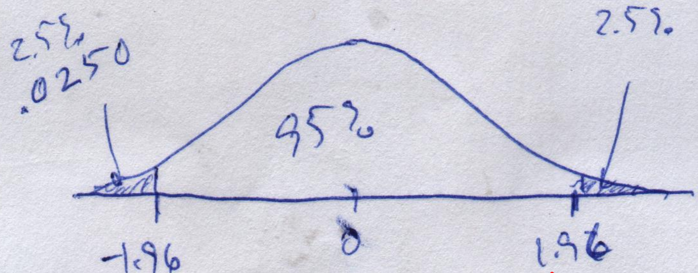
Level of confidence	90%	95%	96%	98%	99%
z critical value	1.645	1.96	2.054	2.326	2.576

This table will be given on the exam.

& the bigger the z-value, the wider the confidence interval!



$$\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$



Steps in a Confidence Interval

Step 1:

confidence required — standard deviation

$$\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Collect $\hat{p}$, z, n

point estimate — Margin of error — determined by the confidence required

Formula Alert!! This formula will be given on the formula sheet.

Step 2:

Actually carry out the calculations, & report the lower and upper limits of the interval.

Step 3: Interpret the interval – explain what it means, level of confidence, etc.

EVERY confidence interval we do will follow these same three steps – start to commit them to memory!!

Example: On the first page of the notes, we were given information about a sample of students from two regional universities in Kentucky. For each school, estimate the proportion of students from Kentucky with 95% confidence.

Northern Kentucky University

Step 1

$\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $\hat{p} = .7117$ $n = \cancel{1244}\ 562$ $z = 1.96$

Step 2

$.7117 \pm 1.96\underbrace{\sqrt{\frac{.7117(1-.7117)}{562}}}_{0.0375}$

$.7117 + .0375 = .7492$

$.7117 - .0375 = .6742$

$\boxed{[.6742 , .7492]}$

Step 3

With 95% confidence we can say the true percentage of KY students at NKU is between 67.42% and 74.92%.

Murray State University

Step 1

$\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $\hat{p} = .7299$ $n = 485$ $z = 1.96$

Step 2

$.7299 \pm 1.96\underbrace{\sqrt{\frac{.7299(1-.7299)}{485}}}_{.0395}$

$.7299 + .0395 = .7694$

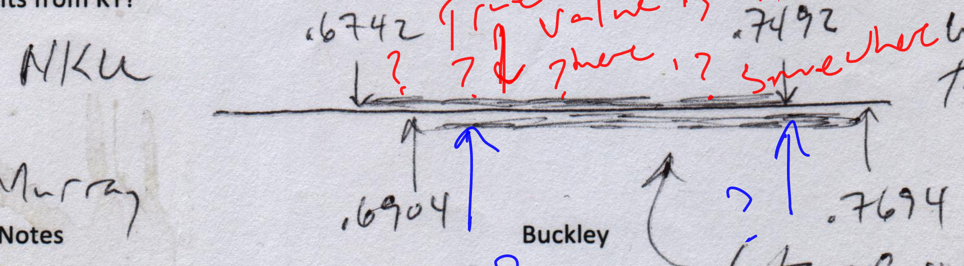
$.7299 - .0395 = .6904$

$\boxed{[.6904 , .7694]}$ ✓

Step 3

With 95% confidence, we estimate that the true percentage of KY-born students at Murray State is between 69.04% and 76.94%.

Based on your two intervals, can you draw any type of conclusion about the two schools with respect to the proportion of students from KY?



We can't conclude that their p's are different.

lots of overlap

Example: The US Commission on Crime wants to estimate the proportion of crimes related to firearms in an area that has one of the highest crime rates in the country. The commission randomly selects 600 files of recently committed crimes in the area and finds that 388 involved the use of a firearm. Estimate the percentage of all crimes in this area in which a firearm was used with 99% confidence.

$z_{crit} = 2.576$ $\quad \hat{p} = \frac{388}{600} = .6467$ $\quad n = 600$

Step 1: $\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $\quad \hat{p} = .6467 \quad n = 600 \quad z = 2.576$

Step 2: $.6467 \pm 2.576\sqrt{\frac{.6467(1-.6467)}{600}}$

$\pm .0503$

$[.5964, .6970]$

Step 4: With 99% confidence we estimate that between 59.64% and 69.70% of crimes in this area involve a firearm.

Would the confidence interval become wider or narrower if we lowered the level of confidence? Would the confidence interval become wider or narrower if we increased the sample size? Explain.

The more confidence you require the larger the box. (z_{crit} bigger → MOE bigger)

less you trust that point estimate

The larger the sample size, the smaller the box. ($\sigma_{\hat{p}}$ smaller → MOE smaller)

z_{crit} & $\sigma_{\hat{p}}$

What *assumptions* must be satisfied for the interval to be valid?

Random sample

$n\hat{p} > 10$ and $n(1-\hat{p}) > 10$

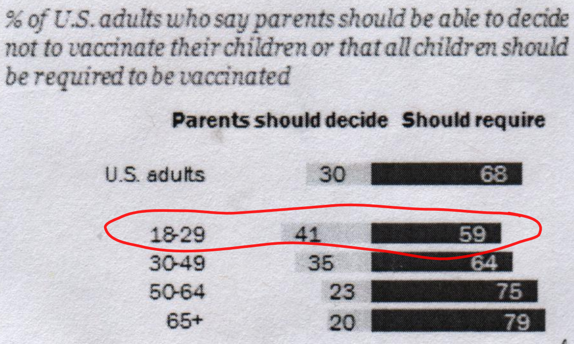
We're estimating p, the parameter using a ~~stat~~ statistic based on a sample.

Calculating a Sample Size

Textbook pages: 496 – 497

The question of how large a sample to take is an important one – we recently decided that the larger the sample size, the smaller the margin of error. When estimating proportions, we can determine the sample size required to achieve a certain margin of error for a given level of confidence. Consider the following example:

In 2014, Pew Research Center conducted a survey to gauge public opinion on the topic of vaccination of children. The study was done as the number of measles cases linked to the California outbreak climbed to over 100. At that time, health officials were urging parents to properly immunize their children, citing unvaccinated individuals as a main contributor to the disease's spread. The results from this survey are summarized below:



The debate over vaccinations continues to receive national attention. To update the estimates, a new survey is planned for this year. Suppose the researchers would like to estimate the proportion of all US adults who feel vaccinations should be required with 95% confidence and a margin of error of no more than $\pm$ 3%. How large a sample should be taken given the estimates provided by the 2014 Pew Research study?

$MoE = .03$ $\quad z = 1.96$ $\quad p = .68$ from the study

$$\boxed{MoE = z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \rightarrow \frac{MoE}{z} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \rightarrow \left(\frac{MoE}{z}\right)^2 = \frac{\hat{p}(1-\hat{p})}{n} \rightarrow$$

$$n\left(\frac{MoE}{z}\right)^2 = \hat{p}(1-\hat{p}) \rightarrow \boxed{n = \left(\frac{z}{MoE}\right)^2 p(1-p)}$$

Always round up.

$z = 1.96$
$MoE = .03$ $\quad p = .68$

$$n = \left(\frac{1.96}{.03}\right)^2 .68(1-.68) = 928.8 \rightarrow \boxed{n = 929}$$

Would the required sample size be larger or smaller to achieve the same marge of error for adults between the ages of 18 and 29?

$$n = \left(\frac{1.96}{.03}\right)^2 .59(1-.59) = 1032.5 \rightarrow \boxed{n = 1033}$$

Sample size n is largest when p approaches $\boxed{p = .5}$:

$$n = \left(\frac{1.96}{.03}\right)^2 .5(1-.5) = 1067.1 \rightarrow \boxed{n = 1068}$$

So: What if there is no prior estimate for the proportion which ~~can be~~ should be used as an estimate?

Take $p = .5$ & you'll get the largest sample possible – so it's a conservative estimate – the answer will be better than required, but at the cost of extra sampling.

Example: The National Center for Educational Statistics keeps track of a variety of measures related to education in the United States. One area of interest is the proportion of full-time undergraduate students at public institutions who receive some form of financial aid. A statistician would like to estimate the percentage of these students receiving financial aid in the 2014-15 academic year.

- The statistician decides to use estimates from the 2006-07 academic year to help gauge how large a sample size should be taken. He would like to estimate the true proportion of full-time undergraduate students at public institutions who receive some form of financial aid with 95% confidence and a margin of error of ±2.75%. If 75% of these students received financial aid in 2006, how large a sample size should be taken in 2014 to estimate the percentage with the desired margin of error?

$z = 1.96$ MOE $= .0275$

$p = .75$

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

$$= \left(\frac{1.96}{.0275}\right)^2 .75(1-.75) = 952.\boxed{46} \xrightarrow{\text{round up!}} \boxed{n = 953}$$

- Realizing the proportion will have changed some in 2014, the statistician decides to take a random sample of 1,025 full-time undergraduate students at public institutions. In those sampled, he found 851 were receiving financial aid. Using this information, estimate the quantity of interest with 95% confidence.

$n = 1025$ $\hat{p} = \frac{851}{1025} = .8302$ $z = 1.96$

Step 1: $\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Step 2: $.8302 \pm 1.96\sqrt{\frac{.8302(1-.8302)}{1025}}$

$.8302 \pm .0230$ (an unexpectedly small box! Smaller than .0275.)

$.0230 < .0275$

$[.8072, .8532]$

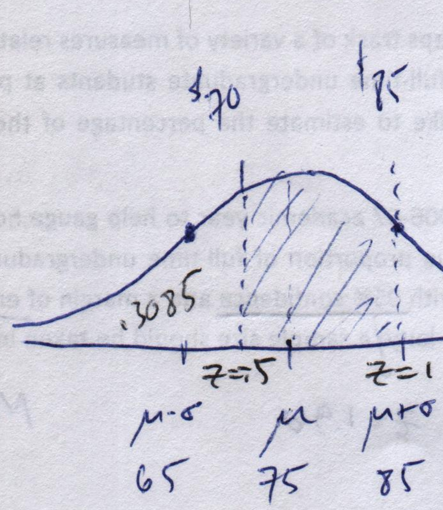
$n >$ required!

Step 3: With 95% confidence we estimate that between 80.72% & 85.32% of full-time undergrads receive financial aid.

- Give two reasons why the margin of error found when calculating the interval was different than the desired margin of error of 2.75%.

#1: we took a larger sample size than "required".

#2: p changed, from .75 to .8302. - since .8302 moved away from .5, we generally require a smaller sample to get the same MOE.



Mean hotel rate is $75, and σ is 10.

What percentage of hotel rates are between $70 & $85?

$$z = \frac{obs - \mu}{\sigma} : \frac{70-75}{10} = -.5$$

$$\frac{85-75}{10} = 1$$

$$\underbrace{.8413}_{\text{Area to the left of } z=1} - \underbrace{.3085}_{\text{area to the left of } z=-.5} = \boxed{.5328}$$

53.28% of hotel rates are between $70 & $85.