Confidence Intervals for Proportions

Let's review a few terms that will aide us in our discussion of Confidence Intervals.

- **Population:**

- **Sample:**

- **Information:**

- **Parameter:**

- **Statistic:**

Recall from past lectures that when we want to estimate the population proportion __________ we often take a random representative sample and use _________________________ to estimate p. However because of _________________________ inherent in each sample we take, often times there is a degree of uncertainty in using $\hat{p}$ to estimate p. To capture this uncertainty in our estimates we can add a **margin of error** to our estimate of our parameter, which is the general form of what's known as a _________________________.

Review:

Recall from our discussion of sampling distributions for sample proportions that under the following conditions:

i. Randomization Condition

ii. 10% Condition

iii. Success/Failure Condition

Then:

We could then standardize our observed values of p using the Normal model and a z-table using the following formula:

Introduction to Confidence Intervals

Using a z-table to find that in 95% of all samples, the sample proportion $\hat{p}$ will be between the following two values:

We use 1.96 instead of 2 because it is more accurate since it is derived from the z-table.

Result One: In other words in 95% of all samples, the ________________________ $\hat{p}$ will be within 1.96 standard deviations of the ________________________ p. We are trying to determine where the sample proportions are in relation to the population proportion, in mathematical language we say:

Result Two: An additional result that 95% of all samples, the ________________________ p will be within 1.96 standard deviations of the ________________________ of $\hat{p}$. Here we are trying to determine where the unknown population proportion is in relation to the sample proportions, in mathematical language:

The first result comes directly from work we did in this module about sampling distributions for sample proportions. The second result simply flips the first result around.i.e.:

-
-

But we have a problem with Result Two, the standard deviation includes the value of p which is unknown. So our solution is to ______________________________________

Final Result

In approximately 95% of all samples, the population proportion p will be within 1.96 standard deviations of the sample proportion $\hat{p}$. In mathematical language we would say the following:

To illustrate this concept, let's look at an example.

Example *In a random sample of 1000 U.S. adults, 38.8% of them state that they believe in the existence of ghosts. Find a 95% confidence interval for the population proportion of all U.S. adults who believe in the existence of ghosts.*

To answer this problem, we will complete the following steps:

i. Identify important parts of the problem

ii. Check conditions for Normality

iii. Calculate the confidence interval

iv. Interpret the confidence interval

Identify the following important pieces of the problem:

- *Population:*
- *Sample:*
- *Parameter:*
- *Statistic:*

Check the conditions for Normality:

i. Random sample condition:

ii. Success/Failure condition:

iii. 10% condition:

Calculate the confidence interval

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}*(1-\hat{p})}{n}}$$

Interpretation:

We are 95% confident the proportion of all adults in the U.S. who believe in the existence of ghosts is between ________________________ and ________________________ percent.

In general, to interpret a confidence interval at any confidence level, we use the following template for our interpretation:

We are __________ (confidence level) confident that the proportion of ______________________ (population description) who/that ________________________ (category or event of interest) is between ________________________ and ________________________ (lower and upper bounds of the confidence interval)

It is important to note that the confidence level represents the percentage of all samples taken randomly from the population where the population proportion is contained within the bounds of the confidence interval. For a single sample, we can't say for sure whether the population parameter is within that particular confidence interval, but we can say we are 95% confident.

Returning to our first example, if we were to take 100 samples of size 1000, and computed a 95% confidence interval for each sample, we would expect approximately 95% of these confidence intervals would contain the true proportion of U.S. adults that believe in ghosts.

Other Confidence Levels

If the conditions for normality are met, then a confidence interval for a population proportion can be computed using the following formula:

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

where the z^* value is chosen based on the desired confidence level:

Confidence Level	**80%**	**90%**	**95%**	**98%**	**99%**
z^*	1.282	1.645	1.96	2.326	2.576

Example *344 out of a random sample of 1,010 U.S. adults rated the economy as good or excellent in a Gallup Poll from October 4-7, 2007. Find a 98% confidence interval for the population proportion of all U.S. adults who believed the economy was good or excellent during this time frame.*

Identify the following important pieces of the problem:

- *Population:*
- *Sample:*
- *Parameter:*
- *Statistic:*

Are the following Normality conditions met?

i. *Random Sample Condition:*

ii. *Success/Failure Condition:*

iii. *10% Condition:*

Calculate the confidence interval

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

Interpretation:

We are 98% confident that the proportion of ______________________________ *who* __________________________________ *is between* ________ *and* ________.

Properties of Confidence Intervals

Recall that the equation for a confidence interval for proportions is:

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

The ____________________ consists of two components: the standard error, and the critical value. The margin of error gives a rough estimate about the uncertainty of our estimates.

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

Confidence Level	**80%**	**90%**	**95%**	**98%**	**99%**
z^*	1.282	1.645	1.96	2.326	2.576

Based on the above formula we can essentially deduce the following:

For a fixed sample size (n)

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

- Smaller confidence level $\Longrightarrow$ smaller z^* $\Longrightarrow$ ________________
- Larger confidence level $\Longrightarrow$ larger z^* $\Longrightarrow$ ________________

Notice how a larger confidence level requires a larger or wider confidence interval! This is because in order to increase our confidence, we need to make "more room" for variation in our estimates.

For a fixed confidence level C%:

Effect of sample size on margin of error (ME):

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

- Smaller samples $\implies$ smaller n $\implies$ ___________________
- Larger samples $\implies$ larger n $\implies$ ___________________

So it would seem that larger samples give more precise estimate of the population proportion!

We really have two goals for confidence intervals, we want a high degree of confidence but a small margin of error. But this seems to be contradictory! So the best thing we can do is determine the sample size needed to obtain the desired confidence level and margin of error.

In our last example about the economy we found a 98% confidence interval for the proportion of all U.S. adults who believed the economy was good or excellent during this time frame. Suppose we took a new sample that was doubled in sample size, but still wanted to find a 98% confidence interval. So instead of 344 adults out of 1,010 adults rating the economy as good is now 688 out of 2,020.

Originally, our 98% confidence interval was between (0.306, 0.376) but with our new sample size, we receive a confidence interval of the following:

$$0.341 \pm 2.326\sqrt{\tfrac{0.341(1-0.341)}{2020}} = (0.316, 0.366)$$

Notice how the width of the confidence interval decreases as our sample size increases!

Sample Size Calculation

Recall that the equation for the ___________________ is as follows:

$$\text{ME} = z^*\sqrt{\tfrac{\hat{p}(1-\hat{p})}{n}}$$

We note that the margin of error relies on the statistic $\hat{p}$ but because we haven't taken a sample yet we don't know what $\hat{p}$ is! In order to find a rough estimate of our sample size we assume that $\hat{p} = 0.5$. So let's solve for our sample size, n.

$$\begin{aligned}
ME &= z^*\sqrt{\frac{0.5^2}{n}} \\
\frac{ME}{z^*} &= \sqrt{\frac{0.5^2}{n}} \\
(\frac{ME}{z^*})^2 &\geq \frac{0.5^2}{n} \\
n(\frac{ME}{z^*})^2 &\geq 0.5^2 \\
n &\geq (\frac{0.5z^*}{ME})^2
\end{aligned}$$

Example *We would like to obtain a 95% confidence interval for the population proportion of U.S. registered voters who approve of President Obama's handling of the war in Afghanistan. We would like this confidence interval to have a margin of error or no more than 3%. How many people should be in our sample?*

Note: That there is not a 1-1 relationship between sample size and margin of error. For example, if we wanted to cut our margin of error in half, we would need four times as many people in our sample.

Example *A sample of 500 pizza delivery orders from a local pizzeria found that 84% of orders were paid for with a credit card (the rest were paid in cash). Find a 90% confidence interval for the proportion of all pizza orders that were paid for with a credit card. The suppose we would like to obtain a 98% confidence interval for the population proportion of pizza delvieries that were paid for with a credit card. We would like the confidence interval to have a margin of error of no more than 3%, what should our sample size be?*