Confidence Intervals for Proportions:

We now want to be able to build confidence intervals when our parameter of interest is not a mean, but a *proportion*; i.e. what proportion of Americans smoke? For one-sample proportions, the confidence interval looks like

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

where $\hat{p}$ is the observed proportion in the sample taken, n is the sample size, and z^* is the critical value corresponding to the confidence level you desire. Let's look at an example.

Example 1: In a 1997 survey done by the Marist College Institute for Public Opinion, 36% of a randomly selected sample of 935 American adults said that they don't get enough sleep at night.

(a) What is $\hat{p}$ in this example? What is n?

(b) Find the z^* value needed for a 95% confidence interval.

(c) Construct a 95% confidence interval for the true proportion of American adults who don't get enough sleep at night, and interpret it in the context of the problem.

We have to be careful to make sure our sample sizes are sufficiently large to justify use of the z multiplier. In general, you should check that

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

Return to Example 1 and make sure that you were justified in using a z value for your confidence interval:

Example 2: Let's try another example. Suppose that a new treatment for a certain disease is given to a sample of 200 patients. The treatment is successful for 166 of the patients. Calculate a 99% confidence interval for the proportion of the total population for whom the treatment would be successful. (Make sure to check that you are justified in using your z value).

What if we want to compare proportions for two different populations? Just like with means, we can construct confidence intervals for the difference in population proportions. In this case, our confidence interval will look like

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

In order to be justified in using z, you should check that $n\hat{p}_1$, $n(1-\hat{p}_1)$, $n\hat{p}_2$, and $n(1-\hat{p}_2)$ are all greater than 10. Let's look at some examples.

Example 3: Students in a statistics class at Penn State were asked "Would you date someone with a great personality even though you did not find them attractive?" Of the 131 women in the class, 61.1% answered "yes". Of the 61 men in the class, 42.6% answered "yes." Construct a 95% confidence interval for the true difference in proportion who answered "yes" in women vs. men. Can we conclude that the two population proportions are likely to be different?

Example 4: In a CBS News survey done in 2009, 95% of $n = 346$ randomly sampled married men said that they would marry their spouses again if they had to do it all over again. In the same survey, 85% of $n = 522$ married women said they would marry their spouses if they had to do it all over again. Calculate a 90% confidence interval that estimates the difference in proportions of married women and men who would marry their spouse again. Can we infer that the population proportions for women and men differ?