Stat 218 - Day 22
Confidence Intervals for a Population Proportion

Yesterday we learned how to conduct a goodness-of-fit hypothesis test involving a population proportion. Today we will learn how to construct a confidence interval to estimate the value of a population proportion.

Recall from Key Result 1 earlier in the quarter that if the population proportion having the characteristic of interest is *p*, and if we repeatedly take random samples from the population, then the sample proportion $\hat{p}$ will vary approximately normally with mean *p* and standard deviation $\sqrt{\frac{p(1-p)}{n}}$. Thus, we can estimate the population proportion *p* by the sample proportion $\hat{p}$, an approximate CI procedure to estimate *p* is: $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. We will consider this procedure to be valid as long as the sample is selected randomly and both $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$.

Example: Kissing the right way
Most people are right-handed and even the right eye is dominant for most people. Molecular biologists have suggested that late-stage human embryos tend to turn their heads to the right. German bio-psychologist Onur Güntürkün conjectured that this tendency to turn to the right manifests itself in other ways as well, so he studied kissing couples to see if they tended to lean to the right. He and his colleagues observed kissing couples in public places such as bus terminals and airports. They found that in a random sample of 124 kissing couples, 80 turned their heads to the right.

(a) Identify (with a symbol and in words) the population parameter of interest here.

(b) Calculate the value of the corresponding sample statistic, and identify the symbol used to represent it.

(c) Construct a 90% CI for the population parameter based on this sample.

(d) Interpret both this interval and the phrase "90% confident."

(e) Based on this CI, does it appear to be plausible that 50% of all kissing couples turn to the right? Explain.

(f) Based on this CI, does it appear to be plausible that 2/3 of all kissing couples turn to the right? Explain.

(g) Conduct a test of whether the population proportion differs from 2/3. Does the result confirm your answer to (f)?

(h) What would happen to the CI if we quadrupled the sample size, and all else remained the same?

Example: Gallup polls
Gallup and other national polls are typically based on a random sample of about 1000 people.

(a) Determine the margin-of-error associated with a 95% confidence interval if the sample proportion turns out to be around .5.

(b) Determine the margin-of-error associated with a 95% confidence interval if the sample proportion turns out to be around .75.

(c) Does the population size enter into these calculations of margin-of-error?

(d) If the proportion is expected to be close to .5, how many people should be interviewed to achieve a margin-of-error (with 95% confidence) of .04?

Applet exploration:
How do we interpret these confidence intervals? Just like before, except applied to the population proportion p rather than the population mean μ. In particular, the interpretation of "confidence level" is the same as always. We can explore this with a Java applet that simulates repeated random sampling from the population.

(a) Open the applet called "Simulating Confidence Intervals for Population Parameter." Set the population proportion (called θ rather than p in the applet) to .5 and the sample size to 200. Ask for 100 intervals with a confidence level of 95%. Click on "Sample" and keep clicking until you have generated 2000 intervals. What percentage of these 2000 intervals successfully capture the population proportion? Is this percentage close to 95%?

(b) How do you expect the intervals to change if the sample size is increased to 800? (Comment both on the width of the intervals and on the percentage that successfully contain the population proportion.) Explain.

(c) Make this change in the sample size and generate 2000 intervals. How did the intervals change?

(d) How do you expect the intervals to change if the confidence level is increased to 99%? (Comment both on the width of the intervals and on the percentage that successfully contain the population proportion.) Explain.

(e) Make this change in the confidence level and generate 2000 intervals. How did the intervals change?

(f) Now change the sample size to 10 and the confidence level to 95%. Generate 2000 intervals. What problem is revealed?

(g) Now change the sample size to 40 and the population proportion to .1. Generate 2000 intervals. What problem is revealed?

(h) Were the technical conditions of this CI procedure satisfied in (f) and (g)?

Alternative procedure:
To compensate for the poor performance of the conventional CI procedure in these situations, an alternative procedure has been proposed that essentially says to pretend that your sample contained two more "yes" responses and two more "no" responses than it really did. In other words, we estimate the population proportion p by a new estimator: $\tilde{p} = \frac{x+2}{n+4}$. Then we form the confidence interval for p by: $\tilde{p} \pm z^* \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$.

Applet exploration (cont.):
(i) Change the applet procedure from "Wald" to "adjusted Wald" and repeat (f) and (g). Are the percentages (of successful intervals) closer to 95% now?

Example: Handedness
Use this alternative procedure on class-generated data to estimate the proportion of Cal Poly students who are left-handed with 95% confidence. Also interpret each interval.