Section 7.1 - Confidence Intervals for Proportions

A **confidence interval** is an interval of values used to estimate the true value of a population parameter.

Our first confidence intervals will be used to estimate a population proportion.

For example...

A Pew Research Center poll found the 70% of 1501 randomly selected U.S. adults believe in global warming. A 95% confidence interval estimate of the population proportion p is

$0.677 < p < 0.723$.

Correct interpretations of this confidence interval

- We are 95% confident that the interval (0.677,0.723) actually contains the true value of the population proportion p.
- If we were to select many different samples of size 1501 and construct the corresponding confidence intervals, 95% of them would contain the true population proportion p.
- The process of computing the confidence interval will result in intervals that contain the true population proportion 95% of the time.

Incorrect interpretations of this confidence interval

- 95% of sample proportions fall between 0.677 and 0.723.
- There is a 95% chance that the true population proportion will fall between 0.677 and 0.723.

Computing population proportion confidence intervals...

Because sample proportions are normally distributed, we can find a confidence interval by computing the z-scores that bound a certain area (the confidence level) under the standard normal curve.

The **confidence level** is the probability $1 - \alpha$ that the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times.

The number $z_{\alpha/2}$ is the z-score with the property that it separates an area of $\alpha/2$ in the right tail of the standard normal distribution.

Confidence Interval for Population Proportion

It turns out that the confidence interval at the level $1 - \alpha$ is given by

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}},$$

where p is the population proportion, $\hat{p}$ is the sample proportion, $\hat{q} = 1 - \hat{p}$, and n is the number of sample values.

The number $E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$ is called the margin of error.

Determining Sample Size

Suppose we want to find the sample size required to estimate p with a desired margin of error.

When $\hat{p}$ is known: $n = \frac{[z_{\alpha/2}]^2 \hat{p}\hat{q}}{E^2}$

When $\hat{p}$ is not known: $n = \frac{[z_{\alpha/2}]^2 (0.25)}{E^2}$