Chapter 19: Confidence Intervals (C.I) for Proportions

One of the primary jobs of a statistician is to estimate characteristics of populations (parameters). What proportion of students drive to school? What is the average weight of a teenager?

To make these estimates, statisticians will select a sample from the population of interest and study the members of the sample. However, because of sampling variability, it is possible that our estimates will be wrong.

In general, there are two types of estimates that can be made: point estimates and interval estimates.

A <u>POINT ESTIMATE</u> is single number that represents our best guess for the population parameter. It is called a "point" estimate since it represents a single point on the number line. For example, based on a recent survey, we estimate that 61% of young adults favor social security reform.

An <u>INTERVAL ESTIMATE</u> is a *range* of plausible values for the parameter. It is called an "interval" estimate since it represents an interval on the number line. For example, based on a recent survey, we estimate that between 58% and 64% of young adults favor social security reform (61% ± 3%).

Using an interval estimate greatly increases the probability we are correct. For example, if I predict the high temperature tomorrow will be between 0 and 150 degrees I have great confidence that I will be correct. Of course, this interval won't help me pick my clothes in the morning. There is a tradeoff between confidence and usefulness.

<u>Making Interval Estimates</u>

Because of the variability inherent in sampling, our point estimates will rarely be correct. After all, different samples will yield different estimates. And, although a point estimate may be our best single guess for the value of a population parameter, it is certainly not the only plausible value.

Thus, statisticians will usually report an interval of plausible values for the population parameter based on the sample. This interval is called a <u>CONFIDENCE INTERVAL</u>. Using a confidence interval gives a much better chance of correctly estimating the parameter.

When the distribution of $\hat{p}$ is normal, the xth% confidence interval for a population proportion (p) is:

Point estimate ± margin of error

$$\hat{p} \pm "inversenorm(\frac{1-\frac{x}{100}}{2})"\sqrt{\frac{\hat{p}\hat{q}}{n}}$$ * ME = CV(SE) margin of error is Critical Value times Standard Error

Why do we use $\hat{p}$ in the standard deviation instead of p?

Note: When we use the sample to estimate the SD of the sampling distribution of $\hat{p}$, the estimate is called the Standard Error (SE). SD = $\sqrt{\frac{pq}{n}}$ and SE = $\sqrt{\frac{\hat{p}\hat{q}}{n}}$

How do we know if the distribution of $\hat{p}$ is approximately normal? (Conditions for a C.I.)

The general form of a confidence interval:

Point estimate ± margin of error

In other words….

Statistic ± (critical value)(standard error of the statistic)

Interpreting the confidence level. In other words, what does it mean to be 90% confident?

- In the long run, 90% of all samples will produce intervals that will capture the true value.
- Or: Before we gather our sample, there is a 90% chance that the sample we use will produce an interval that captures the true value.
- It does NOT mean that there is a 90% chance (or probability) the true value is in that *particular interval*. That particular interval either captures the value or it doesn't.

*The interval moves around the parameter is at a fixed point

The problem of childhood obesity in the United States has grown considerably in recent years. Obesity is among the easiest medical conditions to recognize but most difficult to treat. Unhealthy weight gain due to poor diet and lack of exercise is responsible for over 300,000 deaths each year. The annual cost to society for obesity is estimated at nearly $100 billion. Overweight children are much more likely to become overweight adults unless they adopt and maintain healthier patterns of eating and exercise.

* This article was from a Newspaper article published in 2016

In a recent study of 1050 randomly selected teenagers in the U.S., 231 of them were considered obese. Build a 90% confidence interval to estimate the true proportion of teenagers in the U.S. that could be considered obese. (Be sure to discuss the conditions and assumptions to validate your claim)