MATH 146 INFERENTIAL STATISTICS - 7.1 ESTIMATING A POPULATION PROPORTION - CONFIDENCE INTERVALS

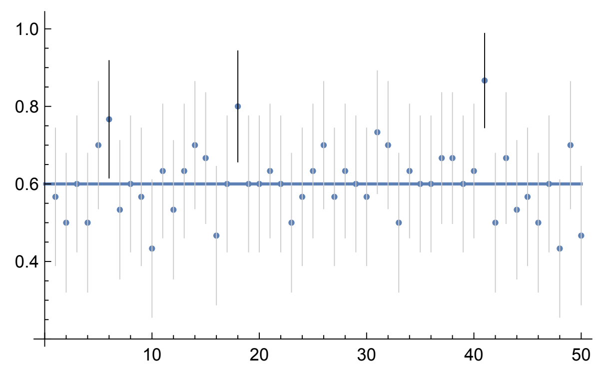
In the previous chapters we used *descriptive statistics* to summarize, or describe, data sets using graphs and statistics such as the mean, standard deviation, mode, probability distributions, etc. *Inferential Statistics* uses sample data to form generalizations or *inferences* about a population.

EXAMPLE 1 Suppose we want to find the percentage, p, of female students enrolled at SVC. Find the proportion of female students in this class.

Q? How confident are you of this point estimate ($\hat{p}$), that is, how close is it to the true percentage of female students? Can you give a range of values that may make you more confident?

A **confidence interval** is a range of values used to estimate the true value of a population parameter. For a given α, the confidence level is the probability, $1 - \alpha$, that the calculated interval from a sample actually contains the population parameter, assuming the estimation process is repeated a large number of times.

For a confidence level of $\alpha = 0.05$, we would expect about 95% of all samples (of the same size) to contain the population proportion. A simulation of 50 samples of size $n = 30$ from a population with a true proportion $p = 0.6$ gave the following estimate intervals:

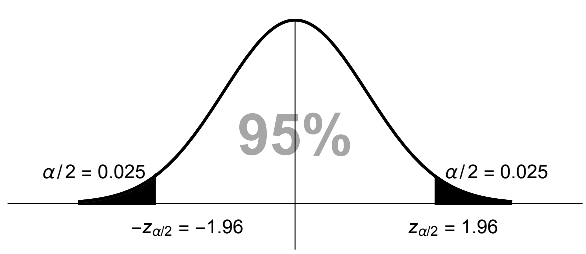


Suppose the point estimate was 17/30, or about $\hat{p} \approx 0.567$, for the proportion of female students. A 95% confidence interval (or at a 5% confidence *level*) would be (0.389, 0.744). Interpret this interval.

If someone *claimed* there were more female students than male students, we could use this confidence interval to informally refute that claim with 95% confidence. Why?

CALCULATING A CONFIDENCE INTERVAL FOR A PROPORTION

A confidence interval at the 5% *confidence level* ($\alpha = 0.05$) means we need an interval that contains 95% of the data, or, for a normal distribution, this means there is 0.025 ($\alpha/2$) area in each tail. For a normal distribution, these critical numbers, are $z_{\alpha/2} = \pm 1.96$



The confidence interval for the parameter p is given as an interval, and can be written as $\hat{p} - E < p < \hat{p} + E$, where E is called the margin of error. For the example above, $\hat{p} = 0.567$, and $E = 0.177$.

DEFINITION The *margin of error* for a proportion parameter with point estimate $\hat{p}$, sample size n, and confidence level α, is:

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$

The interval can also be written as $\hat{p} \pm E$. Typically, round to three decimal places.

REQUIREMENTS

1) The sample must be a simple random sample.

2) The conditions for a binomial distribution are met: a fixed number of trials, trials are independent, there are two categories of outcomes (success, failure).

3) $n\,p \geq 5$ and $n\,q \geq 5$ (using $\hat{p}$ as an estimate for p) This ensures a normal distribution can be used as an approximation to the binomial distribution.

EXAMPLE 2 A simple random sample of 300 registered voters (not a call in, self selected, or observational sample, etc.) at a political polling station resulted in 168 individuals that voted for a democrat. Find a confidence interval at the $\alpha = 0.10$ confidence level for the true proportion of democratic voters.

Compare using your hand calculated value with the TI-84. Press **stat**, **TESTS**, and scroll down **1-PropZInt...**

DETERMINING SAMPLE SIZE FOR A GIVEN CONFIDENCE LEVEL AND MARGIN OF ERROR

a) If $\hat{p}$ is known, solving for n in the margin of error formula above, we get

$$n = \frac{(z_{\alpha/2})^2\,\hat{p}\,\hat{q}}{E^2}$$

b) If $\hat{p}$ is not known, then use

$$n = \frac{(z_{\alpha/2})^2\,0.25}{E^2}, \quad \text{(why?)}$$

EXAMPLE 3 A television station claimed in a recent poll only 34% Seattleites supported a new gasoline tax. The poll had a margin of error of ±3% at a 95% confidence level. Find the sample size of the poll assuming they had prior knowledge that $\hat{p} \approx 0.3$. Round **up** if n is not an integer.

EXAMPLE 4 A sample of 500 people showed that 82% watched the Super Bowl.

a) Find a 95% confidence interval for the proportion of people that watched the Super Bowl.

b) Suppose we had no idea how many people watched the Super Bowl. How many people would we need to randomly sample to have a 98% confidence that our calculated confidence interval contained p within ±3%? (Typical polls have a ±3% error using a 95% confidence interval. Find this sample size.)