9.1 Estimation and Confidence Intervals for a Population Proportion

One aspect of statistics is estimation. The following examples illustrate examples of statistical estimating.

One out of four Americans is currently on a diet
72% of Americans have flown on a commercial airlines

Where did these estimates come from and how can we be certain that they are accurate? After all, the population of Americans is pretty big and these kinds of estimates are usually the result of sampling.

The following questions arise when statisticians talk about estimates.

1. How big of a sample do we need to accurately make estimates for populations?
2. How confident are we in the accuracy of our estimates? Are they accurate representations of the population?
3. If we were able to survey all Americans would we really find that 1 out of 4 Americans is currently on a diet?

Review:
Parameters: Values obtained from data that came from a population (μ, σ)

Statistics: Values obtained from data that came from a sample ($\overline{x}$, s)

Symbols used in proportion problems
P = symbol for the population proportion
$\hat{p}$ = symbol for the sample proportion

For a sample proportion, $\hat{p} = \frac{x}{n}$ and $\hat{q} = 1 - (\hat{p})$, where x = the number of sample units that possess the characteristic of interest in the study and n = sample size.

Point Estimates: A specific numerical value is used as the estimate of a parameter. When using point estimations, $\hat{p}$ is the best estimate of p .

Ex. The president of a university wishes to estimate the proportion of seniors graduating this year. She samples 1020 of the university seniors and determines that 890 of them will be graduating. She could then use this as the estimate of the proportion of all of the graduating seniors at her university. Therefore, since p-hat = 890/1020=87.3% she could say that of proportion of all graduating seniors at this particular university is 87.3 %

In the previous example the statistic 87.3 % is obtained from a sample and is called an estimator.

Can we expect that the sample proportion will always be a good estimator for the population proportion? The answer is no! There will always be some difference between the sample proportion and the population proportion. This difference is called the sampling error. For this reason statisticians prefer to use a second method for estimating. This is known as an **interval estimate** or a **confidence interval**.

An **interval estimate** of a parameter is an interval or range of values used to estimate the parameter. This interval may or may not include the value of the parameter being estimated.

Ex. The president of a university samples 100 students and estimates that the sample proportion, p-hat, is 68% therefore the president could estimate the population proportion to be between between 63% and 73% .

An **interval estimate** is determined according to how confident we are that it is accurate. The more confidence we want in our estimate the larger the interval or range of values should be. This is called the confidence level.

Confidence Interval: A specific interval estimate of a parameter determined by using data from a sample.

Confidence level: The likelihood that an interval estimate contains the actual parameter. The three most common confidence levels are 90%, 95%, and 99%.

To determine a confidence interval for a population proportion two conditions must be met.

1. $n(\hat{p})(\hat{q}) \geq 10$
2. $n \leq 0.05N$

and the following formula must be used:
Suppose that a simple random sample of size n is taken from a population then

$$\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} < p < \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}$$

Rounding Rule: Round to 3 decimal places

The statistician must decide an appropriate level of confidence when using this formula.

90% level of confidence use $Z_{\alpha/2} = 1.645$
95% level of confidence use $Z_{\alpha/2} = 1.960$
99% level of confidence use $Z_{\alpha/2} = 2.575$

Rounding Rule for Confidence Intervals: If you use raw data then round to one more place than the data itself. If you use sample mean and standard deviation then round to the same number of decimal places as the sample mean

Example: A sample of 500 nursing applications to a nursing program included 60 men. Find the 90% confidence interval of the population proportion of men who applied to the nursing program.

When estimating a population you must give consideration to the size of the sample required in order to get a reliable estimate. The minimum sample size formula is a follows:

$$n = \hat{p}\hat{q}\left(\frac{Z_{\frac{\alpha}{2}}}{E}\right)^2$$

Example: What is the minimum number of subjects needed to estimate, with 95% confidence, the number of people who own a computer given 40% of those interviewed had one at home and the estimate should be accurate to within 2% of the true population proportion.