Topic#6 Estimation

The objective of the survey is to make an inference about a population based on some information you have obtained from a sample.

The population characteristics that you make an inference about are usually the mean, total or variance. These numerical measures of the population are called parameters. These numerical measures of the sample are called statistics (or estimators).

For Example: You might use the sample mean to estimate the population mean. You will have a list of sample means. The sample mean is a random variable with a probability distribution. Some of the sample means will be close to the population mean. Some of the sample means might be very far from the population mean (on either side). We want a sampling plan that ensures that the sample mean generate values that center around the population mean and are close to the population mean.

- If the expected value of the estimator is the same as the population parameter, the estimator is said to be unbiased. $E(\hat{\theta}) = \theta$

- The variance of the sample estimator should be small. We do not have a minimum value in this class, but if we have two unbiased estimators to choose from, we will choose the one with the smallest variance! $V(\hat{\theta})$ is small

Once we know which estimator we are using and how its probability distribution looks (normal distribution?), we can access the error of estimation. The error of estimation is the difference between the estimator and the population parameter to be less than some number "B".

- The probability that the error of estimation is less than or equal to B is $1 - \alpha$, where α is between zero and one. $P(|\hat{\theta} - \theta| \leq B) = 1 - \alpha$; $0 < \alpha < 1$

- If the estimator has a normal distribution, and $1 - \alpha$ equals 95%, then $z = +2$ and -2.

- If the estimator does not have a normal distribution, we can still use $z = 2$ because Chebychev's theorem says that 75% of the data will be between -2 and +2.

(z is the number of standard deviations above or below the mean.)

$(\hat{\theta} - B, \hat{\theta} + B)$ is the confidence interval for θ.

We know that the estimator can be in error by as much as "B", so if we add and subtract B to the estimator, we have a range of values (to estimate the population parameter) known as a confidence interval.

Topic#7 Summary

This section of material was a review of basic statistics. The following bullet point list summarizes the topics that will be covered on the next exam. Computations regarding mean, standard deviation and variance may be required. The more rigorous formulas for estimators and regression will not be tested; however, the lessons that you learned in the project will be assessed.

- The mean (a measure of center) and its calculation
- The standard deviation and variance (measures of spread) and its calculation
- Normal distributions, Chebychev's Theorem (Normal 68% 1SD, 95% 2SD) (Chebychev $1-\frac{1}{K}$ %)
- Sampling distributions and how samples means appear as sample size increases
- The mean as an estimator and the shape of its distribution
- Covariance and Correlation, r values & what they mean
- Estimators and how they are evaluated (unbiased, how we choose them)

Sketch a scatter plot of a positive, negative & no correlation.

Sketch a graph of a normal and skewed distribution.

Error of estimation and confidence intervals.