Unit 4 Topic 1: Confidence Intervals and Sample Size

Estimators

- A **point estimate** is a specific numerical estimate of a parameter. The best point estimate of the population mean μ is the sample mean X
- Sample measures are used to estimate population measures. These statistics are called **estimators**.
- A good estimator should satisfy these three properties:
  - The estimate should be an **unbiased estimator**. That is, the expected value or the mean of the estimates obtained from samples of a given size is equal to the parameter being estimated.
  - The estimator should be consistent. For a **consistent estimator**, as sample size increases, the value of the estimator approaches the value of the parameter estimated.
  - The estimator should be a **relatively efficient estimator**. That is, of all the statistics that can be used to estimate a parameter, the relatively efficient estimator has the smallest variance.
- For this reason, statisticians prefer *interval estimates*.
  - An interval or range of values used to estimate a parameter; may or may not contain the value of the parameter being estimated.

Confidence Intervals

- Either the interval contains the parameter or it does not. A degree of confidence (usually a percent) can be assigned before an interval estimate is made.
- For instance, you may wish to be 95% confident that the interval contains the true population mean. A 99% confidence interval would be wider than a 95% confidence interval since to be more confident, you must expand your range.
- The **confidence level** of an interval estimate of a parameter is the probability that the interval estimate will contain the parameter, assuming that a large number of samples are selected and that the estimation process on the same parameter is repeated.
  - E.g. 95%, 99%, etc.
- The **confidence interval** is a specific interval estimate of a parameter determined by using data obtained from a sample and by using the specific confidence level of the estimate.
  - E.g. 23 ± 3; 87 < μ < 97

Formula for the Confidence Interval of the Mean for a Specific α:

$$\bar{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

90% confidence............... $z_{\alpha/2} = 1.65$

$\bar{X} =$ point estimate or given mean

95% confidence................ $z_{\alpha/2} = 1.96$

$\sigma =$ standard deviation of population

99% confidence................ $z_{\alpha/2} = 2.58$ $n =$ sample size

How do we find the $z_{\alpha/2}$??

Example 1:

- A researcher wishes to estimate the number of days it takes an automobile dealer to sell a Chevrolet Aveo. A sample of 50 cars had a mean time on the dealer's lot of 54 days. Assume the population standard deviation to be 6.0 days. Find the best point estimate of the population mean and the 95% confidence interval of the population mean.

Sample Size

- Sometimes we need to know how large a sample must be in order to accurately represent a population.
- Determining an appropriate sample size depends on 3 things:
  - Maximum error of the estimate
  - Population standard deviation
  - Degree of confidence
- When we know the population standard deviation.
  - The formula for finding maximum margin of error is…

$$E = z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

So we just solve this equation for n (sample size) and get…

$$E\sqrt{n} = z_{\alpha/2}(\sigma)$$

$$\sqrt{n} = \frac{z_{\alpha/2}(\sigma)}{E}$$

$$n = \left(\frac{z_{\alpha/2}(\sigma)}{E}\right)^2$$ **Always round up to the next whole number!

This is the formula for the minimum sample size needed for an interval estimate of the population mean.

Example 1

- A scientists wishes to estimate the average depth of a river. He wants to be 99% confident that the estimate is accurate within 2 feet. From a previous study, the standard deviation of the depths measured was 4.38 feet. At least how many measurements does he need?