# Module 6a – Sampling Distributions

Review Against All Odds: Unit 22 (Sampling Distributions)
Review Against All Odds: Unit 24 (Confidence Intervals)

# Stat Procedure Diagram – Where we are

| | | | | |
|---|---|---|---|---|
| | | **Descriptive Statistics (Describing Pops or Samples)** | | **Inferential Statistics (from Samples)** |
| | Variable Types | Display | Describe | Estimation |
| Univariate | categorical (nominal or ordinal)* | Bar Graph/Pie Chart | Counts/Percentages | *When binary/dichotomous:* Confidence interval for proportions |
| | quantitative/continuous | Histogram/Stem & Leaf<br>Box Plot | Mean/St Dev (normal)<br>Median/Min, Q1, Q3, Max (skewed) | Confidence interval for means |
| | | Display | Describe | Significance Tests/Hypothesis Tests |
| Bivariate | 2 categorical | Tables or Bar Graphs | Two-way tables/Crosstabulation | Chi-square test (for goodness of fit) |
| | 1 categorical, 1 quant. | Bar Graphs | Comparison of means/averages | T-test (one sample/group, two samples/groups)<br>ANOVA (two or more samples/groups) |
| | 2 quant. | Scatterplot | Correlation Coef. (Coef. of determ)/ Regression Line | T-test for correlation |
| | | Display | Describe | Significance Tests/Hypothesis Tests |
| Multivariate | Response Variable is Quant. | - | Ordinary Least Squares Regression (OLS) | F-test for overall model<br>T-tests for each explanatory variable |
| | Response Variable is categorical (dichotomous) | - | Logistic Regression | Chi-square tests of significance |

NOTE: Items highlighted in yellow are covered in this course.

**When a categorical variable has two categories, it is called dichotomous.*

# How does sampling work?
(focus on quant/continuous variables)

- Statistics describe samples
- Parameters describe populations (of interest)

- Law of Large Numbers: If you draw observations (your sample) at random from a large population, ***as the sample size increases***, the mean of the sample tends to get closer to the mean of the population.

# Population Distribution vs Sampling Distribution

- Population Distribution: The distribution of the actual values of a population
- Sampling Distribution: The distribution of sample means taken from a population
  - Imagine that you take multiple samples from the population: the means of those samples tend to reflect the mean of the population.

- Leads to: ***Central Limit Theorem***
  - The sampling distribution from a randomly selected sample (when the sample size is sufficiently large) is approximately normal with a mean that approximates the population mean.

# How this applies to samples

- We use the mean of a sample (statistic) to approximate the population mean (parameter)
- We use the standard deviation of a sample (statistic) divided by the square root of the sample size to determine how close our estimate is to the population mean (parameter)

- The second bullet point describes the ***Standard Error*** (aka the standard deviation of the sampling distribution)

# Standard Error

$$SE = \frac{\sigma}{\sqrt{n}}$$

- We can apply our rules about the normal curve to a sampling distribution in the same way we applied standard deviation to the normal curve for population distributions
(e.g. applying the 68/95/99.7 rule)

- In other words, we can use the sample mean and standard error to estimate the population mean.

# Standard Error Example

$$SE = \frac{\sigma}{\sqrt{n}}$$

- A bottling plant fills bottles with an average of 12.2 ounces of soda and a standard deviation of 0.2 ounces.

- We select a sample of 4 bottles. What is our standard error?
- We select a sample of 16 bottles. What is our standard error?

- How does this reinforce the law of large numbers?