

Chapter 4 – Understanding Your Data and Checking Assumptions

Chapter Outline

- I. Exploratory Data Analysis (EDA)
 - A. What is EDA?
 - 1. The first step to complete after entering data and before running any inferential statistics.
 - 2. Computing various descriptive statistics and graphs in order to examine your data.
 - a. Look for data errors, outliers, non-normal distributions, etc.
 - b. Determine if the data meets the assumptions of the statistics you plan to use.
 - c. Gather basic demographic information about the subjects.
 - d. Examine relationships between the variables to determine how to conduct the hypothesis testing.
 - B. How to do EDA
 - 1. Generate plots of the data
 - 2. Generate numbers from the data.
 - C. Check for Errors
 - 1. Examine raw data before entering.
 - 2. Compare some raw data against entered data.
 - 3. Compare maximum and minimum values against the allowable ranges.
 - 4. Examine the means and standard deviations to see if they seem reasonable.
 - 5. Look to see if there is an unreasonable amount of missing data.
 - 6. Look for outliers.
 - D. Statistical Assumptions: explain when it is and isn't reasonable to perform a specific statistical test.
 - 1. Parametric tests
 - a. Usually have more assumptions than nonparametric tests.
 - b. Generally designed for use with data that exhibits approximately normal distribution
 - c. Some parametric tests are more robust in dealing with violations of assumptions than others.
 - 2. Nonparametric tests
 - a. Have fewer assumptions
 - b. Can often be used when assumptions for parametric tests are violated.
 - E. Parametric Tests
 - a.