Lab Module 1 & 2 Review Handout

Important Concepts: Lab Module One

- Organization of Data:
  - Observation: the unit upon which measurements are made (Row)
  - Variables: characteristics being measured (Columns)
  - Values: realized measurements (Table Cells)

- Types of Measurements:
  - Categorical (SPSS: Nominal): Place observations into classes or groups (i.e. Sex, Disease Status, Blood Type, etc.).
  - Ordinal: Assign observations into categories that can be put into rank order. It is important to understand that ordinal variables serve merely as a ranking and do not quantify differences (i.e. Stage of Cancer: I, II, II).
  - Quantitative (SPSS: Scale): Position observations along a numeric scale (i.e. age, glucose level, etc.).

- GIGO: "Garbage In, Garbage Out" emphasizes the importance of proper data collection.

- Types of Measurement Errors:
  - Imprecision: the random inability to get the same results upon repetitions.
  - Bias: Systematic deviation from the truth. When something is unbiased, it is valid (i.e. Undercoverage, Volunteer bias, and Non-response bias)
    - See Figure 1.2 in Ch. 1 and page 19 in Ch.2

- Suggested Practice Problems:
  - 1.1, 1.2, 1.5, 1.6, 1.7, 1.8, 1.10, 1.11 (All from Ch.1)

Important Concepts: Lab Module Two/Ch. 3, & 4

- Stem-and-Leaf Plots
  - Graphical technique that organizes data into a histogram-like display.
  - Splitting Stem Values
  - Truncating Leaf Values
  - Quintuple Split of Stem Values

- Aspects of the Distribution:
  - **Shape:** Refers to the configuration of data points on the graph.
    - Symmetry: refers to the degree to which the shape reflects a mirror image of itself around its center.
    - Modality: refers to the number of peaks on the distribution
    - Kurtosis: refers to steepness of the mound
    - Note: Positive skew (tail toward larger numbers on a number line) and negative skew (tail toward smaller numbers)

        - Know: 1) When mean = median → distribution is symmetrical 2) When mean > median → positive skew. 3) When mean < median → negative skew. See Figure 4.4 on page 71.
        - Outlier: A striking deviation from the overall pattern of the distribution.
    - **Location:**
        - The arithmetic average or gravitational center is a distribution's balancing point.
        - The median is its middle value (n+1 / 2)

    - **Spread:** Informal way to refer to the dispersion or variability of data points.

- Frequency Tables
    - Frequency column contains counts
    - Relative frequency (%) column contains frequency counts divided by the total, with values expressed as a percentage.
    - Cumulative frequency (%) column contains percents that fall within or below a given level.
        - See page 53 for examples.

- Quartiles, 5-Point Summary, Interquartile Range
    - Five-Point Summary
    - IQR

- Boxplots
    - Calculate the lower and upper fence values
    - Identify: Lower and upper outside values
    - Identify: Lower and upper inside values
    - Draw boxplot

- Calculating Variance and Standard Deviation
    - Sum of Squares
    - Variance
    - Standard Deviation
    - Know some basic facts about SD: Page 80, points 1, 3, and 5.

Midterm One Preparation:

- Understand all of the concepts from the labs
- Re-do HW problems and also any new problems for practice