

Lab 2: Descriptive Statistics & Graphical Displays II, and the Normal Distribution

OBJECTIVES: This lab is designed to show you how to calculate basic descriptive statistics & generate some statistical plots. In particular, the histogram, stem plot, and boxplot will be studied and used to answer questions pertaining to the center and spread of a distribution and to the detection of outliers. Further study to the normal distribution will also be covered, including the standardization of, and application of linear transformations to, observations from a normal distribution.

DIRECTIONS: Follow the instructions below, answering all questions. There are some Minitab instructions (help) on the questions. For instance, (Graph-> histogram) means you should go to menu "Graph" first. Then go to the submenu "histogram". Your answers for each of the two questions, including any plots, should be summarized in the form of a brief report (in Word), to be handed in to the instructor by the end of your assigned lab time.

1. Introductory Exploratory Data Analysis . . .

a.) Open the worksheet "grades.mtw" in Minitab and look at the histograms for both the verbal and math scores (Graph-> histogram).

- Note whether the data in each plot is symmetric or skewed, and if skewed, to what direction.
- Are there any observations one would guess are potential outliers? If so, why?
- What is the approximate center of the data?

b.) For each of the two variables (verbal scores and math scores), generate a stem-and-leaf plot (also called a "stemplot") (Graph-> ...) .

- Again, note the symmetry or skewness of this data from this plot. Are outliers evident with this type of plot? (*Hint!:* After looking at the original stemplots, consider checking the "trim outliers" option and note what happens).
- How is the stemplot different from the histogram? In what ways is it similar to the histogram? In particular, be sure to note how the data is depicted, if any additional information about the data is given in the plot, and the advantages / disadvantages associated with this type of plot.
- Further, consider whether or a histogram or stemplot would be better suited for large data sets.

c.) For each of the two variables being studied, generate a boxplot.(Graph-> boxplot)

- What information about the data do the boxplots give that other graphs do not? Describe what each "feature" of the boxplot indicates. (*Hint!:* What famous "summary" of data is associated with this type of plot?)
- What are the advantages / disadvantages of this type of plot?

d.) Finally, calculate and record the basic summary statistics of each of the two variables (Stat-> Stat).

- Summarize the results, including in general, your insights into choosing the mean or median as a measure of location, and how the skewness of the data affects that choice. That is, consider whether the five-number summary, or the mean and standard deviation, is better for describing different types of distributions (e.g., symmetric vs. skewed, outlier vs. outlier-free).

2.) Normally Distributed Data Simulation and Transformation / Standardization . . .

a.) Start a new Minitab worksheet and label the first column "Norm1".
(*Note: Do not close the "grades.mtw" worksheet just yet!*)

b.) Generate a sample of 100 values taken randomly from a *standard normal distribution* (Calc-> Random Data).

- *Recall!:* What constitutes a standard normal distribution (i.e., what is the mean and variance of a standard normal distribution)?
- Generate a histogram of the "Norm1" data you just created. Do the values indeed appear to be normally distributed, and if so, what features of this histogram lead you to this conclusion (e.g., overall shape/appearance)?
- Finally, scale the "Norm1" data so that the standard deviation is now 5 and the mean is now 10.
- (*Hint!:* Using the "Calculator" option in Minitab, create another column of data called "Norm2" (to be stored in C2 (or some empty column) of your worksheet), where "Norm2" will be given by the expression,
 $\text{Norm1} * 5 + 10$)
- Generate the histogram of "Norm2" now, and note the difference between this histogram and the one for "Norm1".

Bonus!

A Look Ahead . . . Studying the Relationship Between Two Variables . . .

Using the "grades.mtw" worksheet from question 1, produce "scatterplots" of the verbal scores vs. gpa, math scores vs. gpa, and verbal scores vs. math scores. What can you say about the relationships between these variables?