

## Quantitative Methods (GEO 441)

### SPSS Mini-Lab 1: Normality Testing

Dr. Paul Marr

Please copy the file **S:\GEO\Marr\Quantitative Methods\SPSS Example Data\Normality Testing.sav** to your portable media.

#### 1. Normality Testing

- a. Open file *Normality Testing.sav*
- b. Histograms (create the following histograms, *Graphs > Legacy Dialog > Histogram*)
  - i. Histogram 1 (Normal with normal curve)
  - ii. Histogram 2 (Leptokurtic with normal curve)
  - iii. Histogram 3 (Skewed with normal curve)
  - iv. Histogram 4 (Extremes with normal curve)
    - The normal curve allows you to gauge how far or close to normal your data are.
    - While this is not a statistical test, it is helpful in understanding the 'shape' of your data.
    - Shape is important since parametric tests work best (lower chance of a Type 1 Error) when the data are symmetrical.
- c. QQ and PP Plots (create the following plots)
  - i. *Analyze > Descriptive Statistics > QQ Plot* (a good general plot)
  - ii. QQ Plot 1 (Normal)
  - iii. QQ Plot 2 (Skewed)
  - iv. *Analyze > Descriptive Statistics > PP Plot* (shows differences close to the mean)
  - v. PP Plot 1 (Normal)
  - vi. PP Plot 2 (Skewed)
    - Both Q-Q and P-P plots also give you a 'detrended' plot. The x-axis shows the observed values, the y-axis shows standard deviations from the mean.
    - Detrended plots are useful in picking out potential outliers. Note the detrended plot for the normal data. It is random about the line. The skewed plot appears to follow a sine curve which is characteristic of non-normal data.
- d. Normality Tests
  - i. *Analyze > Descriptive Statistics > Explore, Display = Plots, Plots = Normality plots with tests*, also check the *Histograms* box and uncheck the *Stem-and-Leaf* box.
    - **IMPORTANT: Do each test independently** (do not add all variables to the *Dependent List*)!!!
  - ii. Test 1 (Normal)
  - iii. Test 2 (Leptokurtic, note the negative results)
  - iv. Test 3 (Skewed, note the inconsistent results, KS positive and SW negative)
  - v. Test 4 (Extremes, note the positive results)
  - vi. Test 5 (Platykurtic, note the inconsistent results, KS negative and SW positive)
    - The K-S test found the extremes and skewed data non-normal. This is likely caused by the test's preference for the area near the mean. The histograms show both data set are skewed.

- The Shapiro-Wilks test found the platykurtic and extremes data non-normal. This is likely due to the test's preference for activity in the data tails.

Notes:

- Neither the Kolmogorov-Smirnov nor the Shapiro-Wilks tests are particularly powerful.
- In fact, according to D'Agostino the K-S test should "never" be used and is merely an "historical curiosity".
- The Shapiro-Wilks test is a bit more powerful, but both are not particularly good at weeding out what data sets are truly normal.
- Remember that the K-S test focuses on the center of the data, so if you suspect problems in this area, this is an ok test to rely on.
- The Shapiro-Wilks test focuses on the tails, and again, if you suspect problems in the tails, this is an ok test.
- However, the D'Agostino test is good for both the center and tails, but must be calculated by hand until SPSS includes it as a test (which is unlikely).



Good news everyone!

Lawrence DeCarlo (Columbia University) has written a macro that will run the D'Agostino Omnibus test in SPSS. The file is can be downloaded from my webpage, along with instructions.