FI=NOT0101B

A COUPLE OF NOTES ON DESCRIPTIVE DATA DISPLAYS

NOTE: For help on any of these, pull up the help in SPSS and read it, or print it out.

Stem and leaf plots

"The difference between adjacent stems is a very round number, such as (1×10^n) or (2×10^n) or (5×10^n) where 10^n represents 10 to some power, e.g., 10, 1, .1, .01, 1,000, 10,000 and so on."

The leaf shows the next digit after the stem. So if the stem is 100,000, then the adjacent leaf values will range from 0 ten thousands to 9 ten thousands.

You need to look at the entire range to find out what the units are that are presented on the stem, and what units are presented on the leaves. **In short, you need to know about the overall distribution if you are to interpret the stem and leaf diagram properly**.

The "leaf" lines in the display do not round values up or down. They show the exact value for that.

Stem and leaf plots are not available as a graph, but instead are available in the EXPLORE procedure.

Box plots and Outside values

Box and whisker plots, and S&L plots, use a consistent definition for a case that is an "outside value."

The interquartile range (25th - 75th percentile) can also be called Hspread; spread between the hinges. In the box plot this is the two ends of the box.

The whiskers spread beyond the box the following amount:

+ or - (1.5 * Hspread)

Therefore you can compute the point where the left whisker ends:

lower hinge (25th percentile) - (1.5*Hspread)

The point where the right whisker ends is:

upper hinge (75th percentile) + (1.5*Hspread)

Points above or below these points are flagged as outside values (*) or far outside values (O). SPSS calls these outliers and extreme values.

Normal P-P and normal Q-Q plots

Hamilton explains these. They are excellent tools for exploring the normality or non-normality of data, on a variable by variable basis. What you are looking for with these is departures from normality, manifested as data points away from the 45 degree diagonal.