SPSS Introduction

Data Entry, Descriptive Statistics,
and Graphs

SPSS 14.0

Note that there is an SPSS manual on the Utts/Heckard CD. This may be of help, but be warned that it is for SPSS 13.0. There may be some differences.

The R Drive

Often we will use SPSS on data sets from the textbook. These datasets are stored on GVSU PCs under the R drive. The R drive needs to be "activated" before you start SPSS in order to open files from it into SPSS. If you start SPSS first, the R drive won't show up and the files will not be visible.

To access the R:\ drive in Windows (XP), you must double-click on the 'R: Data' File Folder within the 'Statistics' folder before opening SPSS. To do this, follow these steps:

3.   Double-click the *Departmental Applications* folder.

4.   Double-click on the *Statistics* subdirectory folder.

5.   Double-click on the *R: Data* file folder.

6.   Notice the *STA 215 3e Utts CD* folder within R. This is the textbook data. Also notice the *Zeitler* folder. I will put datasets here from time to time.

7.   Close the R drive; we will access it from SPSS instead.

Starting SPSS

In the *Statistics* folder, double-click on the *SPSS 14* icon. This will bring up a dialog box asking where the data is coming from.

Nothing can be done without data. One can open an existing SPSS data set (a data set with a *.sav* extension), an existing Excel spreadsheet (*.xls* extension), or type in a new data set. The textbook data files are stored as SPSS files.

Opening Datasets from the R Drive

1. Click *OK* when the initial SPSS dialog screen opens.

2. Change the *Look in:* box to the drive you want (something like Labdata ...Stat... **R**)

3. Open the *STA 215 3e Utts CD* folder, then the *SPSS Data Sets* folder. Or if appropriate, open the *Zeitler* folder, then the 215 *Data* folder.

4. Double-click the name of the data set you wish to open.

SPSS Data Editor

Once you have opened a data file into SPSS, notice that you have two tabs on the bottom: *Data View* and *Variable View*.

Under *Variable View*, note that *Values* can contain information on categorical variables. For instance, 0=male, 1=female.

Also realize that it is necessary to display 2 or 3 decimal places for all quantitative variables. Otherwise, subsequent output will round all statistics to whole numbers. (By default – you can specify otherwise when you request a table, etc.)

Descriptive Statistics

There is more than one way to obtain descriptive statistics with SPSS. First of all, it depends on what type of data you have, and what kind of summary measures you desire. A few examples follow.

Example 1: Categorical variables (one at a time)

➔ Pull down the *Analyze* menu.

➔ Choose *Descriptive Statistics*, then *Frequencies*.

➔ Move the desired variable names into the *Variables* box.

➔ Make sure that the box next to *Display frequency tables* is checked. Don't use any other options. Click OK.

Example 2: Two categorical variables (summarizing frequency counts with a contingency table)

➔ Pull down the *Analyze* menu.

➔ Choose *Descriptive Statistics*, then *Crosstabs*.

➔ Move the desired variable names into the *Row* and *Columns* boxes.

➔ Click OK.

Example 3: Quantitative variables

Note: *Analyze/Descriptive Statistics/Descriptives* gives a very limited selection of summary statistics. What follows gives much more (maybe too much!)

➔ Pull down the *Analyze* menu.

➔ Choose *Descriptive Statistics*, then *Explore*.

➔ Move the desired quantitative variable name into the *Dependent List* box, and if appropriate, the desired grouping variable name into the *Factor List* box. Note: You will only place a variable in the Factor List box if you wish to have descriptive stats broken down by gender, for example. Click OK.

Note: This will produce plots and more info than we need, most likely. Some of the options can be suppressed (turned off).

<u>Example 4:</u> Quantitative variables (again)

What follows is a good way to get most of the descriptive statistics we have learned so far in a neatly organized table. It's easy to produce, and won't result in anything we don't want.

➜ Pull down the *Analyze* menu.

➜ Choose *Descriptive Statistics*, then *Frequencies*.

➜ Move the desired variable names into the *Variables* box.

➜ Click on *Statistics…*

➔ After clicking on the *Statistics* button, pick and choose the stats you want. For example: Quartiles, Standard Deviation, Range, Min, Max, Mean, Median. Un-select "Display frequency tables". Click Continue, then OK.

➔ UNclick the box next to Display frequency tables. This option only makes sense for categorical variables.

This produces a pretty decent table, but there are two problems with it:

➔ We are not able to break down by a categorical variable

➔ We can't control how the stats appear in the table (this is a problem, because when I ask for a five number summary, you need to list it <u>in order</u>!)

Example 5: Quantitative variables (last time!)

This method is harder to do, but will give us exactly what we want.

➔ Pull down the *Analyze* menu.

➔ Choose *Tables*, then *Basic Tables*.

➔ Move the desired variable names into the *Summaries* box.

➔ Move the name of a grouping variable (if appropriate) into the *Subgroups Down* box.

➔ Click on *Statistics…*

➔ After clicking on the *Statistics* button, pick and choose the stats you want from the *Statistics* list on the left, moving them over to *Cell Statistics* on the right by clicking *Add*. Be sure to choose the stats in the order you wish them to be listed in the table. Click *Continue*.

➔ Most likely, the table will be too wide to fit on the page. You can switch the rows and columns of the table like so. Click the *Layout…* button. Change the *Summary Variable Labels* to *Across the top*. Change the *Statistics Labels* to *Down the left side*. Click *Continue*, then *OK*.

➔ The variable names will not appear in the table, for some inexplicable reason. You will need to double click on the table where the labels should go, then type in the names in the appropriate place.

➔ Finally, create some *Titles* if you wish, and click *OK*.

Pie Charts

To create a pie chart for a categorical variable, do the following.

➔ Pull down the *Graphs* menu and choose *Pie…*

➔ By default, *Summaries for groups of cases* will be chosen. This is what we want. Click *Define*.

➔ You can choose the slices of the pie to represent percentages of the total, or frequency counts. *N of cases* is the default; change to *% of cases* if desired.

➔ Move the name of the categorical variable you want to illustrate with a pie chart into the *Define Slices by:* box.

➔ Type in a title under *Titles* if so desired. Click *OK*.

The output window will show the pie chart without any of the frequencies (or percentages). This is not very desirable. Follow the steps below to fix this.

➔ Double click on the graph in order to bring up the *Chart Editor*.

➔ In the *Chart Editor*, click on the *Data ID Mode* icon (looks like a square target). The mouse pointer will change into the icon. Click on each slice to add the frequencies (or percentages).

➔ The new info will echo back to the output window. The graph can be copied and pasted into Word. Use *Copy* and not *Copy Objects,* especially for graphs.

There is another approach to labeling the pie slices. Either way can be problematic – often the labels will overlap, and it's not clear how to specify decimal places and whether or not the % symbol appears (extra credit to the student that figures this out!)

➔ In the *Chart Editor*, click on the pie somewhere. This will make it glow a lovely purple color. Once you have done this, more options become available.

➔ Choose the button that looks like a bar chart (Show data labels). You may need to change *Suppress overlapping labels* to *Display all labels*.

Bar Graphs

To create a bar graph for a categorical variable, do the following.

➜ Pull down the *Graphs* menu and choose *Bar…*

➜ *Simple* is the default type; leave as is.

➜ By default, *Summaries for groups of cases* will be chosen. This is what we want. Click *Define*.

➜ You can choose the bars of the graph to represent percentages of the total, or frequency counts. *N of cases* is the default; change to *% of cases* if desired. We also have cumulative (freq or %) options.

➜ Move the name of the categorical variable you want to illustrate with a bar graph into the *Category Axis:* box.

➜ Type in a title under *Titles* if so desired. Click *OK*.

Stem-and-Leaf

To illustrate quantitative data with a stem-and-leaf display, do the following.

➔ Pull down the *Analyze* menu.

➔ Choose *Descriptive Statistics*, then *Explore…*

➔ Move the desired quantitative variable name into the *Dependent List* box. If appropriate, move a grouping variable into the *Factor List* box.

➔ In order to just obtain the stem-and-leaf plot without a great deal of extra output, choose *Plots* under *Display*.

➔ Click on the *Plots…* button.

➔ Under *Boxplots*, choose *None*. Under *Descriptive*, choose *Stem-and-Leaf*, but not *Histogram*. Click *Continue,* then *OK*. (Ignore the *Case Processing Summary* table.)

Histograms

To illustrate quantitative data with a histogram, do the following.

➜ Pull down the *Graphs* menu and choose *Histogram...*

➜ Move the desired variable name into the *Variable* box.

➜ Type in a title under *Titles* if so desired. Click *OK*.

Note: Histograms (and bar charts) can be fine-tuned in the chart editor, but it is not necessary like it is with pie charts.

Boxplots

You may want to produce a single boxplot, or compare groups with multiple boxplots.

➔ Pull down the *Graphs* menu and choose *Boxplot...*

➔ *Simple* is the default type; leave as is.

➔ By default, *Summaries for groups of cases* will be chosen. This produces side-by-side comparison boxplots. *Summaries of separate variables* should be chosen if a single boxplot is desired, and you are not trying to compare groups. Click *Define*.

➔ Move the quantitative variable name into the *Variable* box. This is the data you are plotting.

➔ If you are comparing groups, move the categorical variable name into the *Category Axis* box. Ignore *Label cases by:*. Click *OK*.

Printing SPSS Output

In general, do not simply print whatever happens to be in the output window. Points will be taken off assignments if you include unnecessary output.

Option 1: Select and delete whatever you don't want from the output window before printing.

Option 2: Multiple select (use the control button) to highlight what you want before printing.

From the output window, pull down the *File* menu and choose *Print*. Using *Print Preview* first is a good idea.

Option 3 (best): Copy and paste what you want into Word. This way, you can re-size tables and graphs so they fit on the page, control how much paper you are using, and create nice-looking reports. Pull down the *Edit* menu. *Copy* and *Copy objects* work differently. You will have to experiment to see what works best for tables and graphs.