**How to open an already created SPSS file**

We will open the file “Workshop Results End File” in SPSS to look at the various components of

the SPSS menus:

* Open SPSS (Lab Applications folder, then Statistics folder, then SPSS 20 or SPSS 22)
* Choose File, then Open, then Data.



* A new window will open:



* Use the “drop down” window next to “Look in” to find the folder where the file is located.
* When you reach the file, click on the file to get the name of the file in the “file name” box, and then click the Open button.

**An overview of the various menu functions of SPSS**

Here we will click on each of the menu functions in SPSS and briefly discuss what they do:

File: Used to create files, open files, and save files.

Edit: Used to cut, paste, clear, and insert variables; also to find data values and find cases.

View: Lets you set how the window looks, etc.

Data: Helpful for sorting observations, choosing a subset of observations, merging files, and for

weighting cases.

Transform: Used to create variables by arithmetic or other operations (e.g., Var3 = var1 + var2),

recoding variables (e.g., putting numeric values into categories).

Analyze: For obtaining descriptive summaries and performing various hypothesis tests and

finding confidence interval estimates.

Graphs: Useful for finding graphical summaries.

Utilities: Various other functions; not sure if you will have occasion to use them.

Add-Ons: Various other statistical functions available in SPSS.

Window: If you want to see multiple SPSS windows at once, this allows you to do so.

Help: Topics and Tutorial are a resource to find out how to do particular processes in SPSS.

**Kinds of SPSS files**

There are three kinds of SPSS files. Their “logos” indicate what they do:

* Data files:  \*.sav file extension. The **data file** is where the data will be. We can create variable and value labels to better “document” our data.
* Output files:  \*.spv file extension. The **output file** includes SPSS syntax to do what we will do here. If we want to do all of this again without the “point and click” we will do today, we’ll want to keep this file so that we can put the code into a syntax file. Also, any graphs and tables we create will be in this file.
* Syntax files (for running SPSS code/programs):  .\*sps extension.

**To Create an SPSS file from scratch**

In the bottom left corner of the SPSS window are two tabs: Data view and Variable view.

* The Data View window is used to enter the data by hand
* The Variable View window is used to set up the columns for entry into Data View.

First, click on the variable view tab. The top of the window looks something like this:



We will walk through each column in Variable View from left to right:

NOTE WELL: Each COLUMN (Variable) in Data View has its own ROW in Variable View.

**Name:** The variable name that appears at the top of the column.

* SPSS variable names can be up to 32 characters in length.
* The first character in a variable name must be a letter
* Subsequent characters in a variable name can be a letter, a numeral, a period (.), a dollar sign ($), the “at” (@), or an underscore.

**Type:** Primarily SPSS sees variables as either String (aka character) or numeric.

* If you want to use another format for a variable, click on the far right end of the Type box to see this:



and then this:



At this point if you want to change something you can.

**Width:** The width of the variable.

* For string variables IMPORTED from Excel, the width is the same as the longest item in the column. For string variables CREATED IN SPSS, the default width is 8.
* For Numeric variables IMPORTED from Excel, the default width is 12. For numeric variables CREATED IN SPSS, the default width is 8.
* To change the width for a variable, simply click in the “Width” box corresponding to the variable and change it. *When entering the data by hand, if the longest data entry for the variable requires more than 8 spaces you MUST increase the width to at least the number of spaces in that longest entry.* LOOK for the longest data entry for the particular variable and count the number of characters (including spaces) for that data entry. If this entry requires more than 8 spaces then increase the width accordingly.

**Decimals:** For string variables the decimals cannot be adjusted: They are irrelevant.

For imported numeric data SPSS uses the data to set the number of decimals.

For numeric variables created in SPSS, the default is *2*.

*For data entered directly in SPSS, find the data entry with the most decimal places and set the number of decimals to that highest number of decimal places*.

For categorical data set decimals at 0.

**Label:** This column is great for documenting variables. A label on the variable allows you to describe the variable. *The label will be used in output*. To put label on the variable, type the variable in the white box corresponding to the variable. Here, we will use the descriptions of the data set on the first page to set the labels.

**Values:** This column is great for documenting the VALUES of variables. Sometimes a variable may be entered as coded data. You may want to put a label on the value of the variable that is more descriptive than the code. For example, the entries might be 0 and 1 representing No and Yes, respectively. In the output it would be nice for No and Yes to appear instead of 0 and 1. To put values on a variable, click on the far right edge of the box to see this:



and then this:



Now follow these steps for EACH variable for which you want to enter labels for the values of the variable:

(a) Type the first value of the variable in the *Value:* box. Hit the *Tab* key. The cursor should be in the

 *Label* box.

(b) Type the explanatory label for the value of the variable. Hit the *Tab* key. The *Add* box should be

 highlighted.

(c) Click the *Add* box (or hit the *Enter* key). The cursor should be back in the *Value* box.

Repeat (a,b,c) for each value of the variable.

**Missing:** Sometimes certain values of a variable imply that in reality there was no response. Instead of leaving it blank, especially if there were different reasons for a no response, different numbers may be used (e.g., 99 may mean “chose not to respond” while 98 many mean “question is not applicable to the person”). The missing column allows the data entry person to tell SPSS what values of the variable are to be treated as ‘missing’ (e.g., not used in tabulating results).

For numeric variables, when a value is not entered SPSS will automatically consider the value to be missing. Also, SPSS will automatically put a period for the missing value. To create a missing value for a numeric variable:

(1) Go to each of the missing values in the data set and put in a space.

(2) Click on the far right edge of the “Missing” box corresponding to the variable so that you see this:



and then this:



(3) Choose what you need and enter the values accordingly. When done, click “OK”.

For string variables, if you have specific values for the variable that represent missing values, follow instructions (1) through (3) above.

How to handle when spaces for a string variable are missing values: When a value is not entered SPSS does not automatically consider the value to be missing. Instead, the value is considered to be “Valid” and is counted as “Valid” for frequency and crosstabs tables. Therefore, if for an observation for a string variable you have nothing entered but want SPSS to treat it as missing you MUST:

* Put a space or some other character for the variable and
* Declare it as missing in SPSS.

To do this, follow instructions (1) and (2) above. Choose “Discrete missing values”. In the first box put in a space. Finally, click OK.

**Column:** Similar to Length. We will not worry about it here.

**Align:** Whether data is left-justified or right-justified. Numeric data is usually right-justified and string data is usually left-justified.

**Measure:** Denotes the kind of data. Nominal and ordinal are used for categorical data. Scale is used for all quantitative data.

For string variables you can choose “Nominal” or “Ordinal”; when choosing ordinal, values are ordered alphabetically. SUGGESTION: For ordinal variables, it is probably smart to make those variables numeric and then use labels to explain what they mean. We did not do this here for Rank and for High\_School, but probably should have.

For numeric variables you can choose “Nominal”, “Ordinal” or “Scale”. If your numeric values are categorical (e.g., Likert Scale data), you should set the measure as “ordinal”.

**Role**: We will not use it here; your instructor has not had occasion to use it.

**Saving our files**

We should save our files often so that we do not lose the work we have already done. Note that there are two windows: A data window and an output window. These are two files with different file extensions.

To save each file (separately!) click **File,** then **Save As** in the upper left corner of the window. A new window will open with **Save Data As** at the top. At this point:

* Use the drop-down next to “Look in” to go to a drive/file/folder that YOU can write to. This will depend on your computer and your operating system.
* Change the file name to one that you want to use. BTW, you can use the same name for both the data file and the output file: The files will be different because the file extensions will be different. The data file will have file extension \*.sav, and the output file will have file extension \*.spv.

NOTE WELL: As long as both windows are open, the results from work done in the data file will be saved in the same output file.

**We will now create an SPSS file with the following items in it.** Note that:

* Gender: 1 = Male, 0 = Female
* Age: 99.00 denotes a missing value



This is what our Variable view (at least the first few columns) will look like; we will explain as we

go why this is the case.



**To import an Excel file**

We first want to open the Excel file that we want to import. Open the Excel worksheet **Sheet1** in the file *N:/SPSS workshop 1 folder/Workshop file.xlsx.*  See the first lines of the data set below:



Note the following:

* The first line of the Excel file contains the variable names.
* Sex: M = Male, F = female
* Rank=class rank: FR = freshman, SO = sophomore, etc.
* High\_school = distance from the high school where they graduated: A = < 50 miles, B = > 50 miles but in MI, C = high school outside of MI.
* Speed = fastest speed one has ever driven a car.
* Seed = how far someone has spit a sunflower seed.
* Q1 = Question with Likert Scale answers: 1 = strongly agree, 2 = agree, 3 = neither agree nor disagree, 4 = disagree, 5 = strongly disagree.

Next, we will follow the steps below to import the file:

 Click **File**, then **Open**, then **Data.**



A new window will open.

* Drop down the **Look in:**  box to the file N:/SPSS workshop 1 folder.
* Drop down the **Files of Type:** box to “Excel (\*.xls, \*.xlsx, \*.xlsm)
* You should now see the file “Workshop file.xlsx”
* Click directly on the Excel file “Workshop file.xlsx” so that the line is highlighted and the words

“Workshop file.xlsx” is in the “File Name” line.

The **Open Data** window should now look something like this:



Click the “Open” rectangle in the lower right-hand corner to open the file.

You will now see the following box:



Observe:

* We want the checkmark beside “Read variable names from the first row of data”, because our

variable names are on the first line of the Excel file.

* If we had other Excel worksheets in the file, we would use the drop-down box next to “Worksheet” to choose the worksheet that we want. In this file, however, there is only one

 worksheet with data so we will leave that as is.

* If we want to restrict the range of columns or rows, we would use the “Range” option to set

that; personally I have not had occasion to use this.

We have this box as we want it. Click “OK”.

After maybe a little fiddling with the data set to make the column titles more clear, the output file (top

file) and the data file (bottom file) should look like this:



**Designating labels, variable types, etc.**

Once you have imported the Excel file, you can follow the procedure from Pages 3-6 of this handout to finish setting up the variable view components of the file. Click on the variable view tab; you will find that some of the Variable View set up has already been done. The window (after enlarging it horizontally) will look something like this:



You can now go through and adjust each column from left to right to add labels, values, and change any measures from Nominal to something else.

The file “Workshop Results End File” that you opened at the start of this workshop is this Excel file with all the adjustments. Go open that file now; we will use it to review some procedures in SPSS.

|  |
| --- |
| **SAVE YOUR FILES! SAVE YOUR FILES! SAVE YOUR FILES!** |

**CATEGORICAL DATA**

We will do the following:

* Frequency Tables for one variable
* Bar Chart for one variable
* Pie Graph for one variable
* Crosstabs table for two variables

**Frequency Tables for one variable**

* Analyze, Descriptive Statistics, Frequencies



You get this window:



To move variables into the “Variables” box,

(1) click on the variable, and then

(2) move on the left arrow key to move the variable over.

Do this for each of the categorical variables. Your box should look like:



Now click “OK”. Tables will appear in the output file.

To copy the tables to a Word document:

* Right click the table
* Choose Copy Special. A new window will open
* In the new “Copy Special” window: Uncheck everything, check “Image”, check “Save as Default” so you only have to hit this window once in a session, then choose “OK”.
* Right click in the Word document
* Choose Paste.

What you get is a JPG. of the table.





**Bar chart for one CATEGORICAL variable**

NOTE WELL! Bar charts are for CATEGORICAL variables, while histograms are for QUANTITATIVE variables. SPSS treats them accordingly, and hence so should you!

Do: Graphs, then Legacy Dialogs, then Bar



A new window will appear. In the new window choose “Define”; do not change anything.

Another new window will appear.

* Click on the desired variable, then click the area next to “Category Axis” to move the variable to there.
* Look under “Bars Represent” and choose what you want the graph to show.
* Click OK.

Below we see the graph for “Class Rank”. COMMENT: We could have created a title in SPSS, but I recommend that you copy and paste the graph into Word and use Word to create the title.



We could (1) change the colors on the graph and/or (2) add the counts or percentages directly onto the graphs. However, because of time we will save that for another workshop. You can in the output window double-click on the graph to open the Chart Editor and then go from there.

**Pie chart for one CATEGORICAL variable**

You can do a pie chart in similar fashion:

* Do: Graphs, then Legacy Dialogs, then Pie
* For the first window choose “Define”
* Move the desired variable to “Define Slices By”
* Choose what you want under “Slices Represent”

Click OK. The graph is like this:



Again, you can change the colors in Chart Editor and (more usefully!) have the percentages or counts put into the graph. That will await another workshop, though if we have time today we will come back to it.

**Crosstabs Table for two variables**

(1) Choose *Analyze*, *Descriptive Statistics*, then *crosstabs*

(2) Move the (independent) variable into the Row box

(3) Move the (dependent) variable into the Column box

(4) Click OK

Below are the results. Note that missing values are not counted in the crosstabs table.





**QUANTITATIVE DATA**

There are at least three basic ways to analyze “one quantitative variable”. We will start with the one that gives the most information in “one fell swoop” get a “kitchen sink full” of statistics so that you can take the information and make your own tables and charts.

Go to **Analyze**, then **Descriptives**, then **Explore**



You will get THIS:



Click on a quantitative variable (read: One with a ruler in front of the variable) and then click on the arrow to the right of “Dependent List” to move the variable to that box. For our purposes move BOTH “Speed” and “Seed” into the Dependent List box.

Next, note the three boxes in the upper right corner: Statistics, Plots, and Options. Click on the statistics box to see the following:



For the sake of argument we will checkmark (click in the box next to them) on all of these just to see what they give you. In reality you probably won’t use them all. Then click “Continue”.

What you will get (from [http://www-01.ibm.com/ support/knowledgecenter/SSLVMB\_21.0.0/com.ibm.spss.statistics.help/idh\_exam\_sta.htm](http://www-01.ibm.com/%20support/knowledgecenter/SSLVMB_21.0.0/com.ibm.spss.statistics.help/idh_exam_sta.htm)):

* Descriptives: These measures of central tendency and dispersion are displayed by default. Measures of central tendency indicate the location of the distribution; they include the mean, median, and 5% trimmed mean. Measures of dispersion show the dissimilarity of the values; these include standard error, variance, standard deviation, minimum, maximum, range, and interquartile range. The descriptive statistics also include measures of the shape of the distribution; skewness and kurtosis are displayed with their standard errors. The 95% level

confidence interval for the mean is also displayed; you can specify a different confidence level.

* M-estimators: Robust alternatives to the sample mean and median for estimating the location. The estimators calculated differ in the weights they apply to cases. Huber's M-estimator, Andrews' wave estimator, Hampel's redescending M-estimator, and Tukey's biweight estimator are displayed.
* Outliers: Displays the five largest and five smallest values with case labels.
* Percentiles: Displays the values for the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles. (NOTE: SPSS uses two methods—Tukey’s Hinges and Weighted Averages. Tukey’s Hinges are used to set outliers for boxplots, while weighted averages are used to calculate quartiles for the Descriptives table and for other SPSS tables of quartiles.

Next, click on the box called “Plots” to get this window:



Some comments:

* Boxplots: For one quantitative variable (e.g., not a “by” gender or some other variable), “Factor Levels together” and “Dependents together” will both give the same boxplot. If you do not want a boxplot, choose “None”. Backgrounds for the Boxplots can be changed with the Chart Editor.
* Descriptives: Gives a stem-and-leaf plot automatically. You can click on the “Histogram” to get a histogram set to whatever class boundaries SPSS chooses to use. To set your own class boundaries, click on the Histogram in the output file to open the Chart Editor, etc.
* Normality Plot with Tests: If you want to test whether it is reasonable to assume that the data come from a normally distributed population, checkmark this. You will get (1) Shapiro-Wilk and Kolmogorov-Smirnov tests, (2) QQ Plots, and (3) detrended QQ plots. Backgrounds can be changed with the Chart Editor.

We will checkmark (click box next to) “Histogram” and “Normality Plot with Tests” so that you can see what you get. Now click “Continue”.

Next, click “Options” to get:



This only become of concern if you have more than one variable in the “Dependent List” box. Hence, to demonstrate, we put both “Speed” and “Seed” in the Dependent List box.

Go down to observations 18 and 26 in the data set. You will see that “Speed” has a value but “Seed” does not:



* “Excludes cases listwise” drops all observations where at least one of the values of the variable is missing. Hence, the 115 (observation 18) and the 79 (observation 26) are dropped for Speed because the value for Seed is missing, and SPSS calculates descriptive statistics for both Seed and Speed with the remaining 24 values.
* “Excludes cases pairwise” essentially uses for each variable all values available for that variable; a value of a variable is not dropped if another variable is missing. Hence, the 115 (observation 18) and the 79 (observation 26) are retained for Speed even though the value for Seed is missing, and so SPSS calculates descriptive statistics for Speed with all 26 available values and Seed with its 24 values.

Choose which one you want, and click “Continue”.

FINALLY, click OK to get all of this in the Output file.

* To copy and paste tables into a Word file, do “right click, copy special, right click, paste”
* To copy and paste graphs into a Word file, you can “right click, copy, right click, paste”
* To change backgrounds on the graph, double-click on the graph in the Output file to open the Chart Editor.