Lab 3 Introduction to SPSS and R, Descriptive Statistics

Overview of SPSS

Opening SPSS. Find SPSS in the programs list on the computer under **IBM SPSS Statistics**
There are several main menus, including **File**, **Edit**, **View**, **Data**, **Transform**, **Analyze**, etc., but most of them we will not use.

Three main windows. Note that different windows have different toolbars.

1. **Data Editor** (file extension is .sav): as a default window, it defines, enters, edits and displays data. This window has 2 important tabs with different views:
   Data View: Cells in the Data View will contain data values, where columns are variables and rows are people (cases or observations).
   Variable View: Variables are defined in the Variable View with several attributes, such as *Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align* and *Measure*
2. **Output Viewer** (file extension is .spv): it displays outputs (results, warnings, errors)
3. **Syntax Editor** (file extension is .sps): it is a text editor used to send commands. Alternative to using the menus for transformations and analyses and recommended for keeping good records of your data changes and analyses.

Working with Data

***Download and open or enter the data*.** We will be working with data exported from Qualtrics, so download the data set for this lab from https://web.pdx.edu/~newsomj/data.htm. (In other circumstances you may need add variable names in the variable view and enter the data in the data view first)**.** Open your data file *File -> Open -> Data..* and navigate to the data location. See the data in the *Data View* tab.

Adding variable labels. Variable labels provide some additional information for brief variable names. In the *Variable View* tab, add variable labels (this helps you keep track of what the shorter variable names mean) into the *Labels* column. For example, for a variable called `Q1`, from a question like “I am often afraid of using computers,” type in "afraid of computers". It is often helpful to use a brief descriptor if the full question is too long.

Adding variable values. Value labels allow you to see what the response option wording was for each numeric value (e.g., 1=strongly disagree). In the *Variable View* tab, click on the *Values* column, then click on the three dots that pop up. In pop-up box, click on the plus sign (**+**) at the right, enter the first value (e.g. 1) in value box and then a label for it to its right (e.g., “strongly disagree”). Click the plus sign again, enter the second value (e.g., 2) and a label to its right (e.g., “disagree”). Then continue until you have completed all of the values. If other variables have the same labels, click on the label box for that variable in the *Variable View* tab, copy and paste for another variable.

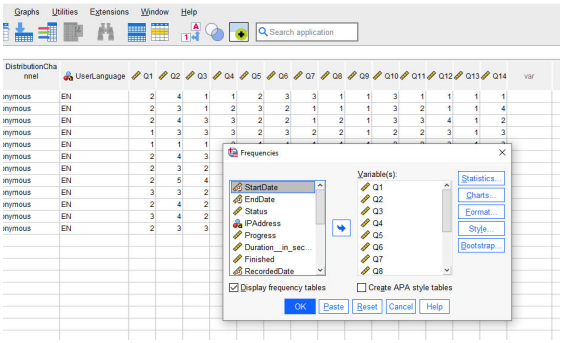
Variable types. The *Type* column in *Variable View* allows you to change the type of variable. For numeric variables, leave them as *Numeric*, which is the default. But for some variables that use words instead of numbers (e.g., job occupations such as “professor” “student”), change the type to *String*.

Saving your data. I strongly suggest that after you have changed your data, save it as a different file name. *File -> Save As*. That way you will always keep an original, unaltered copy of the data that you can always go back to if you make a mistake. The default data type is .sav, which is the SPSS data file format we will be using throughout the course. Notice that you can save in other formats, such as Excel or ASCII (text format).

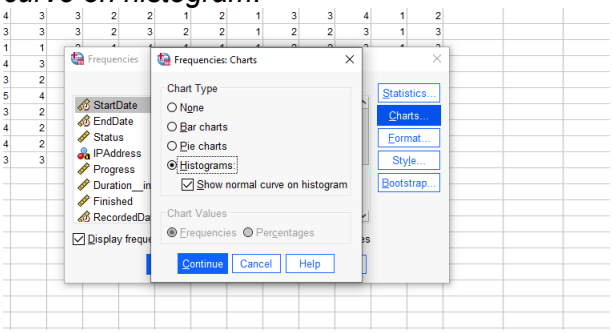
Descriptive Statistics in SPSS

The first step in an analysis of a measure is to get some descriptive statistics of the items. There are two useful types of information to get, (1) *frequency table* and *frequency histograms* for all of the items, and (2) *descriptive statistical values* (e.g., mean, standard deviation)

Frequencies. With your data set open, go to *Analyze* menu choose *Descriptive Statistics -> Frequencies*. On the left hand side of the frequencies box, scroll down and find the scale item variables. Highlight them all and hit the arrow in the center to drag them over.

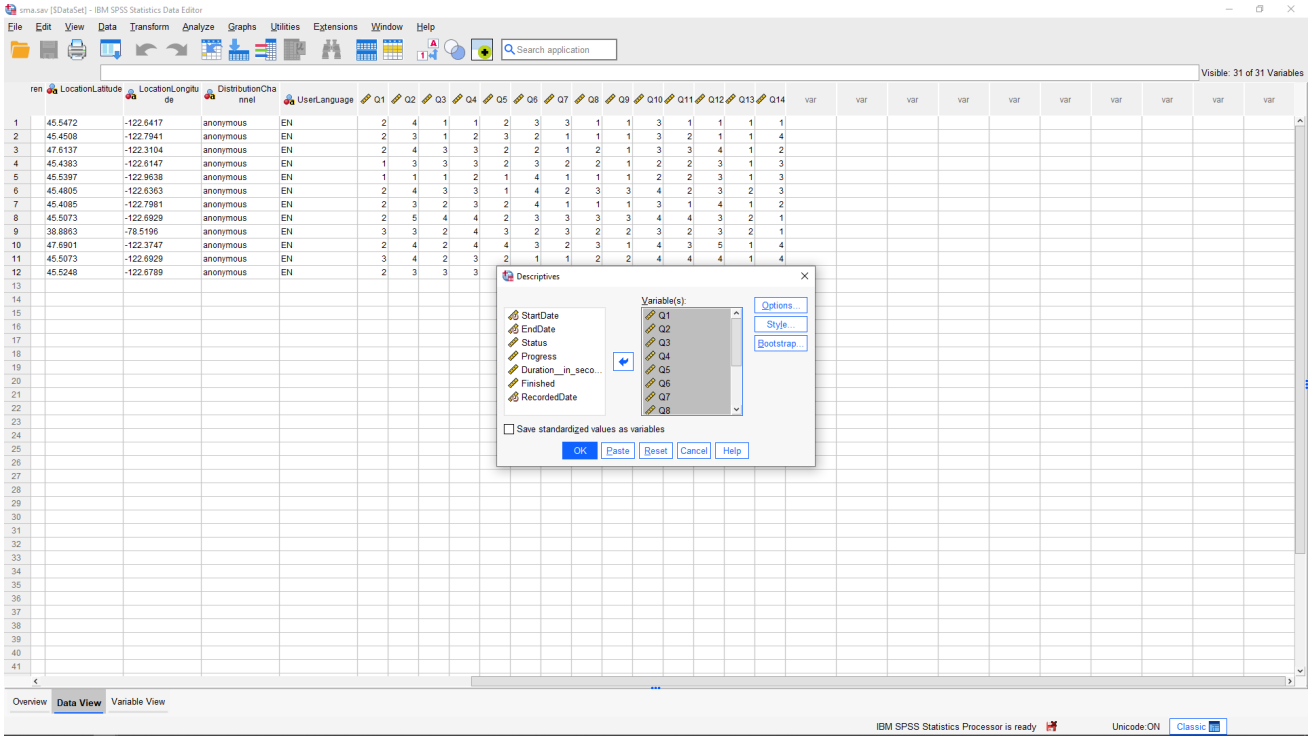


Then click on *Charts* and chose the *Histograms* radio button and check the box for *Show normal curve on histogram*.



Hit *Continue* and then *Ok*

Descriptives. *With your data set still open,* go to *Analyze* menu choose *Descriptive Statistics -> Descriptives.* On the left-hand side, scroll down to find the item variables, highlight them all. Hit the arrow in the middle to move them over to the right-hand side. Choose *Ok*.


sma.sav [DataSet] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Search application
Visible: 31 of 31 Variables
Descriptives
Variable(s):
StartDate
EndDate
Status
Progress
Duration__in_seco...
Finished
RecordedDate
Q1
Q2
Q3
Q4
Q5
Q6
Q7
Q8
Options...
Style...
Bootstrap...
Save standardized values as variables
OK
Paste
Reset
Cancel
Help
Overview
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Classic

Overview of R

What is R? R is an open source language and software environment for statistical computing and graphics. It provides a wide variety of statistical techniques and is very flexible. The R program is free and can be downloaded from https://www.r-project.org/. The R package is a community contributed package and has many modules that are contributed by different individuals. So, most transformations or statistical analyses require that you load different components called "packages" for different types of data manipulations or statistical analyses.

What is Rstudio? Rstudio (https://www.rstudio.com/) is an interface that makes navigating R script or code (same as "syntax" in SPSS) and output a bit easier.

Windows in Rstudio. There are four windows shown in Rstudio by default: text editing window, which can be used for writing a list of commands, the “console” which is where you find output, a History window (which has several tabs in addition to history) that shows prior commands, and a Plot window (which also has several other tabs) that displays output figures.

Opening Data Files

In R, you can work with a variety of data file types. For now, we will always be opening SPSS data files. There are many ways to do this, but I'll illustrate using the `lessR` package. The first step is to install the `lessR` package. This only needs to be done once for each computer. Copy this command into the text editor to start a new script.

```
install.packages("lessR")
```

Note that R is case sensitive (SPSS commands are not). Highlight all of this text and hit the *Run* button at the upper right of this window. You will see a series of statements printed in the output console. Now we can use the package. We begin adding commands by first giving a `library` statement to tell R we want to use the package we just installed. So copy this text (and you can replace or delete the `install.packages` command). Remember everything is case sensitive. Be sure to use forward slashes.

```
library(lessR)
#be sure to change the path for your computer - use forward slashes
d = Read("C:/Users/newsomj/Downloads/sma.sav",quiet=TRUE)
View(d)  #shows the data file
```

Descriptive Statistics in R

There are many packages in R that will provide descriptive statistics, but we will use the `lessR` package, because it is quite easy to use. My approach to using R in Rstudio is always to use a script file and run the full set of code in that file.

Frequencies. To generate a frequencies histogram of all of the variables, the easiest thing is to refer to all of the “data frame” called “d” using the less R function called `Histogram`. Keep the code above starting with lessR and add below `View(d)` this command.

```
Histogram(d)
```

Descriptive statistics. To generate a descriptive statistics, we will use the lessR command `pivot`. We cannot simply refer to the whole data set with this command so must list the variables.

```
pivot(d, c(mean,sd), c(Q1,Q2,Q3,Q4,Q5,Q6,Q7,Q8,Q9,Q10,Q11,Q12,Q13,Q14))
```