

Assignment 6: SPSS

Due Wednesday, October 22, 2014

There is a data file named: SPSS data 10 22 14.xlsx

The data were collected from PSYC 1120. The extension .xlsx indicates that this is an Excel file. The file, which will be used for this exercise, is posted on the Canvas site. There are work stations in O'Neill Library. Also, try accessing SPSS from your own computer using apps.bc.edu and Citrix.

Excel is a great database program for creating data files. It is compatible across PC and MAC platforms, and almost all statistical software packages will accept Excel files. Some will even output data into Excel formatted files. If you are ever collaborating with anyone, or if you want to maintain usable data files for any length of time, I suggest using Excel. Excel is better than storing data in an SPSS file.

Your task for this assignment is to carry out the steps outlined below. The purpose of the assignment is to familiarize you with SPSS.

First open the data file in Excel. Look at it. Each column represents a variable. The convention is that each row corresponds to one subject or case.

****IMPORTANT**** For this file I have jumbled the order of cases so that the SAT math score does not correspond to the SummerIncome. Almost always, each row will contain data for one person, subject, or case. That will be true in later assignments.

Sometimes there will be a column containing an ID number, but not always. For data management, it is a good idea to have an ID number, as long as you maintain the anonymity of the participants.

Notice which columns contain numerical values and which contain strings (letters or letter and number combinations). In this file, there are only numerical values. Are there any missing values? Any questionable values?

Close Excel. Make sure the file is saved somewhere where you can access it. I suggest putting it on your hard drive.

Open SPSS.

Select "Open an existing data source".

SPSS of course starts by thinking you will be opening an SPSS file. SPSS data files have .sav as an extension.

Select Excel as an option for files to look for.

Find the file.

Open it.

You will see a dialogue box asking whether you want to read the first row of information as column names. **YOU SHOULD MAKE SURE THIS BOX IS CHECKED.**

Most often, Excel files will use the first row to label the variables, which are organized in columns. If you do not check this box when opening Excel files, SPSS will read the first row as data, and that will cause problems related to SPSS' misunderstanding what type of information is present in each column (numerical or string).

Sometimes an Excel file will use two rows for labels. You will have to change this before opening the file in SPSS. SPSS can only handle a single row as a label.

Once SPSS has created its own untitled datafile, you should again look it over.

There is Data View, and there is Variable View. Click on Variable View to see how SPSS has interpreted the Excel file.

Are all the variables considered "scale" in the Measure column over on the right? (Scale means more or less the same as "numerical". Even if your data really are only ordinal in nature, I would call them "scale" in SPSS to provide more flexibility in analysis.) If a variable is marked as nominal rather than scale, you may need to fix that. In this file, Hiccups is a nominal variable due to a typographical error in data entry.

If any continuous variables are showing as anything other than Scale in the Measure column, you should take steps. Look at that column under Data View to examine the actual values. Is there something that is not a number? If so, you need to fix that. Return to Variable View. Under Measure, click on the entry to find "scale". On the left, under Type, you will need to have "numeric".

You can try double clicking on the scale entry, for example, to see the options.

You might save the SPSS file with some name you like.

Older versions of SPSS will limit the names of variables to 8 characters. Newer versions of SPSS will allow longer names. I would be surprised if the version you are using will not allow longer variable names.

SPSS will assign its own names to variables if you don't do so. You can change the names later if you like.

For starters, go to Data, and choose Sort cases. Select SummerIncome and move it into the "Sort by" box, and use the default (ascending order), and hit "OK". Check the data file. The cases (rows) are now rank ordered according to height values. The software will do the sorting for the entire rows (not all software does it this way----you need to be careful using sort functions).

If there are clear data entry errors or missing entries, decide what to do. One option is to delete the case(s) from all analyses. Another option is to delete just the offending data point while leaving the other data from that same case. Another option is to replace a missing entry with some value. The literature contains extensive discussion of how to treat missing data. What is crucial is that you inform your reader about any data correction, replacement, deletion, etc. Failure to do so constitutes fraud.

For this assignment, leave the missing cells as empty. Delete the impossible values.

Now go to Transform and then Compute variable. You want to create a new variable that is a log (base10) transform of the SummerIncome variable.

Under Function Group over on the right, select Arithmetic. Then scroll down until you see Lg10, and select that option with a double click. Up above you will see Lg10 (?). Now select footwear from the list on the left and double click. The question mark should not be there anymore. You might have to delete it. Next, you need to generate a name for the transformed variable. I suggest llogSummerIncome. The reason I used two lls is because once I got an error message that the name I had selected was illegal. SPSS can be fussy.

Then hit OK. Look at Data View. There will be a new column with the new values.

Create standard score equivalents for the SummerIncome variable. One annoying thing about the Transform function in SPSS: You cannot calculate z scores using it. To get z scores, you need to go to Analyze, then Descriptive statistics, then Descriptives, and select the Save Standardized Scores option. You will need to tell the computer which variables you would like changed into standard score form. Choose both SummerIncome and SATmath. A new column will appear in the data file for each transformed variable.

Now, go to Analyze. Find Descriptive statistics and, under that, click on Frequencies. Under Frequencies, you will need to tell the computer which variables you would like analyzed. Choose all of them, even the two transformed variables you just created, to see what happens. Under Options, select mean, median, mode, SD, variance, maximum, minimum, skew and SEM (standard error of the mean). Also under Charts, select histogram and also the normal comparison.

Be sure to Unselect Display Frequency Tables. These can be useful, but most often they are just clutter.

Then hit OK.

You will see output. Look at it.

Saving output. There is a chance the graphs will not print. Some workstations do not have a graph printing capability set up. However you should take a look on the screen at the histograms.

Things to notice: the comparison between the SummerIncome and SATmath variables, the z score transform, and the log transforms in terms of skew (graphically, with the normal curve drawn in and numerically). For the raw scores, ask whether the SD is large relative to the mean, and whether the mean and median are different, and whether the value for skew is pretty far from zero. Also, look at the histogram. Do the same for the standard scores. The histogram will look a little different due to the default values for bin size used by SPSS, but the value for skew should be identical. Now, do the same for the log transformed scores. The histogram for LlogSummerIncome is nicer looking, and the value for skew is closer to 0.00, which is good.

Saving the results of your analysis. This can be an annoying process. SPSS will often generate vast quantities of output that you don't need or want. Even worse is that some workstations will not print graphs from SPSS. My suggestion is to use the following strategy. For each portion of the output you want (that is, each individual table and graph), select that portion by clicking it, copy it, and paste it into

an open Word file. Repeat for the next table, etc. You can annotate the Word file as you go along too. It is important to document fully what each portion of output is, what each variable label refers to, the data file used, the date of the analysis, etc.

Save each table and JUST the histograms for SummerIncome and llogSummerINcome to a Word document and save that Word file. Upload that word file as the assignment.

If you run into trouble, ask a TA or the instructor. The difficulties you may encounter will most likely be easy for someone else to address, though not something you could figure out using pure reason. SPSS is not intuitive.

Congratulations. Feel free to play around with the data and to try different menu options under Analyze.

Always: use software to examine data for outliers, weird values that may reflect errors, for the shape of the distribution.