

SPSS Basics

Due at the Start of Lab: None

Rationale for Today's Lab Session

This tutorial is designed to ensure that you can access and use a computer program, SPSS, to begin analyzing data. If you are already familiar with SPSS from your prior coursework, this should be review. If you are unfamiliar with SPSS, this assignment will help you catch up. You will go through this tutorial with your lab instructor. You can work collaboratively on this tutorial but must work independently on the graded lab assignment.

Instructions

Warning

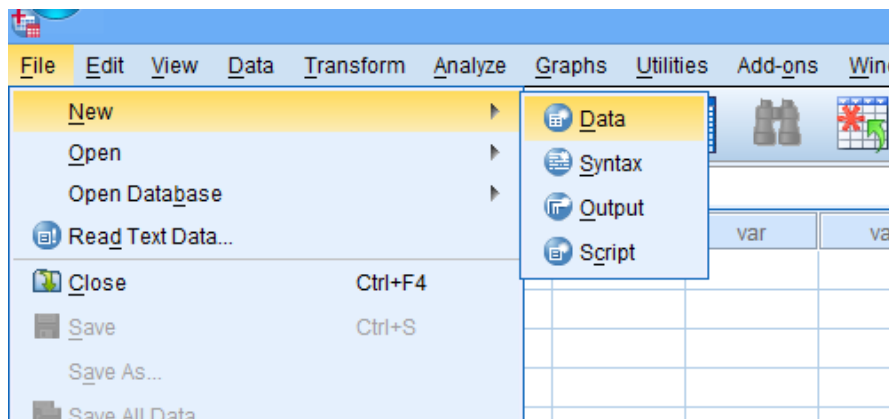
SPSS periodically changes the visual display and organization of menus. The instructions presented in this tutorial may need to be augmented marginally depending on the version of SPSS you are using. If you get stuck, use Google, or ask the lab instructor for help.

Accessing SPSS

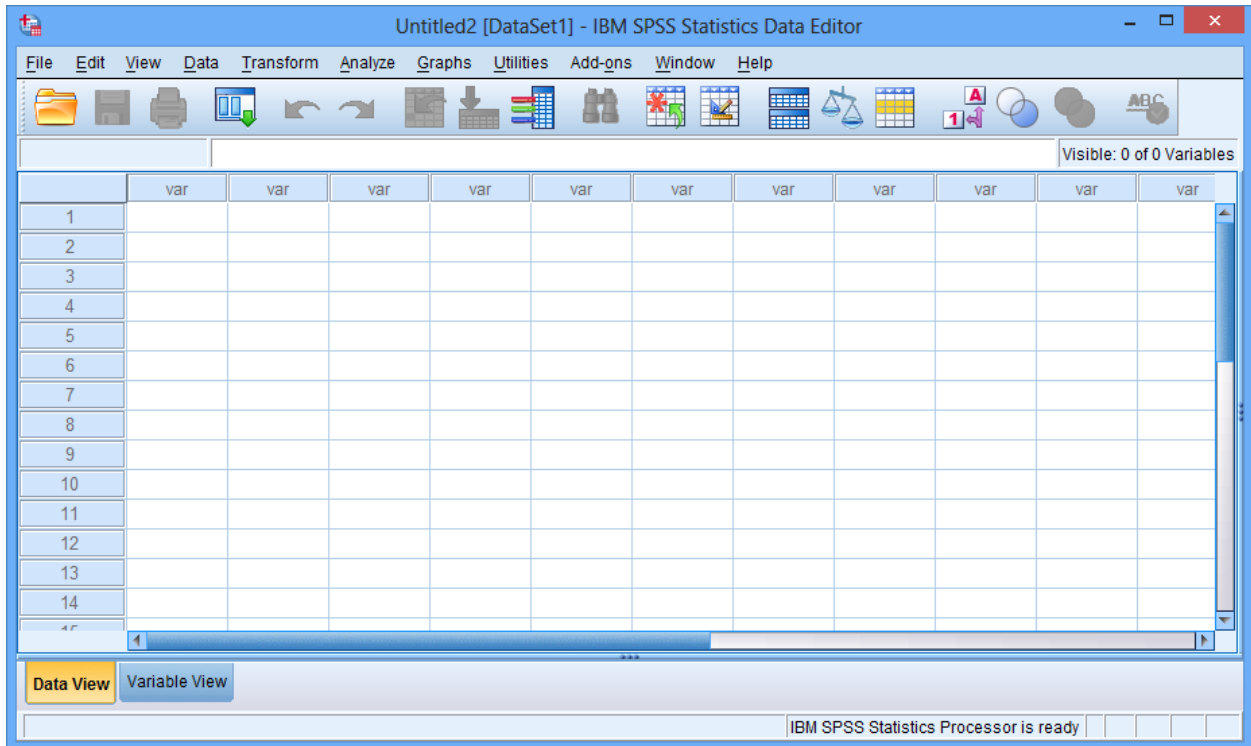
Once you log on, go to the Start menu in the lower left corner of the screen and find SPSS. If you have difficulty finding it, ask the lab instructor for help.

Creating a New Document

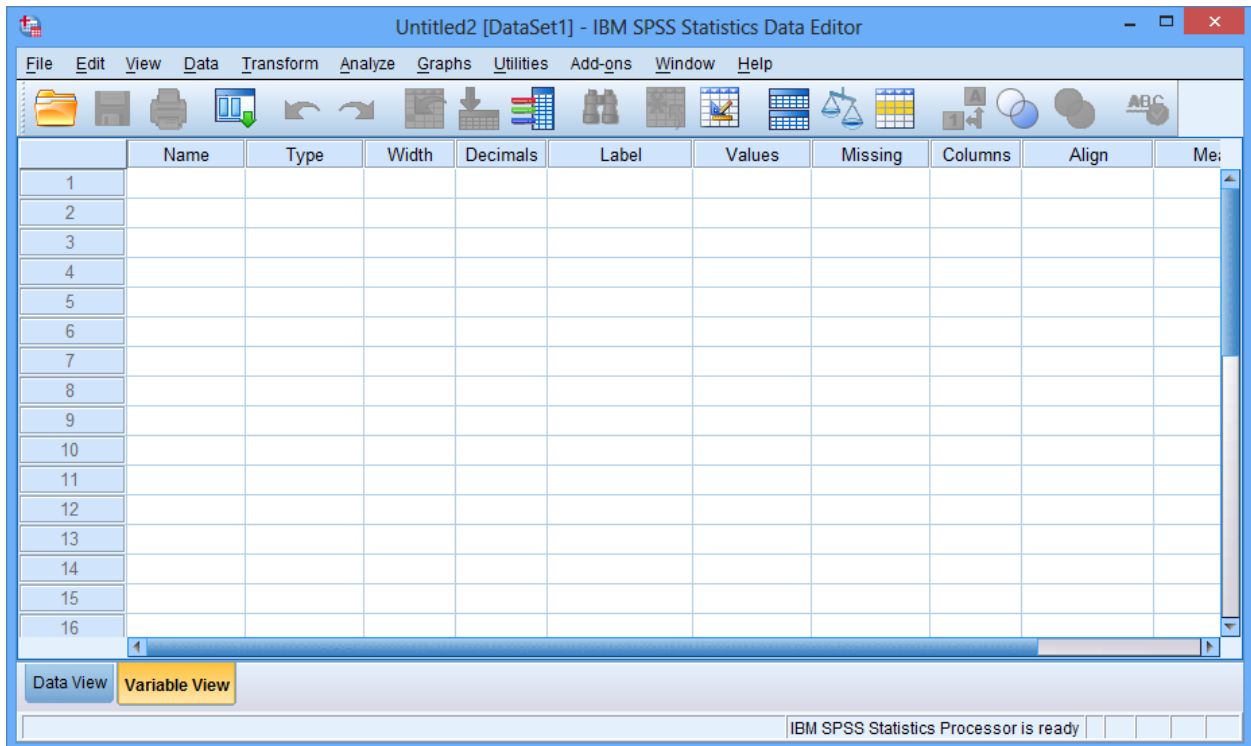
Our goal is to begin with a new SPSS document. If a pop-up window appears when SPSS loads, just click Cancel. Usually a blank document will be open automatically. If not, go to the File menu, point to New, and select Data.



A blank document should look something like the following picture.



Or this....



Explore SPSS

SPSS has two views, called Data View and Variable View. You can switch from one to the other by clicking the Data View or Variable View buttons in the bottom left corner of the SPSS window (the gray or orange buttons on the previous page). They both look pretty similar since the document is blank.

Data View is where you enter data. For example, you might type in each participant's scores on a variety of measures. Variable View allows you to make more advanced changes, such as labeling variables.

Take some time to browse through each of the menus at the top of the SPSS window to familiarize yourself with the program. Here is an overview of each menu.

File. Contains commands for opening and closing files

Edit. Contains Copy/Paste commands and features for searching text

View. Controls the on-screen layout (rarely needed)

Data. Contains commands for sorting the data or grouping the data in various ways

Transform. Used to perform basic calculations or re-code variables

Analyze. Contains most of the functions for running statistical tests (often needed)

Graphs. Contains functions for examining the data visually

Entering Data

Make sure you are in Data View. The Data View button in the lower left should be orange (or highlighted in some way).

When you are in Data View, the numbers along the side represent individual participants. Each column (labeled "var" at the top) represents a different variable.

Enter the data set that appears on the following page. The column names will change from "var" to something new (like VAR00001, VAR00002, etc) when you start entering data. [Note: If you try to enter data and the numbers do not appear, you may be using an unlicensed version of SPSS. Consult with the lab instructor or tech support. You may need to use a different computer or lab].

*Untitled3 [DataSet2] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

20 : Visible: 5 of 5 Variables

	VAR00001	VAR00002	VAR00003	VAR00004	VAR00005	var	var
1	29.00	36.00	35.00	34.00	35.00		
2	22.00	39.00	36.00	34.00	36.00		
3	25.00	36.00	32.00	37.00	35.00		
4	24.00	39.00	39.00	35.00	40.00		
5	29.00	35.00	36.00	32.00	29.00		
6	21.00	36.00	30.00	32.00	39.00		
7	27.00	40.00	34.00	31.00	37.00		
8	29.00	31.00	31.00	30.00	31.00		
9	25.00	36.00	36.00	32.00	24.00		
10	21.00	38.00	46.00	37.00	38.00		
11	28.00	32.00	30.00	35.00	32.00		
12	33.00	37.00	35.00	33.00	29.00		
13	19.00	40.00	30.00	35.00	39.00		
14	19.00	37.00	33.00	39.00	38.00		
15	25.00	30.00	29.00	33.00	38.00		
16	23.00	38.00	32.00	36.00	37.00		
17	24.00	33.00	31.00	39.00	36.00		
18	21.00	34.00	37.00	37.00	38.00		
19	24.00	35.00	34.00	26.00	35.00		
20	20.00	40.00	32.00	33.00	34.00		
21							
22							

Data View Variable View

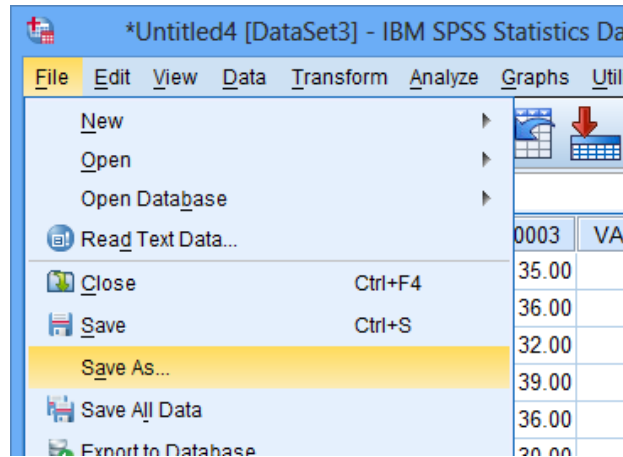
IBM SPSS Statistics Processor is ready

Each row going across represents a different participant. Each column represents the person's scores on a different variable. For example, participant #7 scored a 37 on whatever the 5th variable is (circled above).

Saving Data

It is important to save everything periodically, as computers freeze often. Always save your work. You might catch a mistake once you're home and want to go back to the lab to fix it. If you haven't saved your work, you may have to start over. I suggest using a portable flash drive. To save, go to the File menu, and point to Save As. This will allow you to choose where to save

the file and what to call it. When re-saving a document, you can simply click Save or press Ctrl + S (the Control key and “S” on the keyboard, at the same time).

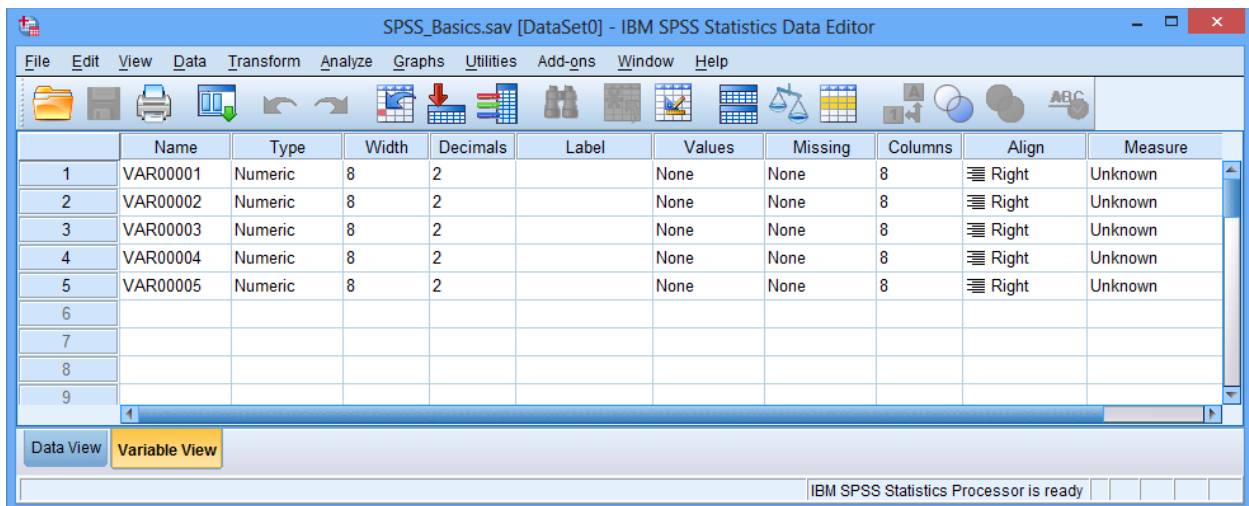


Find the appropriate folder on your flash drive, and save your work as something logical, like “SPSS_Basics.sav”. **Never save any work to the local hard-drive, as this could aid a student in another lab section in plagiarizing your work, placing you both at risk.**

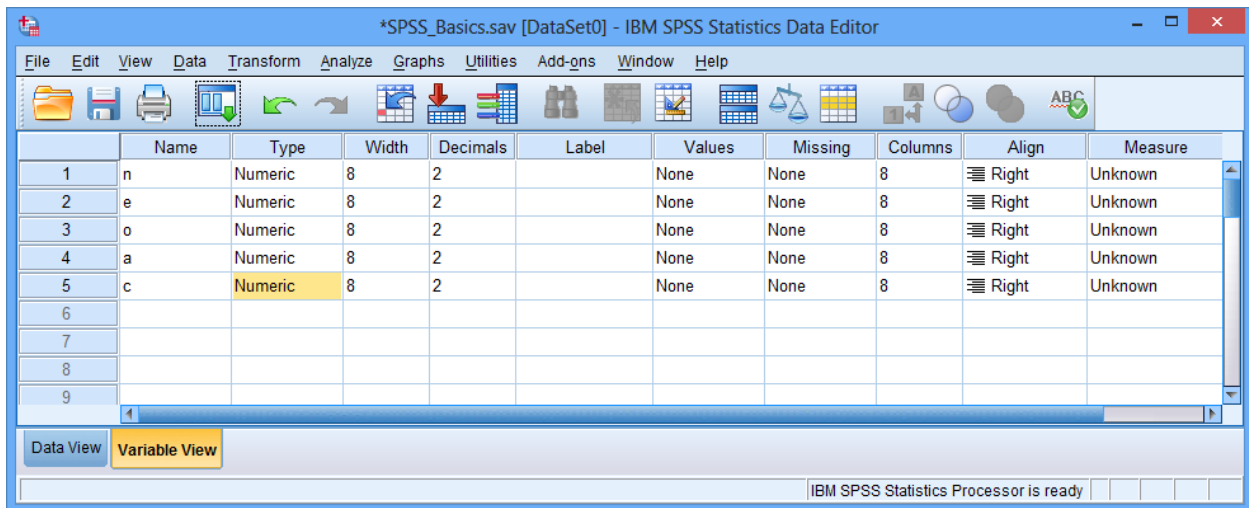
Naming Variables

The data you have entered are from real participants but so far appear meaningless because the variables have no labels.

Switch to Variable View, by clicking the button in the lower left corner. In Variable View, the window should look something like the following.



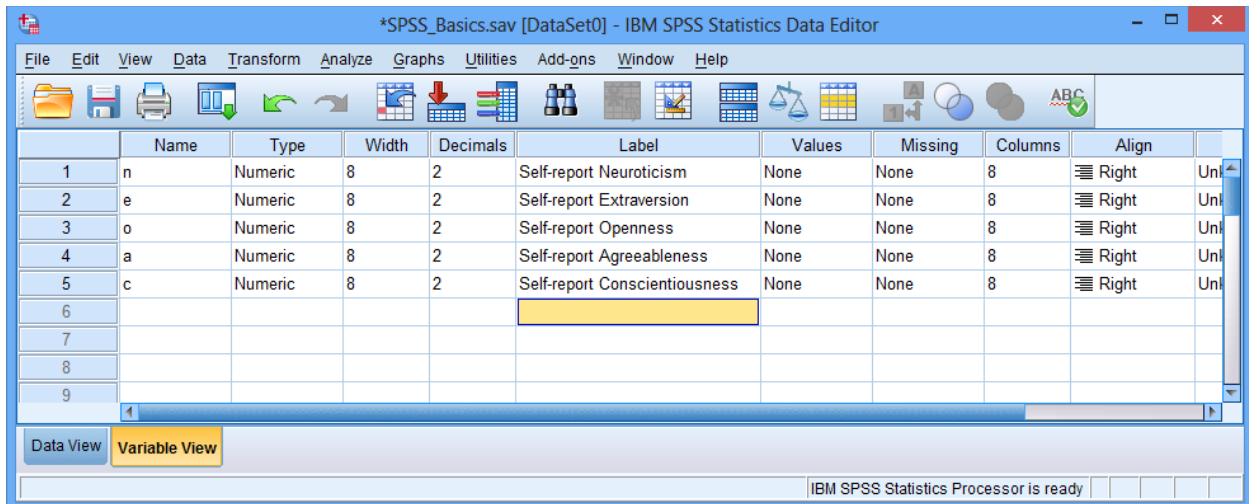
In the Name column, on the left side (where you see VAR00001, VAR00002, etc), you will enter the names for the variables. In this case, we are working with personality variables (neuroticism, extraversion, openness, agreeableness, and conscientiousness. Change the names in the Name column, so they look like this (n, e, o, a, c):



Now switch back to Data View, by clicking the button in the lower left. You should see n, e, o, a, and c as column headings in the data file now. Usually, people use short abbreviations for these column names, and this allows them to quickly and easily look over a data file.

Switch back to Variable View.

Now we are going to enter more detailed descriptions in the Label column. The Label column is where you can provide more detailed labels of variables. The Name information is used for column headings in the data file itself, whereas the Label information is used in our statistical output and printouts. These longer descriptions in the Label column will make it easier for others to understand the variables. Enter the following labels (e.g. “Self-report Neuroticism”) in the Label column:

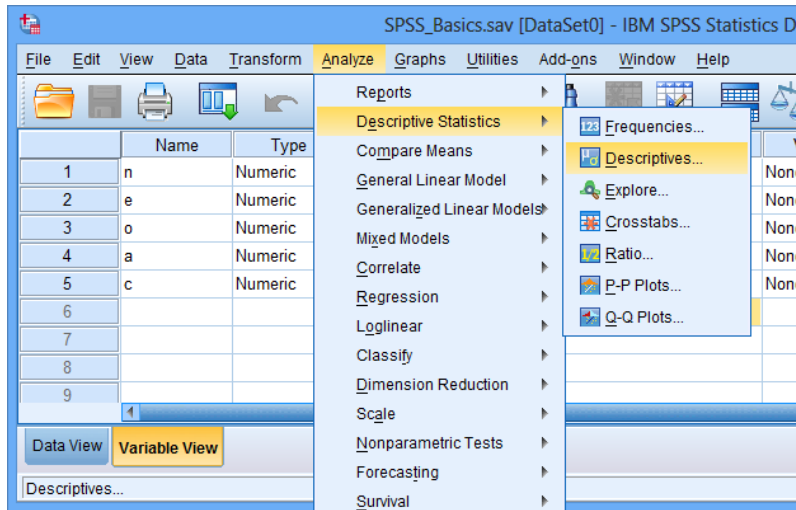


It’s easy to misspell “conscientiousness” (a synonym for work ethic, organization, and tidiness) so watch your spelling.

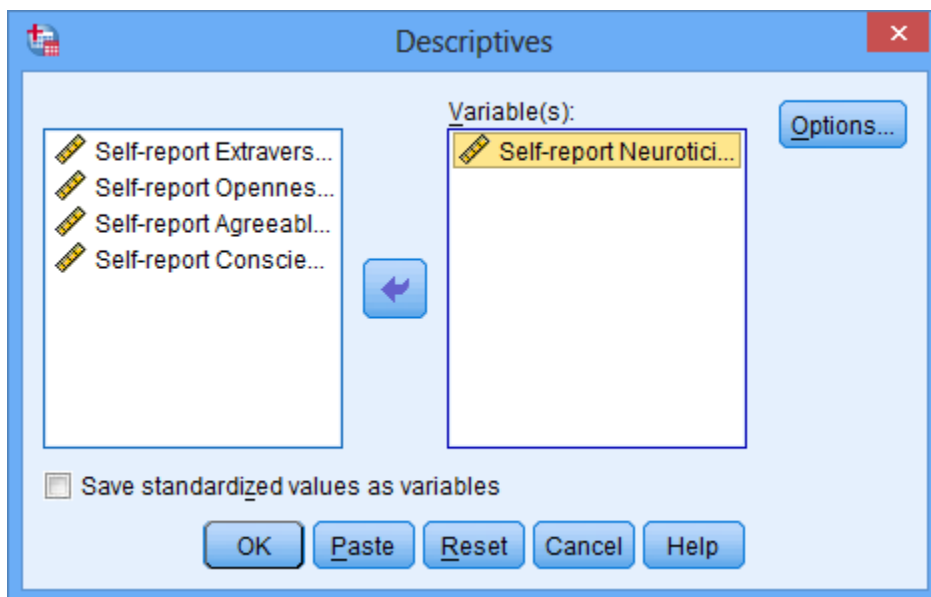
Re-save your data using the Save command in the File menu.

Basic Statistical Operations

There are many ways to get simple (descriptive) statistics in SPSS. Go to the Analyze menu, point to Descriptive Statistics, and choose Descriptives.



A window should pop up. In the left panel, click on the first variable (Self-report Neuroticism). Then click on the arrow button in the middle, moving Neuroticism to the Variable(s) pane on the right. Then, click OK. This will give you some basic statistics on the Neuroticism variable. [Note: SPSS allows you to move multiple variables to the right, if you want to analyze several variables at once. You can also make some choices about which statistics to include, through the Options button, if needed].



When you click OK, the following statistical Output should appear:

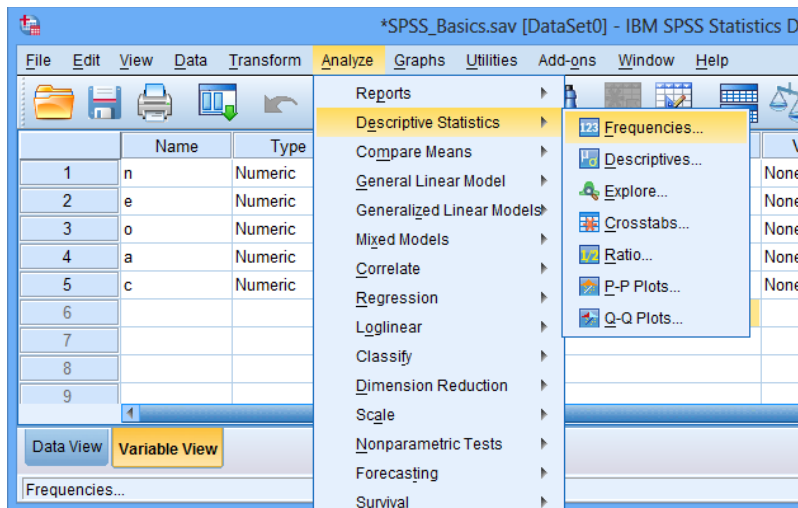
Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Self-report Neuroticism	20	19.00	33.00	24.4000	3.83062
Valid N (listwise)	20				

This tells you that the sample size (N) is 20. Scores ranged from 19 to 33. The mean was 24.4. The standard deviation (a measure of variability) was 3.83.

Frequency Tables

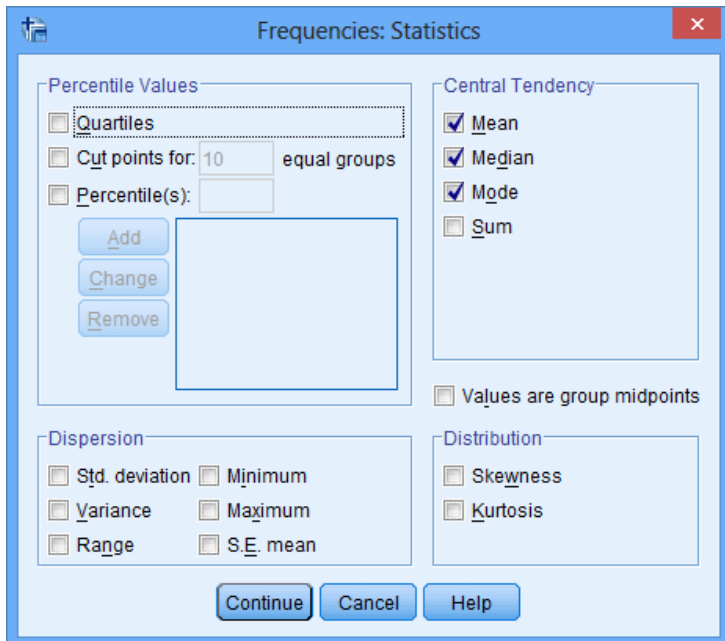
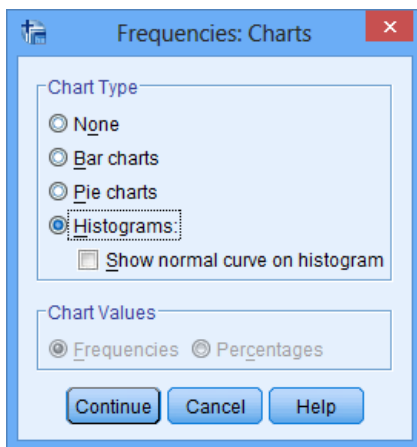
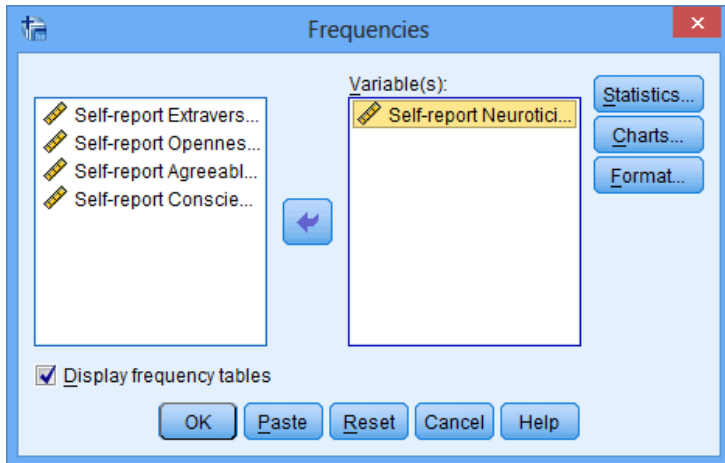
SPSS can also make rough frequency tables. Go to the Analyze menu, point to Descriptive Statistics, and choose Frequencies.



In the window that pops up, move the Neuroticism variable to the Variable(s) pane on the right. This function also has other features.

In the same pop-up window, click on the Charts button. In a new window that pops up, choose Histograms. Then click Continue, taking you back to the first pop-up window. This will include a Histogram with the Output.

Finally, click on the Statistics button on the right. In the window that pops up, select Mean, Median, and Mode. Then click Continue, taking you back to the first pop-up window. In the original pop-up window, click the OK button, producing your Output.



The Output should look something like this. Although typically we make frequency tables that are descending (highest number on top, lowest on the bottom), by default SPSS uses an ascending order (the reverse). Thus, it's always a good idea to look over the Output carefully before answering any questions.

Statistics

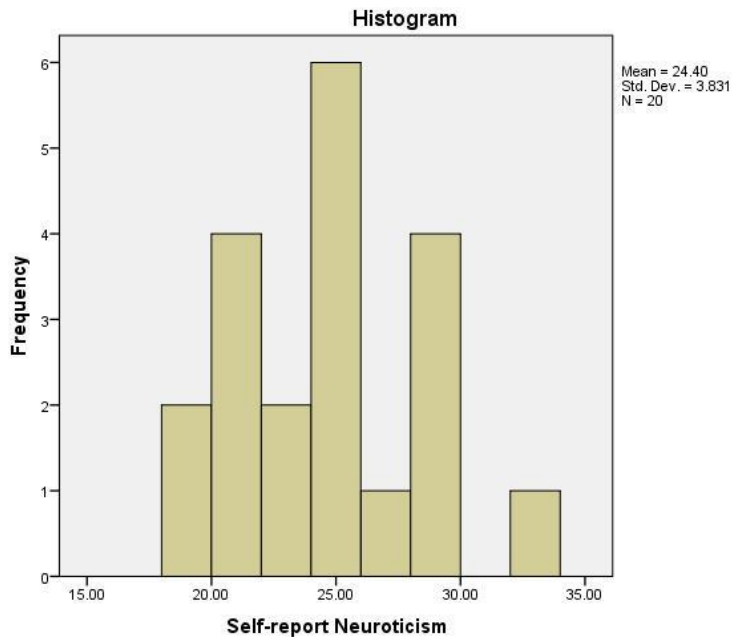
Self-report Neuroticism

N	Valid	20
	Missing	0
Mean		24.4000
Median		24.0000
Mode		21.00 ^a

a. Multiple modes exist.
The smallest value is shown

Self-report Neuroticism

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 19.00	2	10.0	10.0	10.0
20.00	1	5.0	5.0	15.0
21.00	3	15.0	15.0	30.0
22.00	1	5.0	5.0	35.0
23.00	1	5.0	5.0	40.0
24.00	3	15.0	15.0	55.0
25.00	3	15.0	15.0	70.0
27.00	1	5.0	5.0	75.0
28.00	1	5.0	5.0	80.0
29.00	3	15.0	15.0	95.0
33.00	1	5.0	5.0	100.0
Total	20	100.0	100.0	



In this Output, we again see the Mean, Median, and Mode (there are many ways to find these in SPSS). In the left column, we see the possible Neuroticism scores. The next column has the

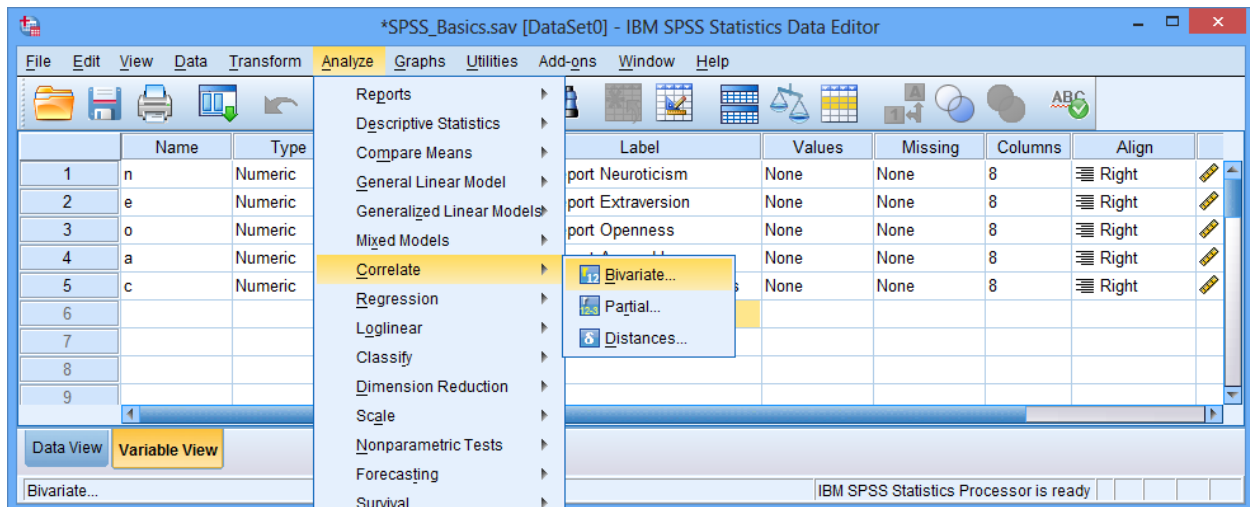
frequency (number) of people who have each score. The Percent column places the frequency into a percentage of the total. The Valid Percent column has identical numbers (they only differ in data files where some participants have missing data – it calculates the percent based on how many people actually completed the survey, rather than how many people are merely in the data file). The Cumulative Percent shows the cumulative percentage. For example, 15% of people scored a 20 or lower. 75% scored a 27 or lower.

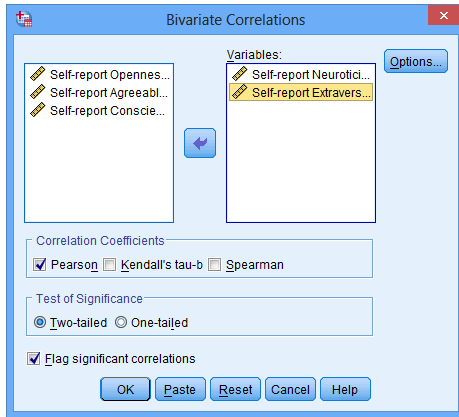
Correlational Analyses

We will cover correlations more in the next unit. This is a brief introduction. Correlations range from -1.00 to +1.00. The sign (+/-) tells how the variables are related. A positive correlation means that as scores on one variable increase so do scores on the other variable. For negative correlations, as scores on one variable increase, scores on the other variable tend to decrease. The numeric value tells you how strong the relationship is. If the number is .10 or higher, the correlation is small; .30 or higher is modest, medium, or moderate; .50 or higher is strong.

Each correlation also includes a *p*-value, which is a probability statistic. The *p*-value more-or-less tells you the probability that a particular correlation would be obtained by chance. A low *p*-value means that the result is probably not due to chance; there is likely a real relationship. A high *p*-value means that an observed correlation is likely a chance finding and is not indicative of a real-world relationship. As an arbitrary cut off, we conclude that a correlation is reliable if *p* is less than .05 ($p < .05$). If $p > .05$, then we must cautiously conclude that the results may have been due to chance. This *p*-value takes into account the size of the sample and the strength of the correlation to quantify whether the result is reliable or due to chance.

To find a correlation in SPSS, go to the Analyze menu, point to Correlate, and choose Bivariate. To practice, we'll find the correlation between Neuroticism and Extraversion. Select each variable and move it to the pane on the right side of the pop-up window. Then, click OK.





Your Output should look something like this:

Correlations

		Self-report Neuroticism	Self-report Extraversion
Self-report Neuroticism	Pearson Correlation	1	-.374
	Sig. (2-tailed)		.105
	N	20	20
Self-report Extraversion	Pearson Correlation	-.374	1
	Sig. (2-tailed)	.105	
	N	20	20

In the area on the right side of the Output, you see four boxes, which contain various statistics. Much of this information is repetitive. For example, the upper left box is the same as the lower right box. The upper right box is the same as the lower left box.

The upper left box (and lower right box) shows the correlation between Neuroticism and itself. Any variable correlates perfectly with itself ($r = 1.00$). The sample size is 20. This box is not particularly useful. It remains a mystery why SPSS includes it.

The upper right box (and lower left box) is more interesting. Three statistics are presented. The value $-.374$ is the correlation between neuroticism and extraversion (labeled ‘Pearson Correlation’ off to the left). It is negative. As neuroticism increases, extraversion decreases. In other words, people who are more emotionally unstable are also less social. The correlation is modest. The next value $.105$ is the p -value (labeled ‘Sig 2-tailed’ off to the left, indicating that the value is for a significance test). If $p < .05$, the finding is reliable. Here the p -value is too large. The finding may be due to chance. In other words, the correlation is not trustworthy, probably because this sample is way too small. The sample size, presented again, is 20.

You are now ready to begin working independently on your homework assignment, “LA1” (see “Due” column on the Course Calendar). Use the APA Style guide on the following pages to assist in reporting results.

Dismissal

Students can leave early if they are able to finish LA1 in its entirety.

APA Style Guide

Note: You have my permission to copy any or all of this writing for this or future assignments.

Style and Rounding

Rules governing rounding vary considerably from discipline to discipline. These guidelines reflect the current norms in psychology.

p-values. Historically, published articles either reported statistically significant findings as “ $p < .05$ ” and non-significant findings as “*ns*” – this was a very imprecise way of reporting the results. All major psychology journals now advocate reporting actual *p*-values when they are provided in text (in tables and figures asterisks are still common). In general, *p*-values should be reported rounded to two decimals (e.g., $p = .08$ or $p = .02$). However, if the *p*-value is less than .01, report three decimals (e.g., $p = .009$ or $p = .002$). If the *p*-value is less than .001, simply report as “ $p < .001$ ” (note, SPSS strangely reports these as .000, but it is impossible to have a probability of zero, so do not report it that way).

Percentages. Usually percentages are rounded to one decimal place (e.g. 88.6% or 1.1%).

Other statistics. In general, all other statistics are rounded to two decimal places (e.g., $M = 1.46$ or $r = -.33$)

Leading zero. If a statistic is a decimal, people usually include a leading zero only if the statistic can commonly exceed 1.0. For example, most descriptive statistics, as well as *t*-scores, *Z*-scores, and *F*-scores commonly exceed 1.0, so even if an observed value is a decimal, a leading zero is included (e.g., $Z = 0.23$ or $t = 0.96$). In contrast, correlations cannot exceed 1.0, so no leading zero is included (e.g., $r = .23$ or $r = .96$).

Italics. Statistical symbols should be italicized (e.g., *M*, *SD*, *r*, *t*, *d*, *F*, *p*, etc.) but not the numbers following them.

Descriptive Statistics

On the 1-9 depression severity scale, the sample reported a mean score of 4.66 ($SD = 1.59$), with 12.3% reporting a “1” (not at all depressed) and 2.1% reporting a “9” (completely depressed).

The sample was predominantly white (94.5%) and college-educated (86.6%), more often female (61.3%), and distributed relatively evenly across the U.S. (North: 24%, South: 30%, Midwest: 20%, West: 26%).

Participants varied considerably in age ($M = 35.5$ years, $SD = 10.2$, ranging from 18 to 77).

Participants identified as Democrats (30.2%), Republicans (19.8%), or Independents (50.0%).

Correlation (Significant, $p < .05$)

Note: Include the correlation, p -value, a description of the direction (more, less, positively, negatively, directly, inversely, etc.), and a description of the effect size (e.g., near-zero/marginal, small/slight, medium/moderate/modest, strong/large/sizeable). If the finding might be confusing to a non-statistician, include a second sentence explaining the finding in simpler terms.

Participants who were more neurotic reported exercising moderately less often, which was statistically significant, $r = -.35$, $p = .02$.

Quarterbacks who were taller had marginally better completion rates, $r = .09$, $p = .04$. Thus, tall quarterback throw completed (caught) passes more often than short quarterbacks.

Anxiety and depression were strongly positively correlated ($r = .71$, $p = .007$). Therefore, it could be difficult to distinguish between whether someone's primary diagnosis should be an anxiety disorder or a mood disorder.

Correlation (Non-Significant, $p > .05$)

Age was not significantly associated with income ($r = .13$, $p = .23$), political views ($r = .01$, $p = .99$), or vocabulary ($r = .06$, $p = .62$).

The present study failed to find an association between wealth and happiness, $r = .08$, $p = .64$.