

Descriptive Statistics Example in SPSS

The datafile **Census.sav** contains data for $N = 360$ individuals on 11 different variables.

Two of the variables for which we have data are *education level* (i.e., the highest degree that each person has completed) and *time spent online per day* (in minutes).

Let's use the descriptive statistics that we've learned to get a better sense of how our sample "looks" with respect to these two variables.

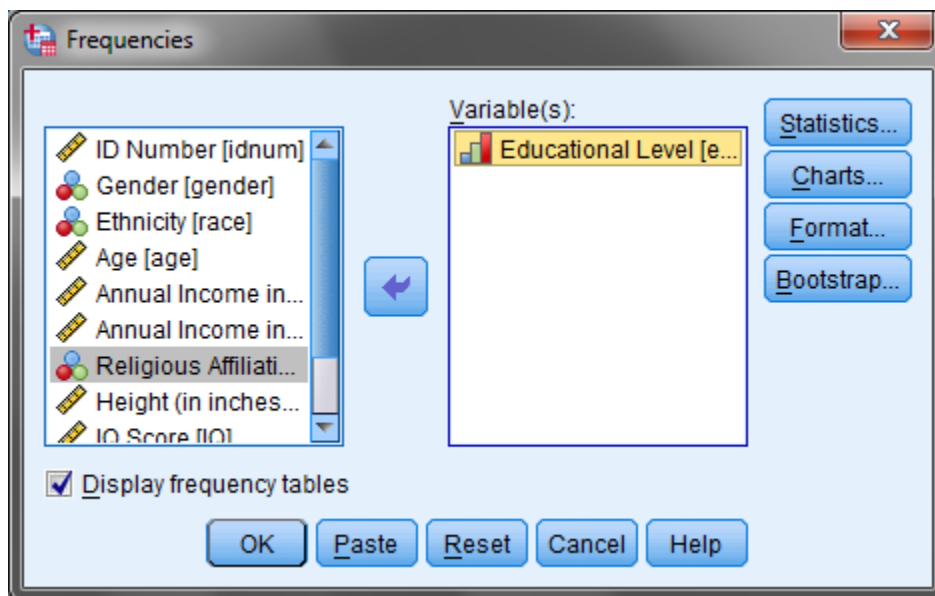
Step 1: What information might we want to know?

Because "education level" is a *categorical* variable, we will likely want to know the *mode*, as well as the *frequency* (i.e., percentage) of people in each category.

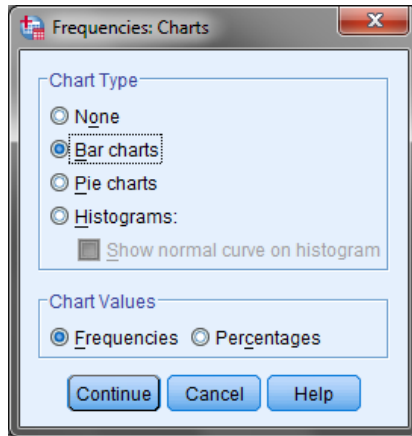
Because "time spent online" is a *continuous* variable, there are many things we might want to know. For example, we might want to know our sample's *mean*, *median*, *range*, *inter-quartile range*, and *standard deviation*. We also might want to know whether the distribution is *skewed*, and whether or not there are any *outliers* in our sample.

Step 2a: How do we get descriptive statistics for categorical variables?

- For *categorical* variables, go to "Analyze" → "Descriptive Statistics" → "Frequencies"



- Select the *categorical* variable for which you want descriptive statistics from the list on the left, and click the arrow button to move it to the “Variable(s)” box on the right.



- If you click the “*Charts...*” button on the right, you can ask SPSS to produce a bar graph of the data for this variable. (This is optional, but is sometimes useful.)

Under “Chart Values,” you can tell SPSS to format this bar graph with either the Frequencies (i.e., the number of people in each category) or the Percentages (i.e., the percentage of people in each category) on the Y-axis.

Step 3a: How do we read the output from SPSS?

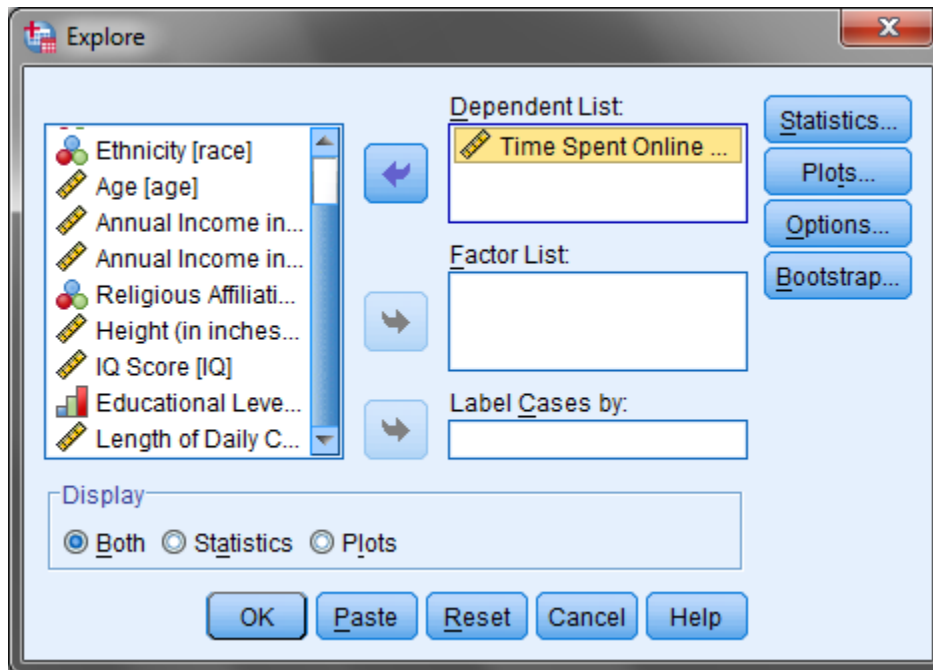
Educational Level					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	High School Degree	72	20.0	20.0	20.0
	College Degree	215	59.7	59.7	79.7
	Masters Degree	47	13.1	13.1	92.8
	Doctoral Degree	26	7.2	7.2	100.0
	Total	360	100.0	100.0	

- 1: There are a total of $N = 360$ people in our sample.
- 2: These are the raw numbers of people in each category. For example, 47 of the 360 people in our sample have a Masters degree, and so on.
- 3: These are the percentages of people in each category. For example, 7.2% of the people in our sample have a Doctoral degree, and so on.

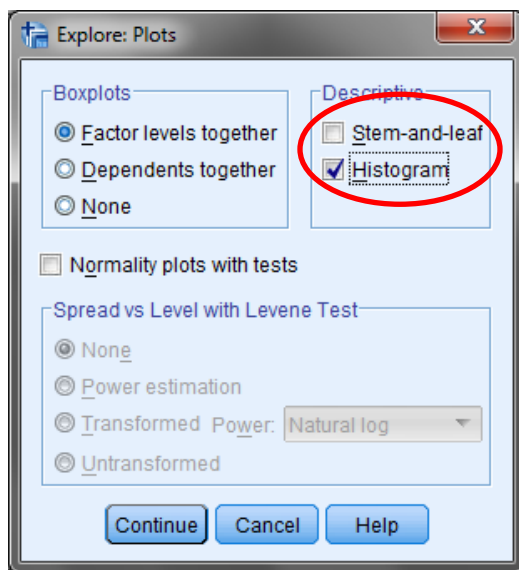
With this information, we can easily figure out the mode, by looking for the largest frequency (or largest percentage). In this case, the mode is “College Degree.”

Step 2b: How do we get descriptive statistics for continuous variables?

- For *continuous* variables, go to “Analyze” → “Descriptive Statistics” → “Explore”



- Select the *continuous* variable for which you want descriptive statistics, and click the first arrow button to move it into the “Dependent List” box.



- If you click the “*Plots...*” button on the right, you will see that the “Stem-and-leaf” box is checked (by default) under “Descriptive.”

If we instead want SPSS to show us a histogram of our data, uncheck the “Stem-and-leaf” box, and check the “Histogram” box.

Step 3b: How do we read the output from SPSS?

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Time Spent Online Daily (in minutes)	360	100.0%	0	0.0%	360	100.0%

4: There are a total of $N = 360$ people in our sample.

Descriptives

		Statistic	Std. Error
Time Spent Online Daily (in minutes)	Mean	57.3750	1.22474
	95% Confidence Interval for Mean	Lower Bound	54.9664
		Upper Bound	59.7836
	5% Trimmed Mean	57.6235	
	Median	58.0000	
	Variance	539.995	
	Std. Deviation	23.23780	
	Minimum	1.00	
	Maximum	109.00	
	Range	108.00	
	Interquartile Range	34.75	
	Skewness	-.172	.129
	Kurtosis	-.513	.256

5: The *mean* amount of time that our sample spends online per day is 57.38 minutes.

6: The *median* time spent online per day is 58.00 minutes.

7: The *standard deviation* of time spent online per day is 23.24 minutes.
[Notice that means and *SDs* are rounded to exactly TWO decimal places.]

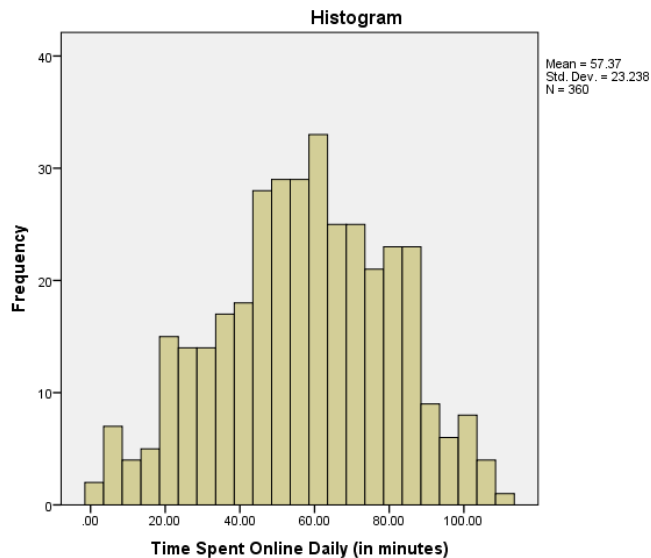
8: The *range* of time spent online per day is 108 minutes (i.e., 1 – 109 minutes).

9: Our *inter-quartile range* is 34.75 minutes.

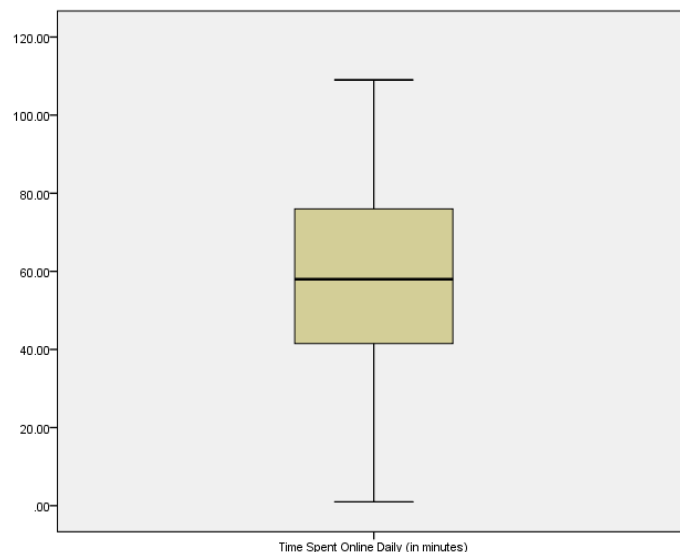
10: Using these two numbers, we can get a sense of whether our distribution is normal or skewed. To do this, we divide Skewness (-.172) by SE_{skew} (.129).

Rule of thumb: If $\text{Skewness} / SE_{\text{skew}}$ is greater than +3.0, our distribution is positively skewed. If this value is less than -3.0, our distribution is negatively skewed. If it is between -3 and +3, we say that our distribution is symmetrical (i.e., not skewed).

In this case, $-.172 / .129 = -1.33$, so we would say that our distribution is symmetrical.



← Here is the histogram of our data for time spent online per day. Notice that it looks pretty normal / symmetrical.



← SPSS also gives us a boxplot of our data. This is where we look to see if there are any **outliers**.

If there were outliers, they would appear as “dots” that fall outside of the “whiskers” of the boxplot.

There are none of these “dots” here, so we don’t have any outliers.