

SPSS: Measures of Central Tendency and Variability. Descriptives¹

The following uses a set of variables from the “1995 National Survey of Family Growth” to demonstrate how to use some procedures available in SPSS PC Version 10.

Measures of central tendency and variability are descriptive statistical tools used to describe the distribution of a variable with just a couple of numbers. Measures of central tendency summarize what is typical or, in the case of the mean, what is average in the distribution. Measures of variability, on the other hand, tell us how much diversity or heterogeneity there is in the distribution of scores.

Requesting measures of central tendency and variability with SPSS:

The NSFG provides us with information on, among other things, the age of the respondents in the sample. In order to summarize the age distribution of these respondents I want to report appropriate measures of central tendency and variability. There are a couple of options for doing this using SPSS.

- **OPTION #1:** The easiest way to get these descriptive statistics is to go to the **Analyze** menu, choose **Descriptive Statistics**, and then **Descriptives....** In the dialog box that appears you can choose any variable in the data set (I'd use *age* in my example) and, once you've hit OK, SPSS will produce descriptive statistics for all of them. Included in these statistics are the number of valid cases, the minimum observed value, the maximum observed value, the mean, and the standard deviation. You will undoubtedly notice that there are some potentially important bits of information missing from the output from the "Descriptives" procedure. Most notably, this procedure does not produce other measures of central tendency (median and mode) or variability (range). These are particularly important if you have variables measured at the nominal or ordinal level. Following is the syntax that you could use to get all of the statistics available with the “Descriptives” procedure.

DESCRIPTIVES

```
VARIABLES=age  
/STATISTICS=MEAN SUM STDDEV VARIANCE RANGE MIN MAX SEMEAN KURTOSIS SKEWNESS .
```

- **OPTION #2:** To get the full slate of central tendency and variability measures you can use the "Frequencies" procedure. At the top of the 'SPSS Data Editor' screen, click on the **Analyze** menu, then **Descriptive Statistics**, and then **Frequencies**. Here again, choose the variable you want statistics for, *age*. Here's the new stuff: once you've selected the variables you want to describe (you can choose as many as you like, by the way), hit the **Statistics** button at the bottom of the dialog window. This will take you to a new dialog box in which you can choose from a variety of measures of central tendency including the mean, median, and mode, and a variety of measures of variability including the range, variance, and standard deviation. After you've selected the measures you want, click on **Continue** and proceed with the frequency-producing process as normal. As before, the output will contain a frequency table for each of the variables you included. This is really unnecessary information for our purposes here, and should be deleted from your output before printing. More important is the first box that appears. It contains all of the descriptive statistics you request for each of the variables included. For example, I requested the mean, median, mode, standard deviation, variance, range, minimum and maximum for the variable *age* and got the following output:

¹Prepared by Kyle Crowder of the Sociology Department of Western Washington University, and modified by Patty Glynn, University of Washington.

Statistics

AGE Age in Years

N	Valid	10847
	Missing	0
Mean		30.5930
Median		31.0000
Mode		32.00
Std. Deviation		8.3072
Variance		69.0103
Range		31.00
Minimum		14.00
Maximum		45.00

Here we can see that the average (mean) age of the women in the sample is not quite 31 years (30.59). The score of the case that falls in the exact middle of the distribution (median) is 31 years, and the most commonly occurring age in the distribution is 32 years. Since this is an interval-ratio level variable with no obvious pronounced skew, I will pay most attention to the mean.

Q1: How do I know that this variable is not dramatically skewed?

A1: In this case, the mean, median, and mode are all fairly close in value, indicating that the distribution is fairly symmetrical. If the distribution were skewed, the mean would be dragged towards the extreme cases and would have a value substantially lower (in the case of a negative skew) or substantially higher (in the case of a positive skew) than the median.

Q2: Why is the mean preferable to the median and the mode?

A2: Unless the distribution is skewed, the mean is preferable to the mode and the median because it takes into consideration the scores for every case in the distribution rather than just the most common score (mode) or the score for the case in the middle of the distribution (median). In other words, the mean takes advantage of all of the available information in summarizing the central tendency of the distribution.

Since I have a variable measured at the interval-ratio level, all of the measures of variability are fair game. The standard deviation indicates (approximately) that the scores in the distribution vary from the mean by an average of about 8.3 years. This corresponds with a variance of 69.01. The youngest person in the sample is 14 years old and the oldest is 45 years old, producing a range of 31 years.

Following is the syntax that could be used to display ALL of the statistics available with the "Frequencies" procedure.

```
FREQUENCIES  
  VARIABLES=age  
  /NTILES= 4  
  /NTILES= 10  
  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM SEMEAN MEAN MEDIAN MODE  
  SUM SKEWNESS SESKEW KURTOSIS SEKURT  
  /GROUPED= age  
  /ORDER= ANALYSIS .
```