

Chapter 6

The Frequencies Procedure: Summarizing Data With Descriptive Statistics

6.1 Introduction

One of the primary reasons for doing research is to be able to make accurate statements regarding the behavior or characteristics of a large number of people. In most cases it is impossible to actually collect data from every member of a target group (referred to as a **population**), so researchers typically collect data from a smaller subset of the population (called a **sample**) and attempt to generalize from the sample to the larger population. For example, 228 participants were selected from the several thousand employees of EZ Manufacturing for the leadership study.

But even when working with relatively smaller samples, it is very difficult for the human mind to comprehend large numbers of individual facts. Few researchers could remember all the individual performance scores of the 228 EZ Manufacturing employees, and it would be difficult to draw meaningful conclusions from this raw data itself. For this reason researchers often begin data analyses by summarizing characteristics of the participants in the sample.

One way to do this is to generate frequency tables of variables. Recall from Chapter 3 that frequency distributions summarize data by listing the number of participants who received scores of the possible values on the variables. Thus, a frequency table of how many employees scored a 1, 2, 3, etc. on the **perform** variable will be much easier to understand and interpret than would a simple listing of all 228 individual performance scores. This process is sometimes called *number crunching*, because large volumes of data are crunched into more manageable, meaningful units. This facilitates drawing conclusions and determining trends in the data (e.g., do most of the employees score towards the high or low end of the performance scale?).

Beyond constructing simple frequency distributions, researchers typically crunch data even further down into a single statistic that is typical of the entire set of scores in some way. These numerical indices are called **descriptive statistics**, because they provide a single number, or index, that best summarizes all of the scores on some dimension. Common descriptive statistics employed are measures of central tendency and variability.

The mean, median, mode, are common measures of central tendency (so called because these indices tend toward the middle, or center, of the distribution). The standard deviation, variance, and range are examples of indices of variability (so called because they describe the how much all the scores vary around the middle of the distribution). These summary statistics go a long towards helping the researcher

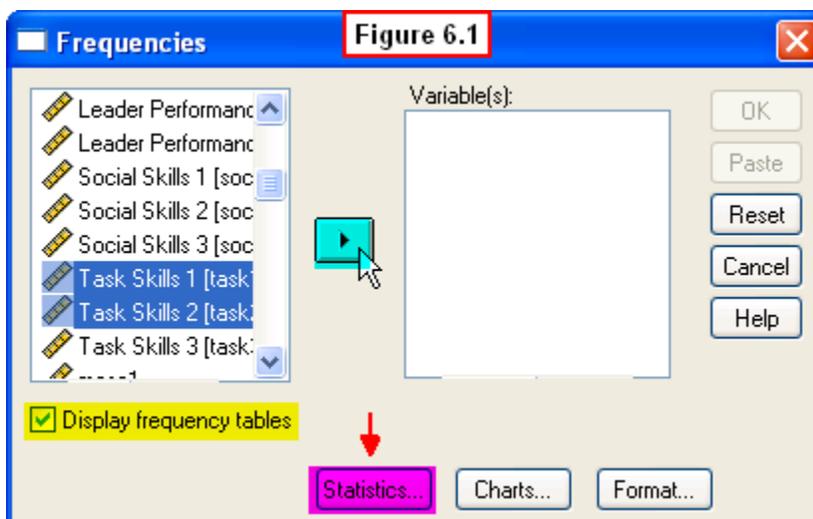
understand the data, and they enable her/him to describe the findings in a brief, precise manner.

Recall that we used the **Frequencies** procedure in Chapter 3 as an introduction to SPSS. In this chapter we will use this procedure to generate frequency tables of two variables in the **ezdata.sav** file. We will also demonstrate how this procedure can be used to generate and interpret descriptive statistics for these variables. Open your **ezdata.sav** file and follow along with the example. You will be asked to do the same analyses on different variables in the file as an exercise at the end of the chapter.

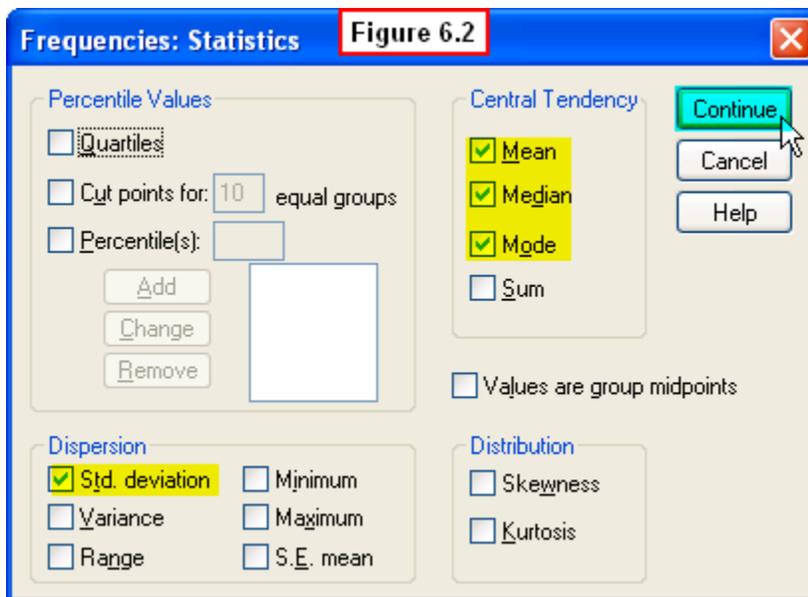
6.2 Running the Frequencies Procedure

Recall our Chapter 4 discussion of social and task skills in leadership. It would be of interest to describe EZ employees on these dimensions. Also recall that you obtained scores on these skills before and after employees attended a leadership training workshop. So it would also be of interest to compare the distribution of employees' scores on these skills before and after the workshop to begin to assess its effectiveness. For this example we will generate descriptive statistics for the before-after scores on task skills, **task1** and **task2**. To begin, select **Analyze, Descriptive Statistics, Frequencies...** from the Data Editor menu.

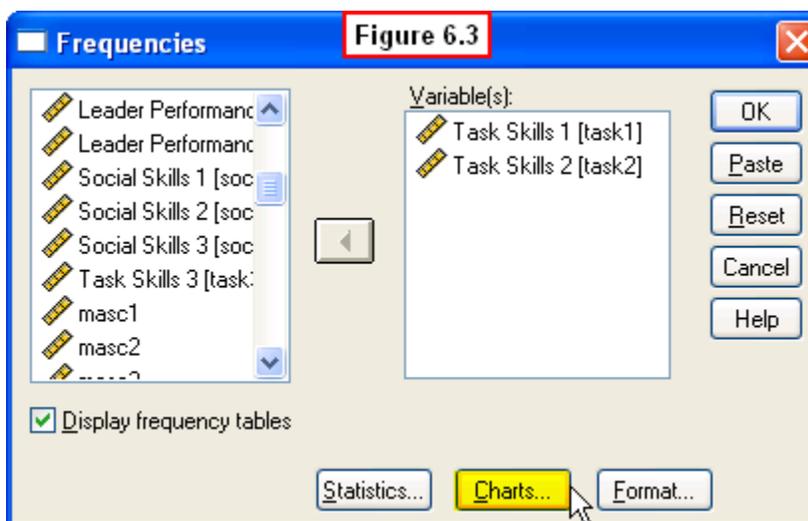
In the left pane of the **Frequencies** dialog window (Figure 6.1), scroll down and highlight **task1** and **task2**. Click the right-arrow in the middle to move these to the **Variable(s):** pane. Make sure the **Display frequency tables** box is checked, then click the **Statistics...** button.



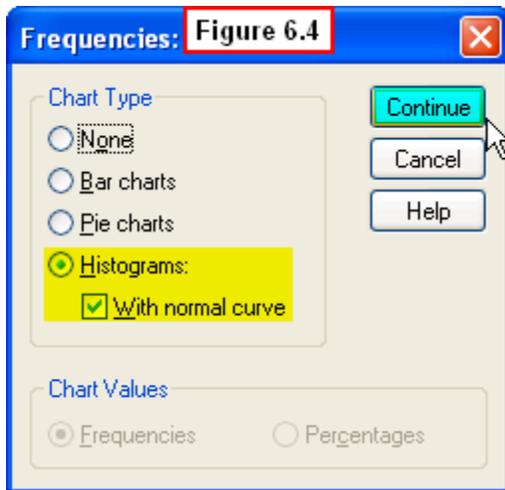
In the **Frequencies: Statistics** window (Figure 6.2), click the Mean, Median and Mode for measures of Central Tendency and Standard Deviation for measures of Dispersion (variability), then click the **Continue** button.



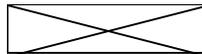
This will return you to the **Frequencies** window. To generate frequency graphs of these variables, click the **Charts...** button at the bottom (Figure 6.3).



In the **Frequencies: Charts** window (Figure 6.4), select **Histograms, with normal curve** (to superimpose a normal curve over the histogram), then click the **Continue** button.



This will return you again to the **Frequencies** dialog window. Now simply click the **OK** button in the upper-right corner of that window.



To view a video of this procedure, visit <http://youtu.be/vGypnHFveTU>

6.3 Interpreting the Output

The descriptive statistics we requested are displayed in the first table of the output file (Figure 6.5). The mean, median and mode of **Task Skills 1** are 5.12, 5.00 and 5.00, respectively. These three measures of central value are in close agreement, indicating that the middle of the distribution happens to be the midpoint of our 9-point scale of task skills. Thus, the typical employee was about average on task skills at the beginning of the study. The standard deviation (1.90) indicates there is variability in task skills (on average, about 2 points from the mean).

Note that after employees attended the leadership training workshop, the measures of central tendency indicate that there was a subsequent overall increase in the mean, median and mode on **Task Skills 2** (5.50, 6.00 and 6.00, respectively). Determining whether or not this is a statistically significant increase in task skills requires use of inferential statistics (a topic we will discuss in a later chapter). However, this comparison of **task1** to **task2** does provide some initial evidence that the workshop was effective, in that the typical employee now has a task skills score one point above the midpoint of the task skills scale. The standard deviation of **task2** (1.94) indicates that there wasn't much change in the variability of **task2** scores compared to that of **task1**.

Figure 6.5

Statistics

		Task Skills 1	Task Skills 2
N	Valid	228	228
	Missing	0	0
Mean		5.1228	5.5000
Median		5.0000	6.0000
Mode		5.00	6.00
Std. Deviation		1.89895	1.94075

Examination of the frequency table of **task1** (Figure 6.6) confirms that the highest frequency of scores was 5 (53 employees, or 23.2%, received this score), and the Cumulative Percent column shows that 59.6% of employees received a score of 5 or lower (and 40.4% scored 6 or higher). We get a sense of the variability in scores by examining the other frequencies/percentages. The scores are fairly evenly distributed within 2 points of this middle, with few scores at either the extreme high or low end of the scale.

Figure 6.6

Task Skills 1

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1.00	9	3.9	3.9	3.9
2.00	6	2.6	2.6	6.6
3.00	34	14.9	14.9	21.5
4.00	34	14.9	14.9	36.4
→ 5.00	53	23.2	23.2	→ 59.6
6.00	34	14.9	14.9	74.6
7.00	33	14.5	14.5	89.0
8.00	16	7.0	7.0	96.1
9.00	9	3.9	3.9	100.0
Total	228	100.0	100.0	

Examination of the frequency table of **task2** (Figure 6.7) confirms that the highest frequency of scores increased to a value of 6 (55 employees, or 24.1%, received this score), and the Cumulative Percent column shows that now only 42.1% of employees received a score of 5 or lower (and now 57.9% scored 6 or higher). Examining the other frequencies/percentages, we see that the scores are fairly evenly distributed within 2 points of this middle, with fewer scores at either the extreme high or low end of the scale. However, there appeared to be a substantial increase in scores at the high end of the scale compared to **task1**. That is, 17.1% have scores of 8 or higher on **task2**, compared to 10.9% with 8 or higher on **task1**. This is additional evidence of the effectiveness of the leadership training workshop.

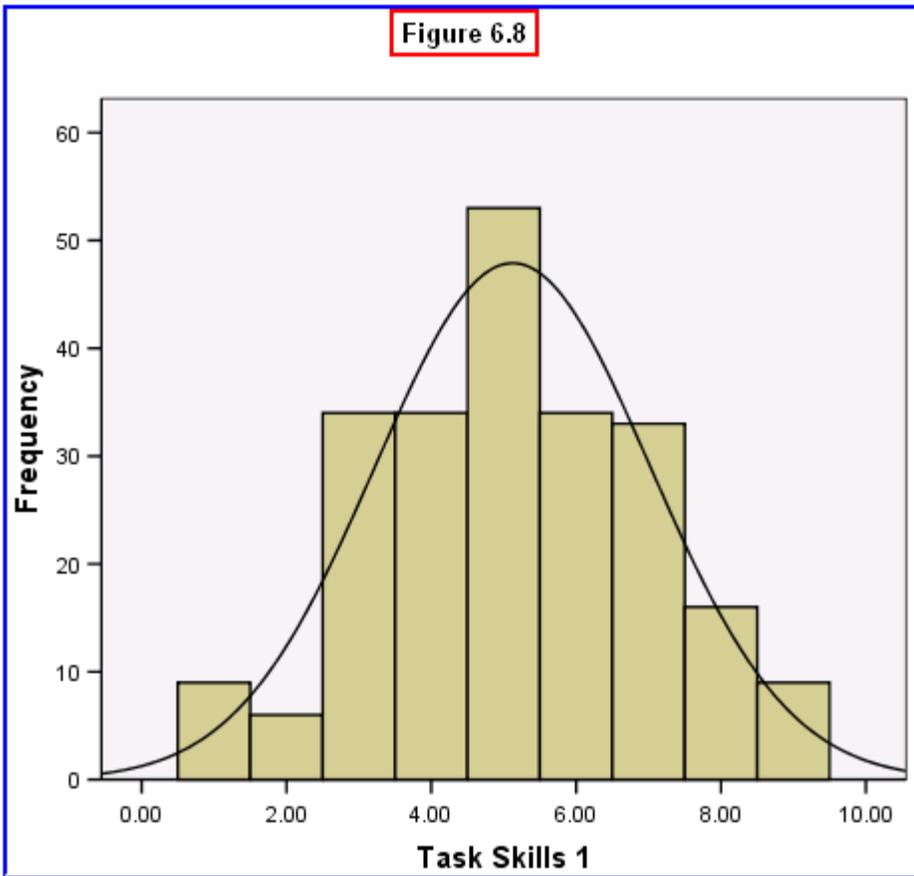
Figure 6.7

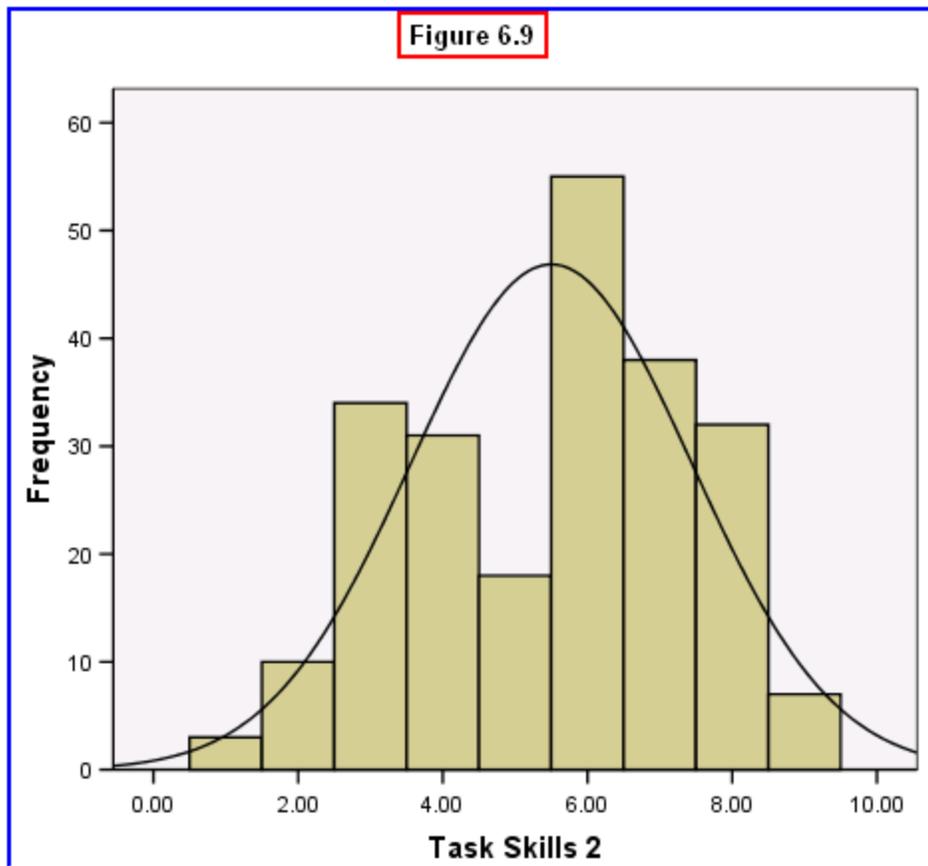
Task Skills 2

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1.00	3	1.3	1.3	1.3
2.00	10	4.4	4.4	5.7
3.00	34	14.9	14.9	20.6
4.00	31	13.6	13.6	34.2
5.00	18	7.9	7.9	42.1
→ 6.00	55	24.1	24.1	66.2
7.00	38	16.7	16.7	82.9
8.00	32	14.0	14.0	96.9
9.00	7	3.1	3.1	100.0
Total	228	100.0	100.0	

The histograms we generated visually depict this shift towards the higher end of the task skills scale (Figures 6.8 and 6.9).

Figure 6.8





Examination of these graphs reveals that the distribution has shifted to the higher end of the scale on **task2** scores. The bars depicting frequencies of scores 6 or higher are much taller in the second figure than in the first, reflecting an increased number of employees with high scores on task skills after attending the workshop.

More could be said about these results, but this has shown the value of the **Frequencies** procedure and computing **descriptive statistics** in helping the researcher gain an understanding of the data from a project such as this. This procedure is often the first step in analyses, and can be done for most variables in the data file. It can lead to some preliminary conclusions early on in the analysis phase, and it is sometimes used for exploratory purposes to generate suggestions for subsequent analyses (e.g., determining whether or not the increase in mean task skills scores is a statistically significant one).

To view a video review of this chapter's procedures, visit <http://youtu.be/vGypnHFveTU>