

Chapter 3

Descriptive Statistics

In Chapter 2, we discussed many of the options available in SPSS for dealing with data. Now we will discuss ways to summarize our data. The procedures used to describe and summarize data are called **descriptive statistics**.

Section 3.1 Frequency Distributions and Percentile Ranks for a Single Variable

Description

The *Frequencies* command produces frequency distributions for the specified variables. The output includes the number of occurrences, percentages, valid percentages, and cumulative percentages. The valid percentages and the cumulative percentages comprise only the data that are not designated as missing.

The *Frequencies* command is useful for describing samples where the **mean** is not useful (e.g., **nominal** or **ordinal** scales). It is also useful as a method of getting the feel for your data. It provides more information than just a **mean** and **standard deviation** and can be useful in determining **skew** and identifying **outliers**. A special feature of the command is its ability to determine **percentile ranks**.

Assumptions

Cumulative percentages and **percentiles** are valid only for data that are measured on at least an **ordinal scale**. Because the output contains one line for each value of a variable, this command works best on variables with a relatively small number of values.

Drawing Conclusions

The *Frequencies* command produces output that indicates both the number of cases in the sample of a particular value and the percentage of cases with that value. Thus, conclusions drawn should relate only to describing the numbers or percentages of cases in the sample. If the data are at least ordinal in nature, conclusions regarding the cumulative percentage and/or **percentiles** can be drawn.

SPSS Data Format

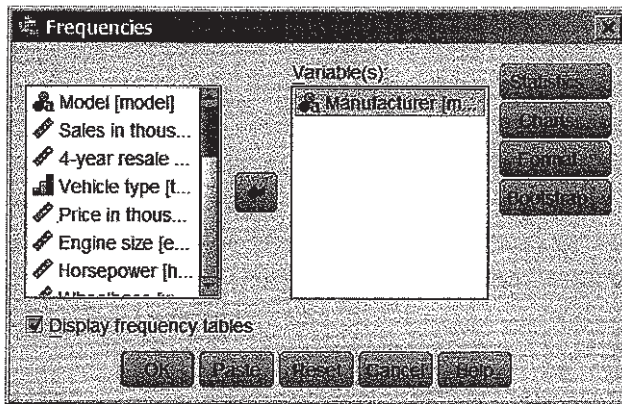
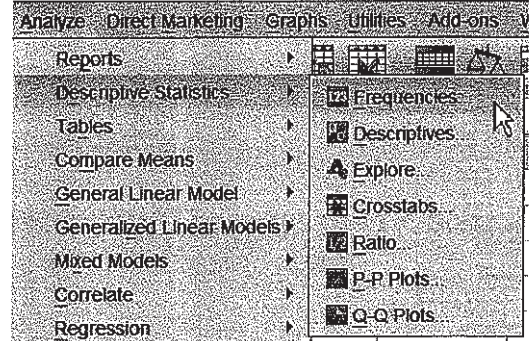
The SPSS data file for obtaining frequency distributions requires only one variable, and that variable can be of any type.

Creating a Frequency Distribution

This example uses the CAR_SALES.sav data file that comes with SPSS. If you are using SPSS Version 20, it will be located at

C:\Program Files\IBM\SPSS\Statistics\20\Samples\English\car_sales.sav

To run the *Frequencies* command, click *Analyze*, then *Descriptive Statistics*, then *Frequencies*. This will bring up the main **dialog box** below. Transfer the variable for which you would like a frequency distribution into the *Variable(s)* blank shown to the right (MANUFACTURER in this case). Be sure that the *Display frequency tables* option is checked. Click *OK* to receive your output.



Note that the **dialog boxes** in SPSS show both the type of variable (the icon immediately to the left of the variable name) and the variable labels if they are entered. Thus, the variable MANUFACTURER shows up in the **dialog box** as *Manufacturer*.

Click *OK* to run the test.

Output for a Frequency Distribution

Statistics		
Manufacturer		
N	Valid	157
	Missing	0

The output consists of two sections. The first section indicates the number of records with **valid data** for each variable selected. Records with a blank score are listed as missing. In this example, the data file contained 157 records.

The second section of the output contains a cumulative frequency distribution for each variable selected. At the top of the section, the variable label is given. The output itself consists of five columns. The first column lists the values of the variable sorted in alphabetical order. There is a row for each value of your variable, and additional rows are added at the bottom for the Total and Missing data. The second column gives the frequency of each value, including

Manufacturer					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Acura	4	2.5	2.5	2.5
	Audi	3	1.9	1.9	4.5
	BMW	3	1.9	1.9	6.4
	Buick	4	2.5	2.5	8.9
	Cadillac	5	3.2	3.2	12.1
	Chevrolet	9	5.7	5.7	17.8
	Chrysler	7	4.5	4.5	22.3
	Dodge	11	7.0	7.0	29.3
	Ford	11	7.0	7.0	36.3
	Honda	5	3.2	3.2	39.5
	Hyundai	3	1.9	1.9	41.4
	Infiniti	1	.6	.6	42.0
	Jaguar	1	.6	.6	42.7
	Jeep	3	1.9	1.9	44.6
	Lexus	6	3.8	3.8	48.4
	Lincoln	3	1.9	1.9	50.3
	Mercedes-Benz	9	5.7	5.7	56.1
	Mercury	6	3.8	3.8	59.9
	Mitsubishi	7	4.5	4.5	64.3
	Nissan	7	4.5	4.5	68.8
	Oldsmobile	6	3.8	3.8	72.6
	Plymouth	4	2.5	2.5	75.2
	Pontiac	6	3.8	3.8	79.0
	Porsche	3	1.9	1.9	80.9
	Saab	2	1.3	1.3	82.2
	Saturn	5	3.2	3.2	85.4
	Subaru	2	1.3	1.3	86.6
	Toyota	9	5.7	5.7	92.4
	Volkswagen	6	3.8	3.8	96.2
	Volvo	6	3.8	3.8	100.0
	Total	157	100.0	100.0	

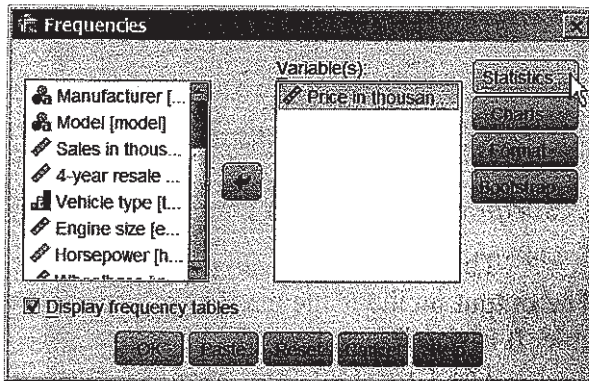
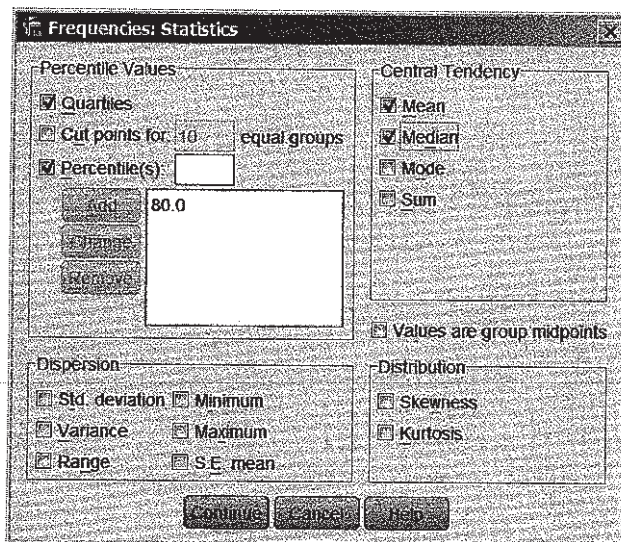
missing values. The third column gives the percentage of all records (including records with missing data) for each value. The fourth column, labeled *Valid Percent*, gives the percentage of records (without including records with missing data) for each value. If there were any missing values, these values would be larger than the values in column three because the total number of records would have been reduced by the number of records with missing values. The final column gives cumulative percentages. Cumulative percentages indicate the percentage of records with a score equal to or smaller than the current value. Thus, the last value is always 100%. These values are equivalent to **percentile ranks** for the values listed. Note: SPSS will provide cumulative percentages even for nominal variables and other variables where it does not make sense.

Determining Percentile Ranks

The *Frequencies* command can be used to provide a number of **descriptive statistics**, as well as a variety of percentile values (including **quartiles**, cut points, and scores corresponding to a specific **percentile rank**). Let us look at statistics for the variable PRICE.

To obtain either the descriptive or percentile functions of the *Frequencies* command, click the *Statistics* button at the top right of the main **dialog box**.

This brings up the *Frequencies: Statistics dialog box* below. Note that the *Central Tendency* and *Dispersion* sections of this box are useful for calculating values, such as the **median** or **mode**, that cannot be calculated with the *Descriptives* command (see Section 3.3). Check any additional desired statistic by clicking on the blank next to it. (In our example here, let us determine the **quartiles**, 80th **percentile**, **mean**, and **median**.)



For **percentiles**, enter the desired **percentile rank** in the blank to the right of the *Percentile(s)* label. Then, click *Add* to add it to the list of **percentiles** requested. Once you have selected all your required statistics, click *Continue* to return to the main **dialog box**. Click *OK*.

Output for Percentile Ranks

The **Statistics dialog box** adds on to the first section of output from the *Frequencies* command. The output contains a row for each piece of information you requested. In the previous example, we checked *Quartiles* and asked for the 80th percentile. Thus, the output contains rows for the 25th, 50th, 75th, and 80th percentiles.

Statistics		
Price in thousands		
N	Valid	155
	Missing	2
Mean		27.39075
Median		22.79900
Percentiles	25	17.89000
	50	22.79900
	75	31.96500
	80	36.21020

Practice Exercise

Using Practice Data Set 1 in Appendix B, create a frequency distribution table for the mathematics skills scores. Determine the mathematics skills score at which the 60th percentile lies.

Section 3.2 Frequency Distributions and Percentile Ranks for Multiple Variables

Description

The *Crosstabs* command produces frequency distributions for multiple variables. The output includes the number of occurrences of each combination of levels of each variable. It is possible to have the command give percentages for any or all variables.

The *Crosstabs* command is useful for describing samples where the **mean** is not useful (e.g., **nominal** or **ordinal scales**). It is also useful as a method for getting a feel for your data.

Assumptions

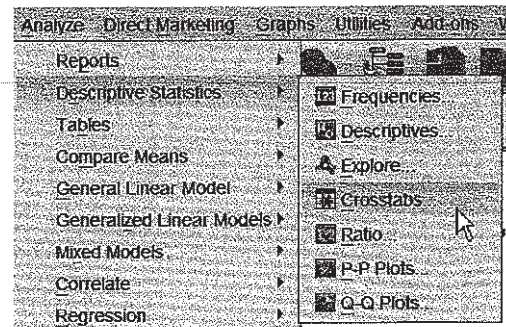
Because the output contains a row or column for each value of a variable, this command works best on variables with a relatively small number of values.

SPSS Data Format

The SPSS data file for the *Crosstabs* command requires two or more variables. Those variables can be of any type.

Running the Command

This example uses the *SAMPLE.sav* data file, which you created in Chapter 1. To run the procedure, click *Analyze*, then *Descriptive Statistics*, then *Crosstabs*. This will bring up the main *Crosstabs dialog box*, which is shown on the next page.¹



¹ The *Exact* button is not available in SPSS Base and is not used in this text.