4. Descriptive Statistics: Measures of Variability and Central Tendency

Objectives

♦ Calculate descriptive for continuous and categorical data
♦ Edit output tables

Although measures of central tendency and variability were presented as separate chapters in the *Fundamentals* text, they are presented together here because they are options located in the same command windows in SPSS. Descriptive statistics are calculated using the **Analyze** menu. Most are calculated using either the **Descriptives** or **Frequencies** command under **Descriptive Statistics**. When calculating descriptives for more complex designs including more than one independent variable, you can also use the **Means/ Compare Means** or the **Descriptive Statistics/ Crosstabs** command which allow you to calculate descriptive statistics of subgroups.

It is always important to take a moment to think about the type of data you are using and what descriptive statistics will be most useful given the type. For continuous or measurement data, you typically report measures of central tendency and measures of variability. For categorical data (i.e., nominal data) you typically report the frequency of each value. Though you don't typically report the frequencies for continuous data, it is often useful to observe the frequency distributions or histograms of continuous distributions to note if they are normal or skewed.
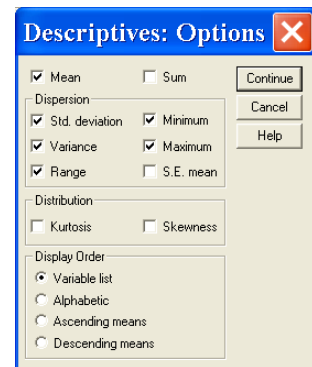
Descriptive Statistics

Let's begin by calculating descriptive statistics for the data in Appendix D which can be found on the web as appendix.dav. (In some editions of these books the file is referred to as "Appendix Data Set" or as Add.dat or as ADD.dat.) In this data set, I think of ADD symptoms, IQ score, English grade, and GPA as continuous variables. We'll calculate measures of central tendency and variability for each of these.

✔ **Open** *appendixd.sav*.

✔ In the **Analyze** menu, select **Descriptive Statistics** and then **Descriptives**.
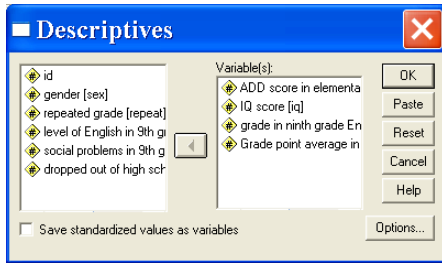
✔ Select each of the continuous variables by either double clicking them, which automatically puts them in the **Variable** list, highlight them one at a time by single clicking them and then clicking the arrow to shift them into the variable list, or by holding the control key down while highlighting all of the variables of interest and then shifting them into the variable list all at once by clicking the arrow. Then click **Options**.

✔ Select each of the measures you've been learning about (**Mean**, **Std. deviation**, **Variance**, **Range**, **Minimum** and **Maximum**). Then, select the **Display Order** you would prefer. This will determine the order they appear in for the resulting table. I like them in the order I indicated in the **Variable list**. Then click **Continue**.


✔ In the main descriptives dialog box, check the box that says **Save standardized values as variables**. SPSS will calculate z scores for each of the variables using the formula you learned about and append them to the end of your data file. Click **Ok**. The resulting output will look like this. Note that the variable labels are used rather than the variable names. Remember, we specified this as the default in Edit/Options/Output Labels.
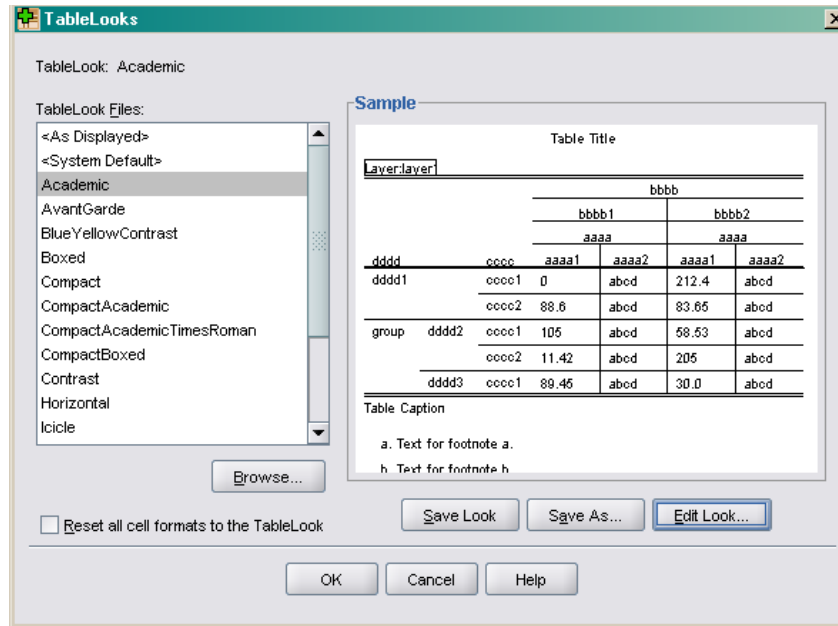
Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
ADD score	88	59.0	26.0	85.0	52.602	12.4222	154.311
English Grade	88	4	0	4	2.66	.945	.894
GPA	88	3.33	.67	4.00	2.4562	.86143	.742
IQ	88	62.00	75.00	137.00	100.2614	12.98496	168.609
Valid N (listwise)	88						

✔ Double click the table so you can edit it. As was the case with graphs, SPSS has many options to edit statistics in tables as well.   Let's try some of them.

✔ Under **Pivot**, select **Transpose Rows and Columns**. Which orientation do you prefer? I like the first since it's more conventional, so I will **Transpose the Rows and Columns** again to return to the original orientation.

✔ Now, click on **Format/Table properties**.   Take a moment to view all of the options in this dialog box. "General" allows you to specify the width of row and column labels. "Footnotes" allows you to chose numeric or alphabetic labels and subscript or superscript as the position for those labels. "Cell formats" allows you to change the font style and size, color, and the alignment. "Borders" allows you to add or remove borders around rows, columns, and even cells. "Printing" allows you to select options such as rescaling tables to fit on paper. After you've viewed the options, hit **Cancel**.
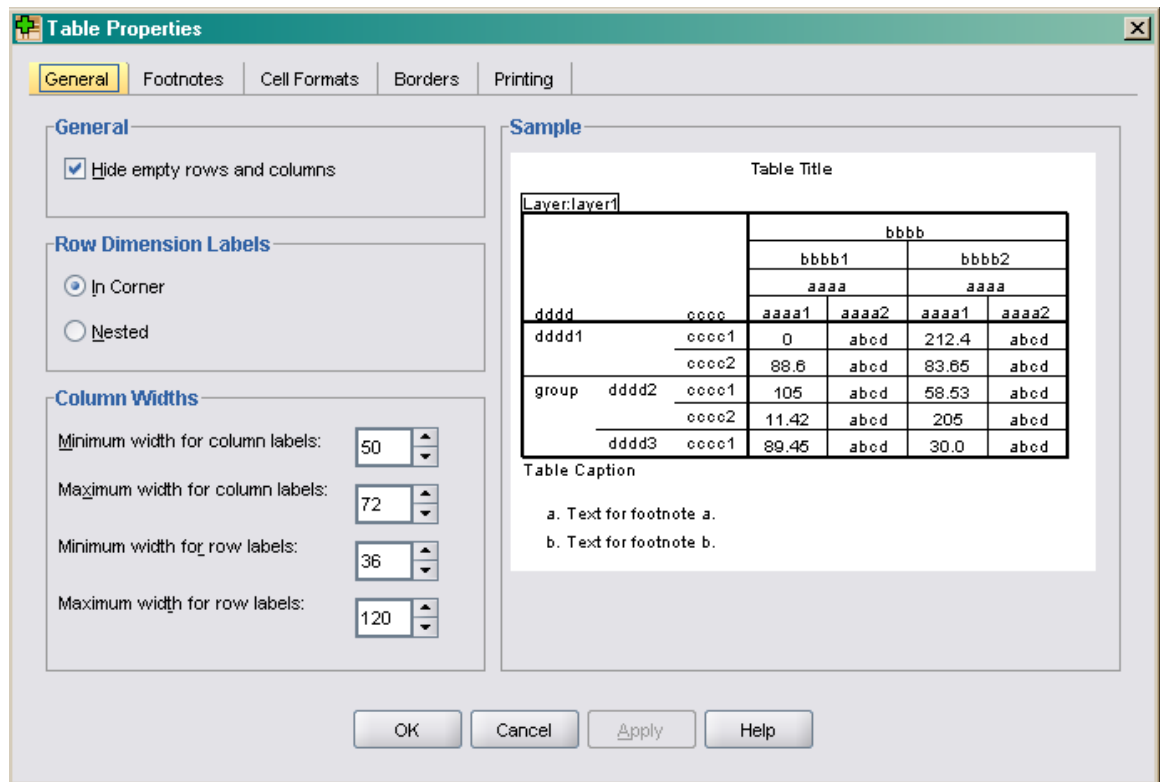
✔ Now, select **Format/Table Looks**. Scroll through the TableLook Files and look at the samples. Select one you like and click **Ok**. I chose **Academic**.

✔ The resulting table is below. I could edit each individual cell by double clicking on it and then edit the text. For example, I could alter each statistic to include 2 decimal places if I wanted. You try it.

Descriptive Statistics

|	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
ADD score	88	59.0	26.0	85.0	52.602	12.4222	154.311
English Grade	88	4	0	4	2.66	.945	.894
GPA	88	3.33	.67	4.00	2.4562	.86143	.742
IQ	88	62.00	75.00	137.00	100.2614	12.98496	168.609
Valid N (listwise)	88						

✔ Now, click on **Window/SPSS Statistics Data Editor** and look at the standardized values (z scores) SPSS added to your file. A brief portion of the Data Editor appears below. You can see that SPSS named each variable with a z. SPSS also labeled the new variables. Check this out in Variable View.
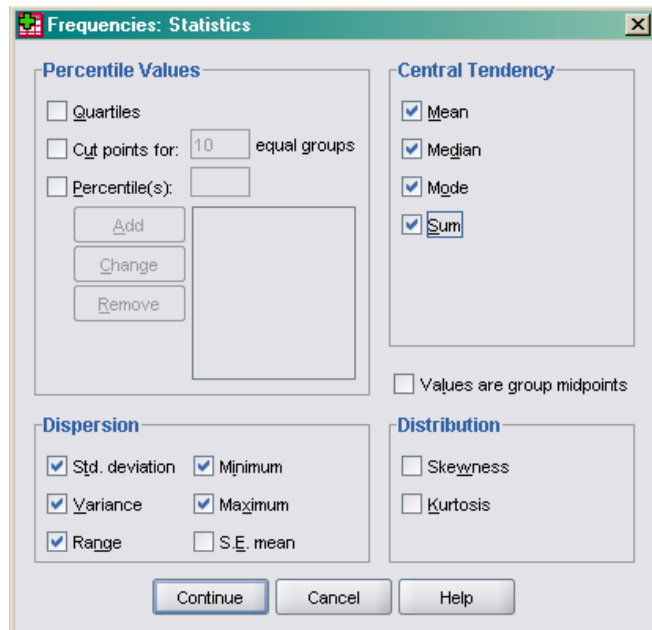
Frequencies

Now, we'll use the frequencies command to help us examine the distributions of the same continuous variables.

✓ Select **Analyze/Descriptive Statistics/Frequency**.
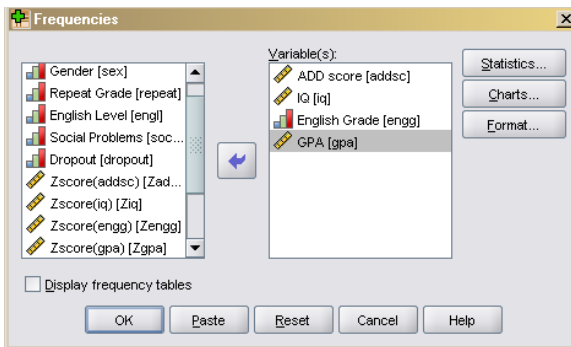
✓ Put the variables of interest in the **Variable** list box. Unselect **Display frequency tables**, because this will be a list of the frequency of every value. (Ignore what looks like an error message.) Click on **Charts**, select **Histogram with normal curve**

✓ Click on **Statistics.** This dialog box has all of the same options we selected under Descriptives earlier. However, the Descriptives dialog box did not include the median and mode. Select all of the statistics of interest and click **Continue.** Then, click **Ok**. A sample of the output follows.

Frequencies

Statistics

		ADD score	IQ	English Grade	GPA
N	Valid	88	88	88	88
	Missing	0	0	0	0
Mean		52.602	100.2614	2.66	2.4563
Median		50.000	100.0000	3.00	2.6350
Mode		50.0	95.00	3	3.00
Std. Deviation		12.4222	12.98496	.945	.86143
Variance		154.311	168.609	.894	.742
Range		59.0	62.00	4	3.33
Minimum		26.0	75.00	0	.67
Maximum		85.0	137.00	4	4.00
Sum		4629.0	8823.00	234	216.15

Histogram

ADD score

Mean =52.6
Std. Dev. =12.422
N =88
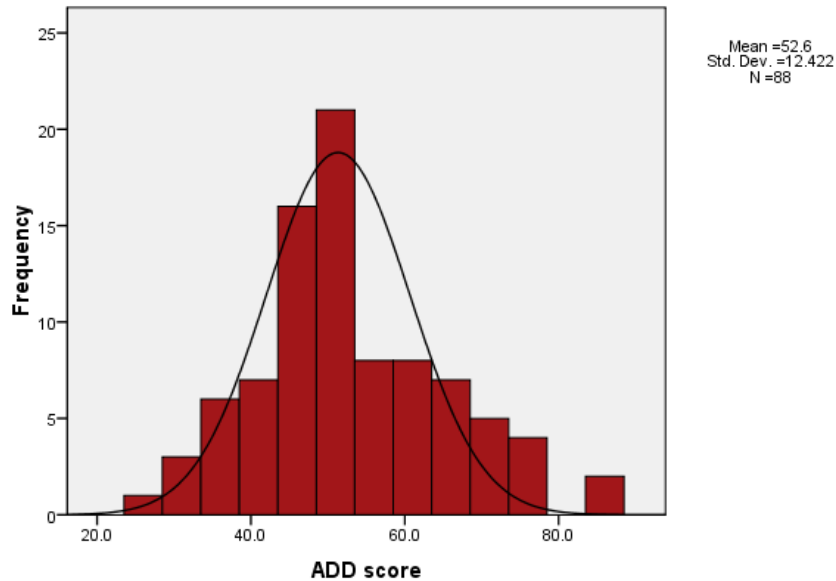
Take a moment to review the output. It looks like ADD is somewhat normally distributed, though a bit negatively skewed. Looking at your own output, are the other variables normally distributed? I also remember now that English grade is nominal too. Variables were scored as A, B, C, D, and F, though coded as 1 - 4. As noted in the text,

we could analyze this as continuous data, but it seems that reporting the frequencies rather than measures of central tendency and dispersion may be more appropriate for this variable.

As before, you can edit the tables or the graphs by double clicking on them. One difference we have seen between the **Descriptives** and **Frequencies** options is that **descriptives** only include mean for measures of central tendency whereas **Frequencies** include the mean, median, and mode. Further, **Descriptives** does not have any built in graphing options, but **Frequencies** does.

Now let's use Frequencies to describe categorical data.

✔ Select **Analyze/Descriptive Statistics/Frequencies**.

✔ This time, put gender, level of English class, English grade, repeated a grade, social problems, and drop out status in the variable list. Select **Display frequency table**. Since there is a finite number of values, we want to know how many people fit in every category. Click on **Statistics** and unselect all of the options because we decided that measures of central tendency and variability are not useful for these data. Then click **Continue**. Next, click on **Charts**. Click on **Bar chart** and select **Percentages** as the Chart Values. Click **Continue** and then **Ok**. A sample of the resulting output is below. Take a moment to review it.

repeat

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 Did not repeat	76	86.4	86.4	86.4
	1 Did repeat	12	13.6	13.6	100.0
	Total	88	100.0	100.0	

English Level

|		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	College Prep	14	15.9	15.9	15.9
	Geneeral	64	72.7	72.7	88.6
	Remedial	10	11.4	11.4	100.0
	Total	88	100.0	100.0	

English Grade
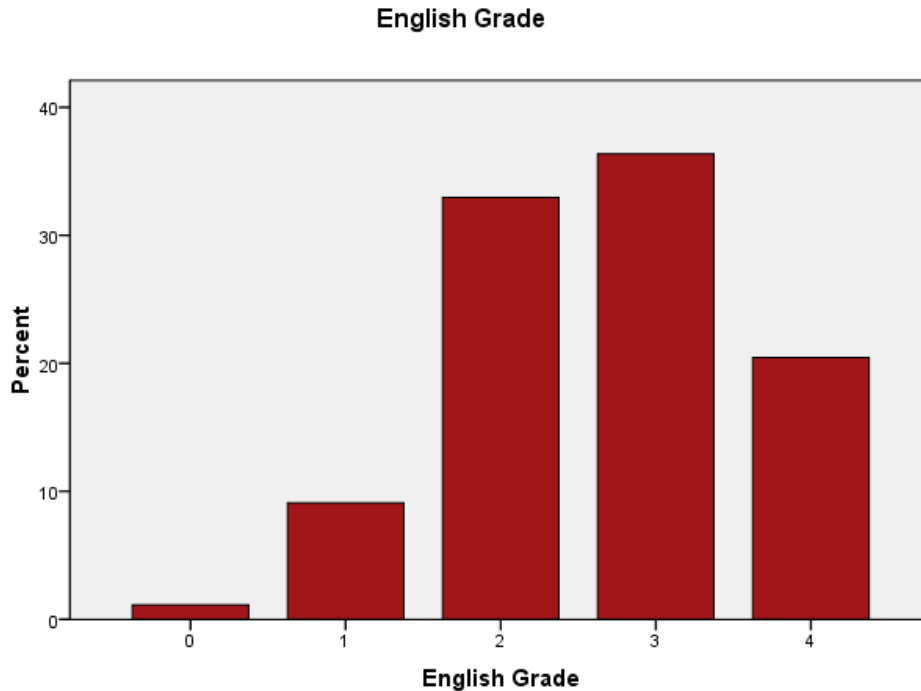
Notice that the frequency tables include a column labeled **Percent** and another labeled **Valid percent**. This is an important distinction when you have missing cases. The percent column indicates the percent of cases in each category out of those cases for which there is complete data on the variable. Valid percent indicates the percent of cases in each category out of the total number of cases, even if some data are missing. For example, imagine a sample of 100 students. Fifty cases are women, 40 are men, and 10 are missing the data. The percent of men would be 44.4%, but the valid percent of men would be 40%. Which do you believe is the more accurate way to describe the sample? I'd argue the valid percent. Now let's move on to a more complicated type of frequency table.

Crosstabs

Sometimes we need to know the number and percent of cases that fall in multiple categories. This is useful when we have multiple categorical variables in a data set. For example, in the data set we have been using, I'd like to know what percent of dropout and nondropout students had social problems. We'll use crosstabs to calculate this.

✔ Click Analyze/Descriptive Statistics/Crosstabs.
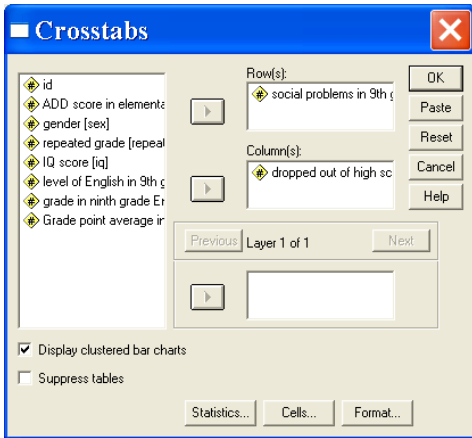
✔ Select social problems for **Rows** and dropped out for **Columns**. Click on **Cells** and select **Observed** for **Counts**, and select **Row**, **Column**, and **Total** under **Percentages**. The click **Continue**. Let's select **Display clustered bar charts** to see if we find this option useful. Then, click **Ok**. The output follows. You can edit both the table and the chart as you have learned.

Social Problems * Dropout Crosstabulation

|			Dropout 0	Dropout 1	Total
Social Problems	0	Count	73	5	78
		% within Social Problems	93.6%	6.4%	100.0%
		% within Dropout	93.6%	50.0%	88.6%
		% of Total	83.0%	5.7%	88.6%
	1	Count	5	5	10
		% within Social Problems	50.0%	50.0%	100.0%
		% within Dropout	6.4%	50.0%	11.4%
		% of Total	5.7%	5.7%	11.4%
Total		Count	78	10	88
		% within Social Problems	88.6%	11.4%	100.0%
		% within Dropout	100.0%	100.0%	100.0%
		% of Total	88.6%	11.4%	100.0%

Bar Chart
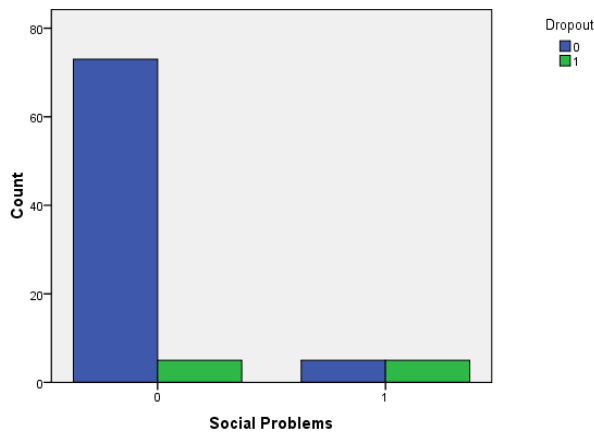
Both the table and the graph show that of those youth with social problems, an equal number did and did not ultimately drop out. This suggests that social problems in ninth grade and drop out status are independent, something we can test later using chi square.

Compare Means

Now, let's consider a case where we want to describe a continuous variable but at different levels of a categorical variable. This is often necessary when you are comparing group means. For example, we can compare ADD symptoms for males and females. Let's try it together.

✔ Select **Analyze/Compare Means/Means**. Notice this is the first time we haven't selected Descriptive Statistics in this chapter.
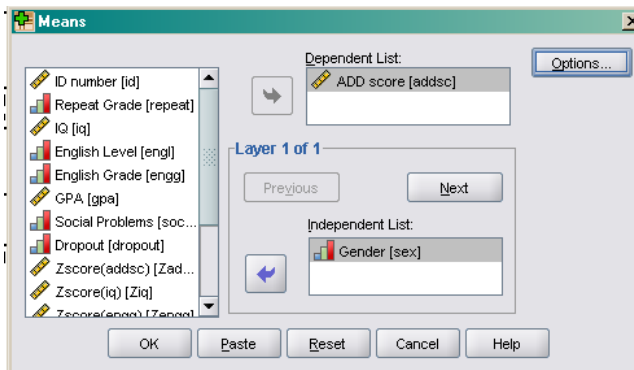
✔ Select ADD score for the **Dependent List** and Gender for the **Independent List**. Click **Options**. Notice that mean, standard deviation and number of cases are already selected under statistics. Add any other descriptive you are interested in, then click **Continue** and then **Ok**. The output follows.
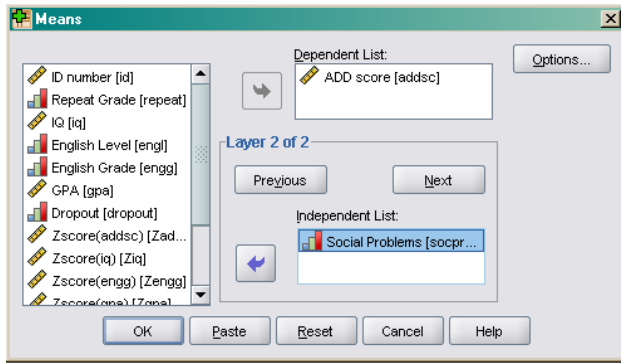
Report

ADD score

Gender	Mean	N	Std. Deviation
Male	54.291	55	12.9023
Female	49.788	33	11.2048
Total	52.602	88	12.4222

Do you think males and females differed in their ADD symptoms?

Let's try another more complicated example. This time, let's calculate descriptive statistics for ADD symptoms broken down by gender and whether or not a child had social problems.

✔ Select **Analyze/Compare Means/Means**.

Report

ADD score

Gender	Social Problems	Mean	N	Std. Deviation
Male	No social problems	52.250	48	11.6135
	Social Problems	68.286	7	13.4129
	Total	54.291	55	12.9023
Female	No social problems	48.900	30	11.1830
	Social Problems	58.667	3	8.0829
	Total	49.788	33	11.2048
Total	No social problems	50.962	78	11.4941
	Social Problems	65.400	10	12.4918
	Total	52.602	88	12.4222

Notice that this table gives you the marginal descriptives (i.e., the descriptive for gender independent of social problems and vice versa) under totals and the cell descriptives (i.e., the descriptives at each level of the variables-e.g., for boys with social problems).

✔ **Exit** SPSS. There is no need to save the Data File since we haven't changed it. It is up to you to decide whether or not you would like to save the output file for future reference.

We've reviewed a variety of options for calculating descriptive statistics depending on the type of data and the kinds of questions. We've also seen that many of the graphs we reviewed in Chapter 3 are options in the subcommands under Descriptive Statistics. In the following chapters you will discover that descriptive statistics are an option embedded within many other analyses dialog boxes (e.g. t-test, ANOVA, etc). Try the following exercises to be sure you understand all of the various options for calculating descriptives and to help you identify your own preferences.

Exercise

1. Using *merge1.sav* calculate the mean, median, mode, range, variance, and standard deviation for the following variables: self-esteem, anxiety, coping, and health. Create a histogram for anxiety. Note how you did each.

2. Using the data in *appendixd.sav*, calculate the frequency and percent of females and males who did and did not have social problems.

3. Using the data in *appendixd.sav*, calculate the mean, variance, and standard deviation for GPA broken down by social problems and drop out status.