1

DESCRIPTIVE STATISTICS

What follows is a quick reminder of descriptive statistics. Quiz yourself on descriptive statistics at the end of this chapter! You can run all the examples in this book using a data file posted at the website: **study.sagepub.com/winter**

HOW, WHEN, AND WHY DO WE USE DESCRIPTIVE STATISTICS?

First, let's make sure you understand the concept of *descriptive statistics*. As you can probably figure out based on the name, descriptive statistics describe the data. There are three essential characteristics of descriptive statistics we need to discuss: scales of measurement, measures of central tendency, and measures of variability (or spread).

1. Scales of measurement are different ways to classify the measurement of a study variable. There are four categories we will consider: nominal scales, ordinal scales, interval scales, and ratio scales.

   Nominal scales are based on assigning items to categories. For example, you can have yes versus no categories, or male versus female categories, or Honda versus Toyota versus Subaru versus Ford. For nominal scales, there is no better or worse, higher or lower—just different categories. (After all, which is higher: males or females?)

   Ordinal scales have more order to them. That is, they are ranked. Thus there might be a better or worse ranking here (pizza is ranked highest, salad second highest, sandwiches third highest, haggis lowest for food

preference). We might know the order, but we may not know how spread out those preferences are. That is, maybe pizza, salad, and sandwiches are all ranked very high in preference but haggis is ranked really, really low! Or think about a marathon race. The first, second, and third place finishers may come in a few seconds apart while the fourth place finisher is more than a minute behind.

Interval scales have order and equal intervals between items. That is, we can look at a scale ("Rank this food on a scale from 1 to 9, with 1 being not at all preferred and 9 being highly preferred"). Here, we have order: 9 is high while 1 is low. We also know that the difference between 1 and 2 is the same as the difference between 8 and 9. There are set distances between all numbers that are always equal.

Ratio scales have order, equal intervals, and a zero point. A zero indicates a total lack of the property. For example, you can get 0 correct on a 10-question test. Correctness can thus range from 0 to 10, but it cannot go below 0. It still has order (5 correct is higher than 3 correct) and intervals (the difference between 1 and 3 correct is the same as the difference between 6 and 8 correct). Time is another good ratio—there can be a zero time, but you cannot go below "no time."

2. Measures of central tendency refer to a single score that shows the central number in a set of numbers. I bet you already know these! There are three kinds of central tendency measurements: mode, median, and mean.

   The **mode** refers to the most frequently occurring number. In the set of numbers 12, 15, 15, 17, 18, 20, the mode is 15. In some cases, we may have multiple modes. For the set of numbers 12, 15, 15, 17, 17, 18, 20, we have a bimodal set (both 15 and 17 are modes). We use modes mostly for nominal scales (if participants are male, male, male, female, female, male, then our mode is male—this gender occurs 4 out of 6 times).

   The **median** refers to the middle number. In the set 12, 15, 15, 17, 17, 18, 20, our median is 17 (12, 15, and 15 come before it, while 17, 18, and 20 come after it). In the set 12, 15, 15, 17, 18, 20, there are two central numbers: 15 and 17. We simply add 15 + 17, and then divide by 2. 15 + 17 = 32 / 2 = 16. We can use the median for ordinal, interval, and ratio scales (since all are based on order).

   The **mean** is the average number. In the set 12, 15, 15, 17, 17, 18, 20, the mean is 16.29. That is, 12 + 15 + 15 + 17 + 17 + 18 + 20 = 114 / 7 =

16.29. We use the mean only for interval and ratio data since we need scale items to have equal intervals.

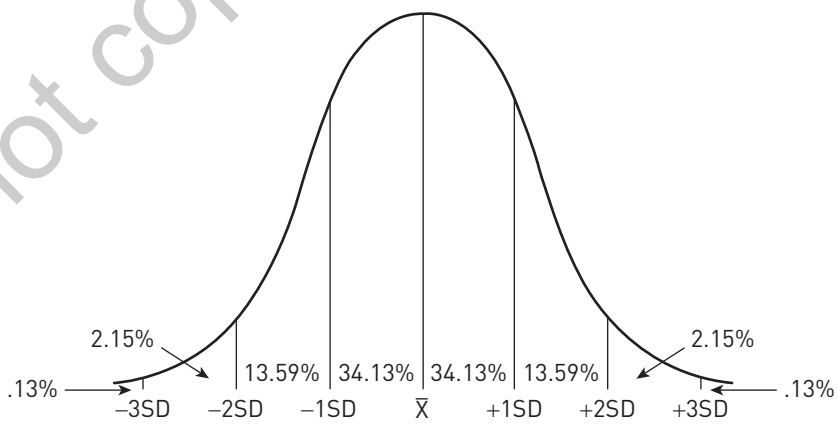
3. Measures of variability (or measure of spread) tell us how numbers spread out around the mean. That is, how much do they vary? There are three kinds of variability I want to discuss here: range, variance, and standard deviation.

   The **range** is the simplest measure of variability. You simply subtract the lowest score from the highest score. In the set 12, 15, 15, 17, 17, 18, 20, the range is 20 minus 12, or 8. Unfortunately, it doesn't tell us much more than the highest and lowest number, and thus it is susceptible to outlier issues. In the set 12, 15, 15, 17, 17, 18, 15,000, the range is 14,988, but that last number is huge! It is somewhat misleading to keep this large number in the data set as most numbers are within 12 to 18.

   Variance tells us about the variation in the data set. The larger the number, the more variability. Finding variance is the first step in looking at the standard deviation.

   The **standard deviation** is the square root of the variance. This is the statistic you will use frequently when discussing your descriptive statistics. It relates to the normal curve, which you can see in the following figure.

FIGURE 1.1



Let's say we have a normal curve that represents 100% of all cases in our sample. Fifty percent fall at or below the mean, while 50% are at or above the mean. The standard deviation (SD) breaks this down even further. That is, 34.13% of all cases fall between the mean and one standard deviation above the mean. In addition, 34.13% of all cases fall between the mean and one standard deviation below the mean. Thus 68.26% of all cases fall between minus one and plus one standard deviations of the mean. For two standard deviations, another 13.59% of the sample is counted. That is, if we look at two standard deviations above and below the mean, this accounts for 95.44% of the sample (13.59 + 34.13 + 34.13 + 13.59 = 95.44). Three SDs adds another 2.15 on each end, and so on.

What the standard deviation tells us is how many participants fall within that percentage. Think about age of participants. Imagine we have a sample ranging from 10 to 36 years old. It's a big range, right (36 – 10 = 26)? Further imagine that our mean (average) age is 24. If we know the standard deviation, we have even more information about the sample. Let's say our standard deviation is 3 years (which we get by calculating the variance from a set of participant ages and taking the square root). If one standard deviation is 3 years, and our mean age is 24, then we know that 34.13% of our sample ranges in age from 24 to 27 (the mean + the standard deviation, or 24 + 3 = 27). We also know that 34.13% of our sample ranges from 21 to 24 (the mean minus the standard deviation, or 24 – 3 = 21). Thus we know that 68.26% of our sample ranges in age from 21 to 27 (mean of 24 +/– 3). As you can see, a standard deviation of 3 shows that a lot of our sample clusters around that young adult range. Only 31.74% falls above 27 years old or below 21 years old. Now let's say we have a standard deviation of 10 with a mean of 24. Here, 68.26% of our sample ranges in age from 14 to 34. This is a much larger spread, right? 31.74% percent of our sample is older than 34 or younger than 14. As researchers, we prefer to have less spread, as it shows less variance overall, i.e. give us more precise statistical estimates.

An Example: Social Influence and College Textbooks

A robust finding in social psychology is that when people have insufficient information about how to behave or what decision to make, they rely on others as a source of information. Imagine we ask college student participants how much money they spent on books last semester. We have them fill out their name and the amount of money they recall spending at the bottom of a book survey after the names and amounts listed by 10 prior participants. Unknown to our participants, we alter

the amounts recalled by the 10 prior participants, creating one condition with a high average book price ($150) and one condition with a low average book price ($75). We predict that those who see high dollar amounts from prior participants will use that information as an anchor point and recall spending a similar high amount themselves. Those in our low dollar amount condition will anchor to that lower amount. We collect signatures and dollar amounts from 10 real participants randomly assigned to the High Dollar (HD) condition and 10 real participants randomly assigned to the Low Dollar (LD) condition. Consider the data:

High Dollar Condition ($)	Low Dollar Condition ($)
134	77
156	65
134	87
132	100
164	88
127	68
135	86
134	73
133	69
151	87
$\Sigma HD = 1{,}400$	$\Sigma LD = 800$
Mean = 140	*Mean* = 80
Median = 134	*Median* = 81.5
Mode = 134	*Mode* = 87
Standard Deviation = 12.33	*Standard Deviation* = 11.28

<HD and LD should be subscripts, per text below>

Σ, or the symbol for Sigma, means "the sum of." Thus Σ_{HD} is the sum of the scores for the High Dollar Condition. That is, 134 + 156 + 134 + 132 + 164 + 127 + 135 + 134 + 133 + 151 = 1,400. There are 10 scores here, so we divide 1,400 by 10 to get 140, our mean. The median is the middle number (in this case 134 + 134 / 2 = 134). The mode is the most frequently occurring number (in this case 134 appears three times). The standard deviation is 12.33, which means that 34.13% of all scores fall between the mean (140) and 1 standard deviation above

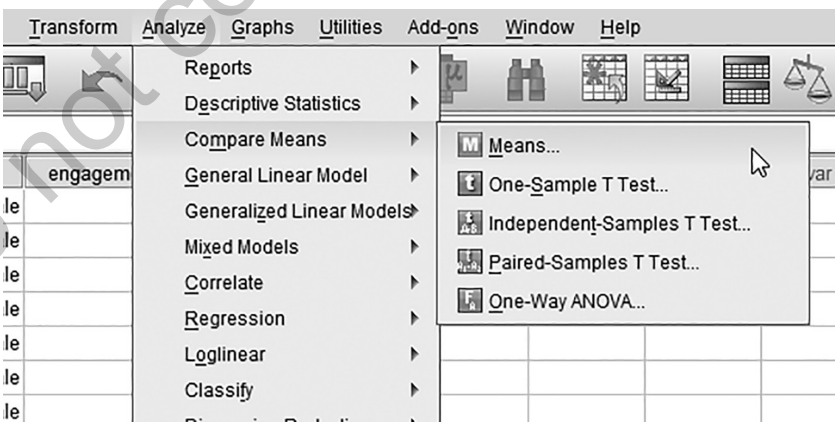
the mean (140 + 12.33 = 152.33), while another 34.13% falls between the mean (140) and 1 standard deviation below the mean (140 – 12.33 = 127.67). Thus 68.26% of the sample in the High Dollar Condition recalled spending between $128 and $152, which is close to our high dollar anchor amount of $150.

We do the same thing for the Low Dollar Condition, giving us a mean of 80 (800 / 10 = 80), a median of 81.5 (77 + 86 / 2 = 81.5), and a mode of 87. Our standard deviation is 11.28, with most Low Dollar Condition participants (68.26%) recalling spending between $68.72 and $91.28 on their books (once again, pretty close to our low anchor). Keep in mind that the variance here is the standard deviation squared, or 11.28 × 11.28 = 127.23. Since variance is not on the same scale as the standard deviation, it is less useful in psychology studies.

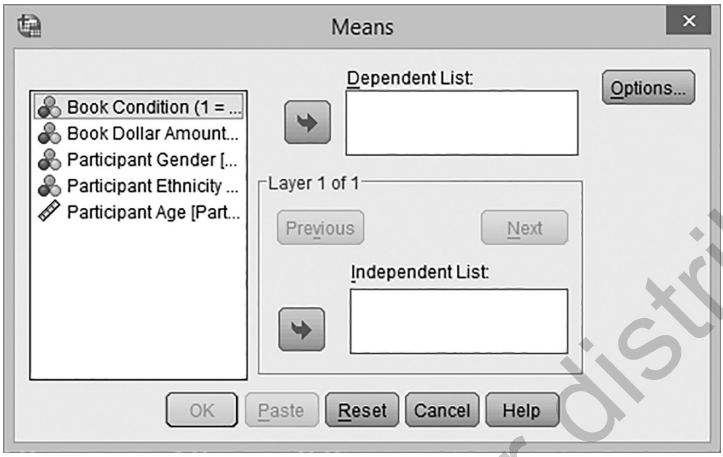
Let's run this in SPSS. There are several different ways to get descriptives, so we'll look at three different SPSS procedures: means, descriptives, and frequencies. **(Note: If you are writing a participant section, I highly recommend frequencies.)**

1. SPSS: OUR BOOK SURVEY STUDY (USING THE MEANS PROCEDURE)

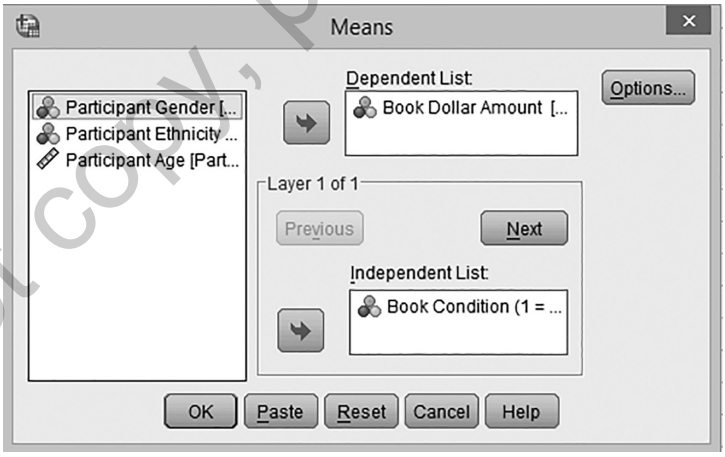
1. The first method I want to show you has the advantage of looking at means and standard deviations within different levels of the independent variable. That is, maybe I want to see the mean and SD for the High Dollar Condition compared to the Low Dollar Condition. To run this test, you should use the Means procedure. First, click Analyze > Compare Means > Means . . . on the top menu as shown in the following screenshot.



You will be presented with the following:

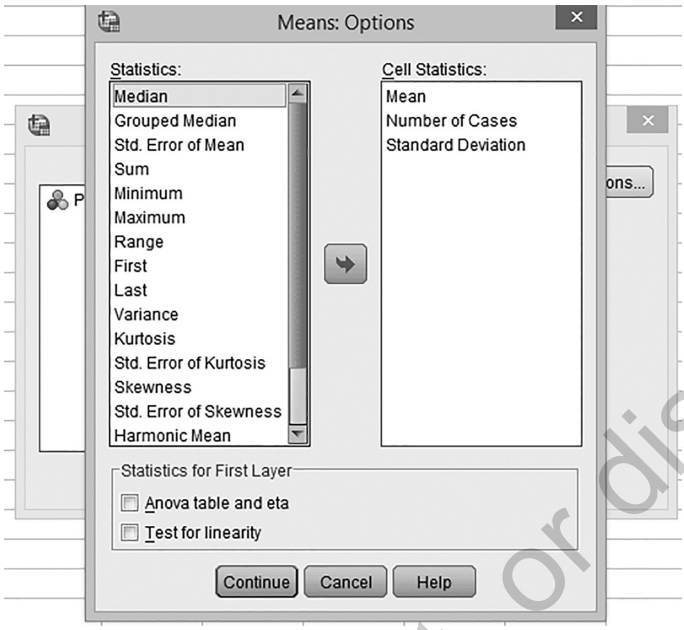


2. Put the **Book Condition (1 High, 2 Low)** variable into the Independent List box and the **Book Dollar Amount (in $)** variable into the Dependent List box by highlighting the relevant variables and pressing the buttons.



3. Next, click the Options button.

You will be presented with the following screen:



4. As you can see, the Mean, Number of Cases, and Standard Deviation are preselected for you. You can get other statistics if you want from the left column, but we will stick with those three items only. Select Continue and then OK.

OUTPUT OF THE MEANS PROCEDURE IN SPSS

You will see two tables containing all the data generated by the means procedure in SPSS.

Case Processing Summary

This table provides the *N* for us. It isn't all that important.

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Book Dollar Amount * Book Condition (1 = High, 2 - Low)	20	100.0%	0	0.0%	20	100.0%

Report

The report, however, gives us our mean and standard deviation for each group (high and low) as well as the total mean and standard deviation across condition. (Please note this Total row, as we will discuss it in our next few procedures).

Report

Book Dollar Amount

Book Condition (1 = High, 2 - Low)	Mean	N	Std. Deviation
High	140.00	10	12.329
Low	80.00	10	11.284
Total	110.00	20	32.859

REPORTING THE OUTPUT FOR THE MEANS PROCEDURE

We report the statistics in this format:

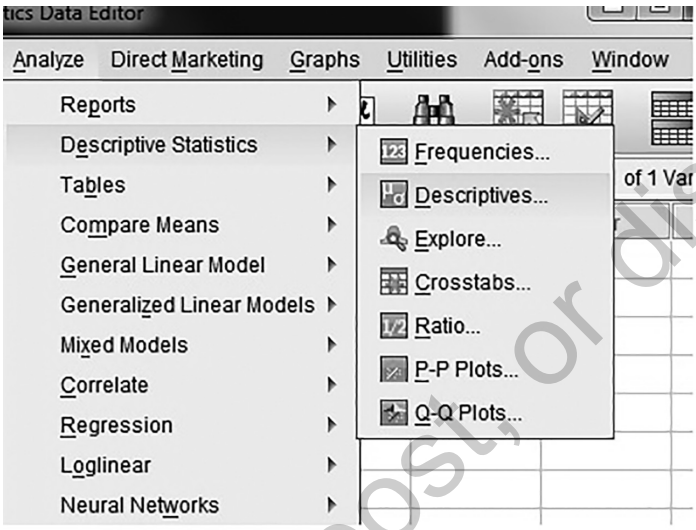
> Students in the High Dollar Condition recalled spending an average of $140 (*SD* = 12.33) on books. Students in the Low Dollar Condition recalled spending an average of $80 (*SD* = 11.28).

That's our first way of getting means. Not too hard, right? When getting the mean and standard deviation for interval- or ratio-dependent variables like dollar amount, time, height, and weight, the Means procedure is useful. Of course, there are other ways of getting descriptive statistics that are a part of other statistical tests. For example, when running a *t*-test or an ANOVA, descriptives may be automatically given to you in the output, or you can click on the Options button in those SPSS menus to get means, standard deviations, and ranges.

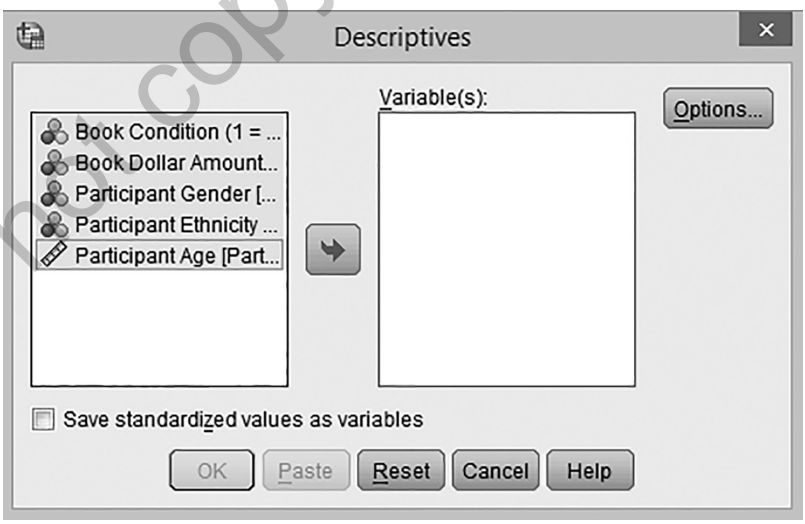
However, let's say you don't want to look at dollar amounts recalled across different conditions. You can still use the Means procedure, but this time leave the Independent List box empty. You can then see the total mean and standard deviation across all participants (not separated by High or Low Dollar Condition). For example, the mean across all participants in the book survey study is $110 (*SD* = 32.86). An alternative way to look at descriptives is by using the Descriptives procedure in SPSS. It will similarly give you descriptive information for the entire sample (not broken up by condition)

2. SPSS: OUR BOOK STUDY (USING THE DESCRIPTIVES PROCEDURE)

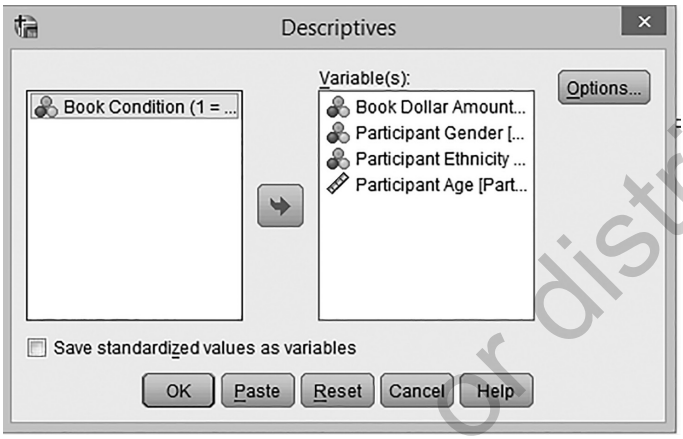
1. Click Analyze > Descriptive Statistics > Descriptives . . . on the top menu as shown in the following screenshot.



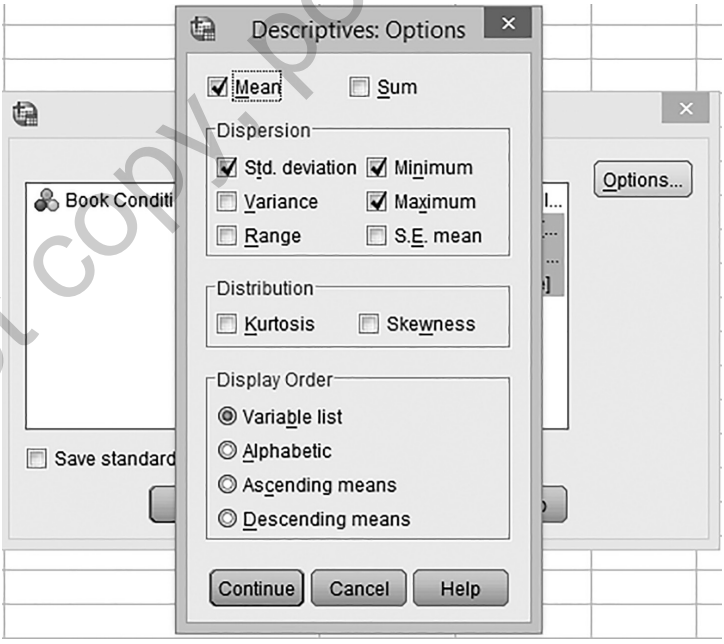
You will be presented with the following (notice there is no longer an Independent List box):



We will move variables that we want descriptive statistics for over to the Variable(s) box. In this case, we can look at several dependent variables (Book Dollar Amount, Participant Gender, Participant Ethnicity, and Participant age). Let's move all of our dependent variables over . . .



Now, click on Options. You will see several potential choices. Click the ones I chose.



Click Continue, and then OK.

OUTPUT OF THE DESCRIPTIVES PROCEDURE IN SPSS

When using Descriptives, you will see only one box.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Book Dollar Amount	20	65	164	110.00	32.859
Participant Gender	20	1.00	2.00	1.5500	.51042
Participant Ethnicity	20	1.00	4.00	2.0500	1.05006
Participant Age	20	18.00	24.00	20.3500	2.03328
Valid N (listwise)	20				

REPORTING THE OUTPUT FOR THE DESCRIPTIVES PROCEDURE

See the Book Dollar Amount row? The mean (110) and standard deviation (32.86) here are identical to the Total from the means report we saw in our prior test (the Means procedure). We don't get the information separately for the High and Low Dollar Conditions. This is a limitation of the Descriptives procedure in SPSS when you want to look within different levels of your independent variable.

But consider our next three rows related to participant gender, ethnicity, and age. We often report descriptive statistics for demographic information, but I hope you see the problem with looking at gender and ethnicity here. Before we look at that problem, let's first look at age. As you can see, our sample ranged in age from 18 (minimum) to 24 (maximum). The mean age was 20.35 with a standard deviation of 2.03. In a results section, I would write this as follows:

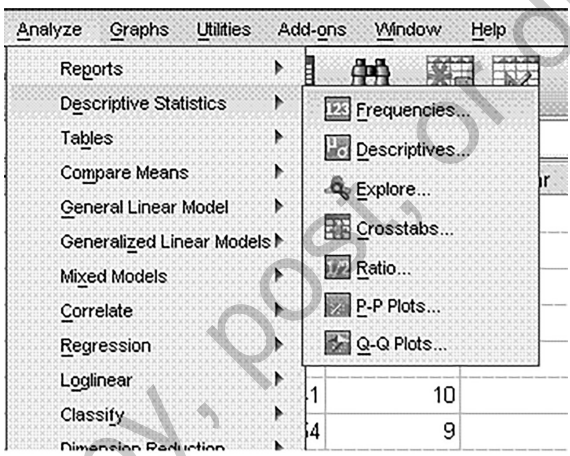
The sample ranged in age from 18 to 24 ($M = 20.35$, $SD = 2.03$).

Gender and ethnicity, however, are a little odd here and are actually inappropriate to interpret within the Descriptives procedure. Consider gender. The minimum is 1 and the maximum is 2. In my data set, I assigned males to be 1 and females to be 2, so obviously it ranges from 1 to 2, but I could easily have assigned females to be 1 and males to be 2. With nominal scales, we aren't really interested in the numbers, just the category. Also note that I have a mean of 1.55 for gender, but does that make any sense? Is there any such thing as an average gender? Not really, right? Gender is a nominal variable (a category). The same thing applies to ethnicity. Our average ethnicity here is 2.05, but since I have five categories (Caucasian, African American, Hispanic/Latino, Asian, and Other), an average

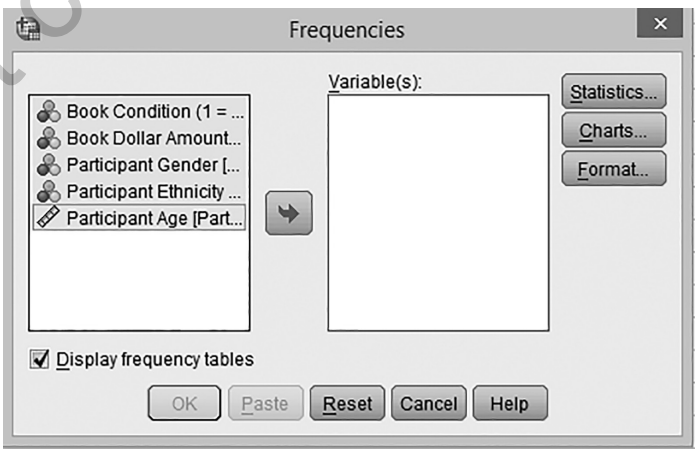
ethnicity of 2.05 is meaningless. We cannot use the mean for gender or ethnicity. Instead, we rely on the mode, which tells us how many data points fall into each category—that is, how many men are there and how many women are there, and how many of each ethnicity are represented? To get the mode, we need to run a different type of SPSS test: Frequencies.

3. SPSS: OUR BOOK STUDY (USING THE FREQUENCIES PROCEDURE)

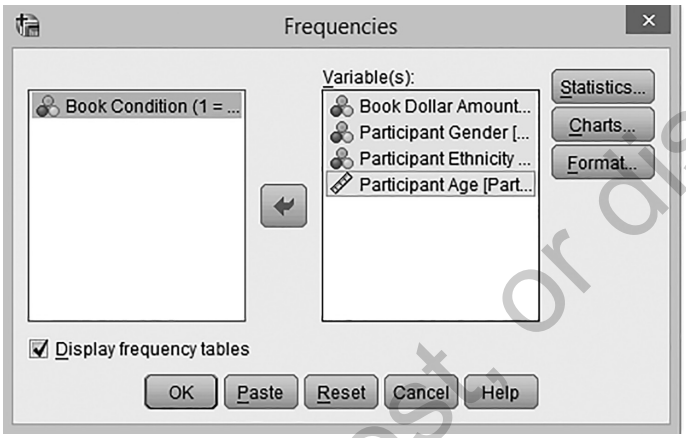
1. Click Analyze > Descriptive Statistics > Frequencies . . . on the top menu as shown in the following screenshot.



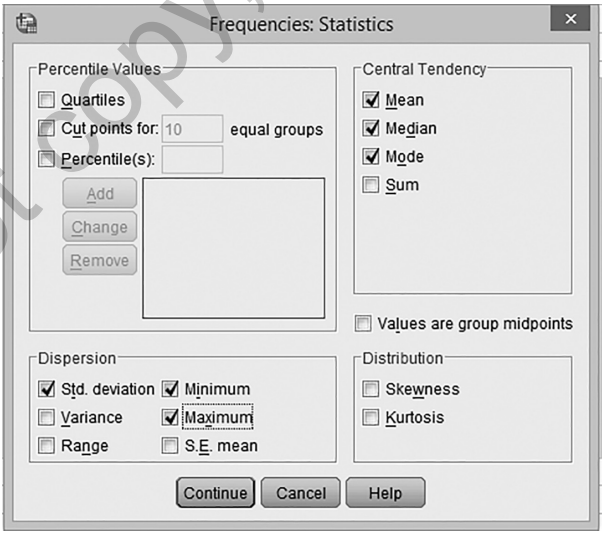
You will be presented with the following:



As with the Descriptives procedure, move the variables that we want descriptive statistics for over to the Variable(s) box. In this case, I prefer to look at my dependent variables (Book Dollar Amount, Participant Gender, Participant Ethnicity, Participant Age). Again, you should notice that there is no place to list your independent variable in this Frequencies procedure. Thus you will not be able to get means and standard deviations for each of your separate conditions (High versus Low). For now, move all of the dependent variables to the variables box.



Now, click on Statistics and select the items I chose. You see that you have several options here, including the mean, median, mode, range, and standard deviation. You can even look at the minimum and maximum.



Now, click Continue and then OK to get to the output.

OUTPUT OF THE FREQUENCIES PROCEDURE IN SPSS

You will see several tables in your output, including one overall Statistics table followed by Frequency tables for each of the variables you added to the Variable(s) box.

Statistics

Statistics

		Book Dollar Amount	Participant Gender	Participant Ethnicity	Participant Age
N	Valid	20	20	20	20
	Missing	0	0	0	0
Mean		110.00	1.5500	2.0500	20.3500
Median		113.50	2.0000	2.0000	20.0000
Mode		134	2.00	1.00	18.00
Std. Deviation		32.859	.51042	1.05006	2.03328
Minimum		65	1.00	1.00	18.00
Maximum		164	2.00	4.00	24.00

Here you can see descriptive statistics for Book Dollar Amount, Participant Gender, Participant Ethnicity, and Participant Age. Once again, we see the mean Total information for Book Dollar Amount ($M = 110$, $SD = 32.86$), but it doesn't break it down by condition (High versus Low). We also see our descriptive information for age ($M = 20.35$, $SD = 2.03$), which duplicates the test we ran in the prior section. Finally, we have means and standard deviations (and other information) for gender and ethnicity. At this point I hope you're thinking, "Means and standard deviations for gender and ethnicity make no sense!" You're right—they don't. The only helpful columns in this Statistics table are Book Dollar Amount and Participant Age. So let's look at some other tables to get descriptive information about gender and ethnicity.

Frequency Table

Participant Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	9	45.0	45.0	45.0
	Female	11	55.0	55.0	100.0
	Total	20	100.0	100.0	

As the Participant Gender table shows, it appears to be a fairly even split here; there are 9 males and 11 females. Thus 45% are male and 55% are female. Between the statistics table and the frequency table, we can now write out our participant data. Now, consider ethnicity.

Participant Ethnicity

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Caucasian	8	40.0	40.0	40.0
	African American	5	25.0	25.0	65.0
	Hispanic / Latino	5	25.0	25.0	90.0
	Asian	2	10.0	10.0	100.0
	Total	20	100.0	100.0	

We have eight Caucasians (40%), five African Americans (25%), five Hispanics/Latinos (25%), and two Asians (10%). In this table, there are zero Others, so that row is omitted. We now have enough information to write up our participant section in a journal-style methods section.

REPORTING THE OUTPUT FOR THE DESCRIPTIVES PROCEDURE

There were 20 participants in the study, including 9 men and 11 women, who ranged in age from 18 to 24 ($M = 20.35$, $SD = 2.03$). There were eight Caucasians (40%), five African Americans (25%), five Hispanics/Latinos (25%), and two Asians (10%).

Of course, you will see frequency tables for Book Dollar Amount and Participant Age as well. They look like this:

Book Dollar Amount

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	65	1	5.0	5.0	5.0
	68	1	5.0	5.0	10.0
	69	1	5.0	5.0	15.0
	73	1	5.0	5.0	20.0
	77	1	5.0	5.0	25.0
	86	1	5.0	5.0	30.0
	87	2	10.0	10.0	40.0
	88	1	5.0	5.0	45.0
	100	1	5.0	5.0	50.0
	127	1	5.0	5.0	55.0
	132	1	5.0	5.0	60.0
	133	1	5.0	5.0	65.0
	134	3	15.0	15.0	80.0
	135	1	5.0	5.0	85.0
	151	1	5.0	5.0	90.0
	156	1	5.0	5.0	95.0
	164	1	5.0	5.0	100.0
	Total	20	100.0	100.0	

Participant Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18.00	5	25.0	25.0	25.0
	19.00	3	15.0	15.0	40.0
	20.00	3	15.0	15.0	55.0
	21.00	4	20.0	20.0	75.0
	22.00	1	5.0	5.0	80.0
	23.00	2	10.0	10.0	90.0
	24.00	2	10.0	10.0	100.0
	Total	20	100.0	100.0	

As you probably know, though, these tables don't provide as much help for interval or ratio data compared to tables that give us means and standard deviations.

Conclusion

Frequencies and Descriptives are the building blocks for future statistical tests. Make sure you can distinguish between nominal, ordinal, interval, and ratio scales as well as different measures of central tendency (mean, median, and mode).

To run tests yourself, download the Crash Course in Descriptive Statistics SPSS file at **study.sagepub.com/winter**.

Crash Course in Descriptive Statistics: Quiz Yourself!

(Answers in back of book)

Instructions: Imagine you ran frequencies to write out your participant data for a study. You inputted data into SPSS, which gave you the following output. Using these charts, answer the questions on the final page.

Statistics

		Participant Gender	Participant Ethnicity	Participant Age
N	Valid	144	144	143
	Missing	0	0	1
Mean		1.5069	2.0972	20.9930
Median		2.0000	2.0000	21.0000
Mode		2.00	2.00	22.00
Std. Deviation		.50170	1.01264	2.81719
Minimum		1.00	1.00	12.00
Maximum		2.00	6.00	29.00

(Continued)

(Continued)

Frequency Table

Participant Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	71	49.3	49.3	49.3
	Female	73	50.7	50.7	100.0
	Total	144	100.0	100.0	

Participant Ethnicity

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Caucasian	22	15.3	15.3	15.3
	Hispanic	111	77.1	77.1	92.4
	African American	3	2.1	2.1	94.4
	Asian American	2	1.4	1.4	95.8
	Other	6	4.2	4.2	100.0
	Total	144	100.0	100.0	

Participant Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	12.00	3	2.1	2.1	2.1
	14.00	1	.7	.7	2.8
	15.00	3	2.1	2.1	4.9
	16.00	4	2.8	2.8	7.7
	17.00	1	.7	.7	8.4
	18.00	9	6.3	6.3	14.7
	19.00	13	9.0	9.1	23.8
	20.00	20	13.9	14.0	37.8
	21.00	22	15.3	15.4	53.1
	22.00	26	18.1	18.2	71.3
	23.00	19	13.2	13.3	84.6
	24.00	13	9.0	9.1	93.7
	25.00	4	2.8	2.8	96.5
	26.00	3	2.1	2.1	98.6
	27.00	1	.7	.7	99.3
	29.00	1	.7	.7	100.0
	Total	143	99.3	100.0	
Missing	System	1	.7		
Total		144	100.0		

1. Choose the correct range, mean, and standard deviation for participant age written in correct APA format.

   A. Participants ranged in age from 11 to 28 ($M = 20.99$, $SD = 2.82$).

   B. Participants ranged in age from 12 to 30 ($M = 20.99$, $SD = 2.82$).

C. Participants ranged in age from 12 to 29 ($M = 20.99$, $SD = 2.82$).

D. Participants ranged in age from 12 to 29 ($M = 2.82$, $SD = 20.99$).

E. Participants ranged in age from 12 to 29 ($M = 20.99$, $SD = 22$).

2. Choose the correct frequency information for gender.

A. There were 71 men (39%) and 73 women (61%).

B. There were 71 men (49%) and 73 women (51%).

C. There were 73 men (50%) and 71 women (50%).

D. There were 71 men (51%) and 73 women (49%).

E. There were 73 men (49%) and 71 women (51%).

3. Which of the following is the correct frequency information for ethnicity?

A. In this sample, 21 participants were Caucasian (15%), 112 were Hispanic (77%), 3 were African American (2%), 3 were Asian (1%), and 6 did not provide their ethnicity (4%).

B. In this sample, 22 participants were Caucasian (19%), 111 were Hispanic (67%), 3 were African American (4%), 2 were Asian (1%), and 3 did not provide their ethnicity (4%).

C. In this sample, 20 participants were Caucasian (76%), 11 were Hispanic (21%), 3 were African American (2%), 6 were Asian (1%), and 6 did not provide their ethnicity (4%).

D. In this sample, 22 participants were Caucasian (15%), 111 were Hispanic (77%), 3 were African American (2%), 2 were Asian (1%), and 6 did not provide their ethnicity (4%).

E. In this sample, 15 participants were Caucasian (15%), 117 were Hispanic (77%), 3 were African American (2%), 2 were Asian (1%), and 6 did not provide their ethnicity (4%).

4. Which of the following scales was used for gender (nominal, ordinal, interval, or ratio)?

A. Gender is a nominal variable.

B. Gender is an ordinal variable.

C. Gender is an interval variable.

D. Gender is a ratio variable.

E. None of the above.

(Continued)

(Continued)

5. For which (if any) of the three dependent variables in this data set (gender, age, ethnicity) would you report the standard deviation?

   A. Age, since it is measured on an interval scale

   B. Gender, since it is measured on an interval scale

   C. Ethnicity, since it is measured on an interval scale

   D. All of the above

   E. None of the above, since the standard deviation is inappropriate to use for all three variables