CHAPTER 4

Making Sense of Data: Measures of Central Tendency and Variability

After reading this chapter, you will be able to

- Calculate the three measures of central tendency and the three measures of variability
- Interpret the information conveyed by each measure of central tendency and variability
- Report the three measures of central tendency and variability in APA style
- Identify considerations when selecting a measure of central tendency
- Describe why variability is an essential part of research
- Generate the three measures of central tendency and variability using SPSS
- Interpret SPSS output with respect to the three measures of central tendency and variability

In the previous chapter, we discussed ways to organize and help us make sense of large quantities of data. Indeed, this is the essence of descriptive statistics. In this chapter, we will continue to add to our collection of descriptive statistical tools. Whereas the previous chapter emphasized tabular and graphical displays of data, this chapter and the next one will emphasize quantitative summaries of data. In this chapter, we will learn about what are called *measures of central tendency* and *measures of variability*. Much like an artist needs paint to paint a picture, a researcher needs measures of central tendency and variability to describe a dataset concisely.

In this chapter, we will begin by introducing the three measures of central tendency, how to calculate each one, and how they are reported in research studies. Next, we will discuss four considerations in choosing a measure of central tendency to describe a dataset. We will then discuss how to use the software program Statistical Package for the Social Sciences (SPSS; SPSS Corporation, Chicago, IL) to generate these measures and interpret the output that SPSS gives us. After completing our discussion of the measures of central tendency, we will turn to three measures of variability. We will begin this discussion by explaining what is meant by "variability" and why it is an essential part of any research study. After learning how to calculate and interpret each measure of variability, we will use SPSS to generate them and interpret the corresponding output.

When learning about frequency tables and graphs in the previous chapter, we used a research study that Laura Wendt (2013) conducted. We will continue to use her study and data throughout most of this chapter. When we learn how to use SPSS, we will introduce a new study and use its data.

MEASURES OF CENTRAL TENDENCY

Three Measures of Central Tendency

As you saw in the previous chapter, research studies tend to contain a lot of data, and we need to make sense of it all somehow. You also know from the previous chapter that we could make a frequency distribution table or a frequency distribution graph of these data, and in doing so, it would help us understand what these data "look like" so that we can describe them quickly and understandably to other people. Now we are going to add to our collection of tools to describe data. In the first half of this chapter, we will examine measures of central tendency. In the second half, we will examine measures of variability.

I bet that when you first saw Table 4.1, which again contains the burnout score for each participant in Wendt's (2013) research, you wanted to throw your hands up in the air and walk away. At first glance, this table is a mess of random numbers between 15 and 75, the response range used on Wendt's measure of academic burnout. However, we can assign one number to describe all 108 datapoints. Of course, any one number used to describe so much data will have its limitations, but let's examine three ways we can assign one number to these data.

The purpose of a measure of central tendency is to provide one number that describes a large set of data. There are three measures of central tendency: mean, median, and mode. Let's discuss each one, including an examination of how they are calculated, by using a much smaller set of data than what Wendt (2013) used.

Mean

The **mean** is the arithmetic average of a set of numbers. It is the most commonly reported measure of central tendency in published research. To calculate a mean, we add together all of the scores in the dataset; then we divide by the number of scores in the dataset. Let's use the following (small) dataset to calculate its mean:

6, 5, 4, 6, 7, 8, 3, 5, 9, 2

To calculate the mean, sum these numbers (to get 55) and divide the sum by the number of scores (10). Formulaically, we have:

$$\frac{\Sigma x}{N}$$

where

Σ means "summation"

x is an individual score

N is the total number of scores in the dataset

The mean is thus:

$$\frac{6+5+4+6+7+8+3+5+9+2}{10}$$

$$\frac{55}{10} = 5.50$$

Table 4.1 Burnout Score for Wendt's (2013) 108 Participants

36	37	51	42
38	40	41	40
48	36	44	34
30	30	36	28
31	32	34	37
16	42	50	39
34	40	25	28
25	28	41	28
42	35	43	52
35	35	31	39
37	49	47	46
45	38	34	50
27	34	55	40
54	20	53	36
20	28	38	23
32	16	37	32
34	22	43	24
35	25	41	47
33	19	46	25
33	32	53	45
31	42	38	40
49	49	31	45
20	39	40	37
31	31	35	42
28	40	45	28
31	49	45	38
35	31	39	28

Mean: arithmetic average of a dataset.

Recall from Chapter 1 the difference between parameters and statistics. Parameters are numbers based on the population characteristics; statistics are numbers derived from a sample that was drawn from a population. We are currently discussing statistics, but starting in Chapter 6, we need to incorporate certain parameters into our discussion. In this case, we need to distinguish some symbols you will see in psychological research (which almost always involves statistics) and how they correspond to certain parameters.

In psychological research, you will see a mean reported as an italicized capital *M*. Sometimes the mean will be reported as $\overline{x}$. These are sample statistics. However, when we begin to incorporate population parameters into our discussions, the symbol for the population mean is the Greek letter μ (pronounced *mew*). We will reiterate this distinction when the time comes, but as you will soon be seeing symbols for the mean in psychological research, I wanted to make sure you know what you are reading.

Median

The **median** is the "middle" number in a dataset. It is the number that divides a dataset in half so that 50% of the scores are greater than that number and 50% of the scores are less than that number. The median is often reported in research when the mean does not do a good job of describing a dataset. This occurs when the scores in the dataset tend to be toward the high end or toward the low end of the distribution of scores. In such cases, a few "extreme" scores would distort the mean because the mean uses all data points in its calculation. We will discuss a real-life example of such a situation in the next section of this chapter.

Median: middle score in a dataset that divides the dataset in half so that an equal number of scores fall above and below that score.

For now, to calculate a median, let's return to the small dataset we used to calculate the mean:

6, 5, 4, 6, 7, 8, 3, 5, 9, 2

To determine the median, we need to arrange the scores in a dataset from lowest to highest. We then find the score that divides the distribution of scores in half. Here are the scores arranged in order from lowest to highest:

2, 3, 4, 5, 5, 6, 6, 7, 8, 9

We now need to find the score that divides the dataset in half so that 50% of the scores are greater than that number and 50% of the scores are less than that number. With an odd number of scores (and a small dataset), this is easy to do. However, when we have an even number of scores as we do here, there is not a readily available score that divides the dataset in half. In this case, we take the two middle scores and find the mean of these two numbers. These two numbers are highlighted here:

2, 3, 4, 5, 5, 6, 6, 7, 8, 9

The mean of 5 and 6 is 5.5. Thus, our median is 5.5.

We said that the median is a "better" measure of central tendency when the mean does not do a good job of describing its dataset. Let's add an 11th score to our dataset to illustrate this point:

2, 3, 4, 5, 5, 6, 6, 7, 8, 9, 45

With this additional score of 45, our mean becomes 100 / 11 = 9.09.

What is our median? It is now 6 (the first 6 in the distribution). Which measure of central tendency does a better job of describing the dataset? The mean of 9.09 is misleading as only one score is greater than 9.09. When there are a few extreme scores, called **outliers**, the median is preferred to the mean as a measure of central tendency because the median is less affected by these extreme scores.

Outlier: score that is extremely high or extremely low compared with most other scores in a dataset.

Mode

The **mode** is the most frequently occurring score in a dataset. If you are looking at a frequency distribution graph, the mode will be the highest point on such a graph. When we have nominal data, the mode is the only measure of central tendency that can be used. It does not make sense to calculate a mean for a variable such as respondent's sex or religious affiliation. We don't see the mode reported often in research studies. There are two reasons why this is the case. First, the mean and median tend to do a better job of describing large sets of data because they both use all scores in their calculations. Second, although the mode is used with nominal data, as we discussed in Chapter 2, researchers generally prefer data quantified with a scale measurement because there are more statistical tools that can be used with such data, which we will see later on in this class.

Mode: most frequently occurring score in a dataset.

In our dataset, what is the mode? Realize that this is a bit of a trick question:

6, 5, 4, 6, 7, 8, 3, 5, 9, 2

The scores of 5 and 6 both occur most frequently (each occurs two times). Thus, the mode is 5 and 6. In this case, we have a *bimodal* distribution, meaning that there are two modes. A dataset with one mode is called a *unimodal* distribution, and a dataset with three or more modes is called a *multimodal* distribution.

Be aware that the mode is the score(s) that occurs most frequently, not the frequency of that score(s). If you said the mode was 2, that is indeed how often the most frequently occurring scores of 5 and 6 occurred, but it is not the mode.

Table 4.2 contains a summary of the three measures of central tendency.

Reporting the measures of central tendency in research

When reporting measures of central tendency, researchers will do so either in the text of an article or in a table that presents such statistics. For instance, returning to Wendt's (2013) research on burnout, she could have reported scores on the burnout measure in the text using these symbols:

For the mean, it would be reported as "M = 36.50"

Table 4.2 Summary of the Three Measures of Central Tendency

	Measure of Central Tendency		
	Mean	**Median**	**Mode**
Definition	Average number	Middle number	Most frequent number(s)
Appropriate for	Scale data	Ordinal and scale data	Nominal, ordinal, and scale data
Weakness	Influenced by extreme scores		Uses only two numbers in its calculation

Table 4.3 Descriptive Statistics for Variables in Wendt's (2013) Research

Variable	*M*	*Mdn*	Mode
Participant Age	19.7 years	21.0 years	21.0 years
Burnout	36.5	36.5	28.0 and 31.0
Dysfunctional Perfectionism	32.8	33.0	33.0
Role Overload	40.4	41.0	33.0, 34.0, and 36.0

Notes. For burnout, scores could range from 15 to 75. For dysfunctional perfectionism, scores could range from 11 to 55. For role overload, scores could range from 13 to 65.

Abbreviations. *M* = mean; *Mdn* = median.

For the median, it would reported as "*Mdn* = 36.50"

For the mode, it would be reported as "mode = 28.0 and 31.0"

Notice that the symbols for mean and median are italicized. It is general practice to italicize statistical symbols. We do not have a symbol for the mode; thus, it is not italicized.

In Wendt's (2013) research, she reported her measures of central tendency in a table, which you can see in Table 4.3. Notice the *Notes* at the bottom of the table. These are important because they give you information that is needed to interpret the numbers in the table. For example, knowing the possible range of scores on each variable is important because, as was the case here, those possible ranges were different for each variable. Thus, we cannot say, at this point in the class, that participants' scores were higher on role overload than on burnout even though each measure of central tendency is greater for role overload than it is for burnout.

Let's check our understanding of the three measures of central tendency before moving on to consider issues in selecting which one is the "best" way to summarize a dataset.

LEARNING CHECK

1. A researcher surveyed 10 undergraduate psychology majors about their study behaviors. The following is a list of the number of hours they spent studying on the weekend:

6	5	3	4	9
7	3	7	8	3

Calculate the mean.

$$A : \frac{\Sigma x}{N}$$

$$\text{Mean} = \frac{55}{10}$$

$$\text{Mean} = 5.50$$

Calculate the median.
Remember to arrange the scores from lowest to highest:

3, 3, 3, 4, 5, 6, 7, 7, 8, 9

A: 5.50 (which is the mean of the middle two numbers of 5 and 6)

Calculate the mode.

A: 3

2. In the previous chapter, you saw this frequency graph, which displays the distribution of scores on a midterm exam. Use the graph in Figure 4.1 to calculate each measure of central tendency.

A: mean = 83.0

median = 85.0

mode = 85.0

3. What is an "outlier?" Why do outliers tend to affect the mean more than the median?

A: An outlier is a score that is extremely high or extremely low compared with most other scores in a dataset. The mean uses all numbers in its calculation, so an outlier will pull the mean up or drag it down. The median needs only the middle number or middle two numbers in a dataset, so an outlier will not be as likely to affect the median as it is to affect the mean.

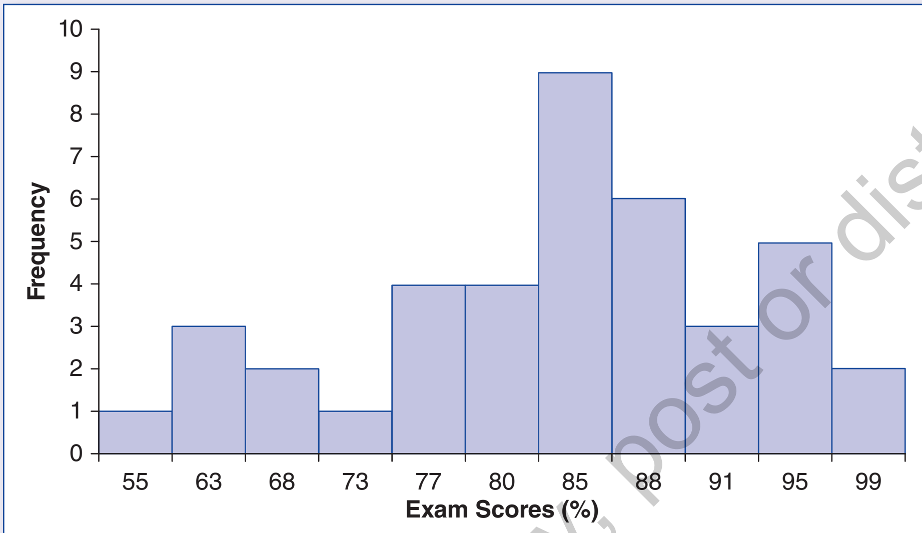
4. The value of one score in a dataset is changed from 20 to 30. Which measure(s) of central tendency is (are) certain to change?

a) the mean
b) the median
c) the mean and the median
d) the mode

A: a

(Continued)

(Continued)



5. Explain why "the mean" is the correct answer to the previous question.

   A: The mean must use each and every score in its calculation, so it has to change when one score is added to the dataset. The median does not necessarily change because the middle number (or middle two numbers) are not guaranteed to change by adding one score to the dataset. The mode does not change unless its addition makes that score the most frequently occurring score in the dataset.

6. You read a research article that reports the following information: "Scores on the sexism measure (M = 5.5, Mdn = 5.0, mode = 5.0) were higher than scores on the measure of racial prejudice (M = 4.8, Mdn = 4.5, mode = 5.0)."

Use this information to answer the following questions:

a) What is the mean of the sexism measure?

   A: 5.5

b) What is the mean of the racial prejudice measure?

   A: 4.8

c) What is the median of the sexism measure?

   A: 5.0

d) What is the median of the racial prejudice measure?

   A: 4.5

e) What is the mode of the sexism measure?

   A: 5.0

f) What is the median of the racial prejudice measure?

   A: 5.0

Choosing a Measure of Central Tendency

In this section, we will discuss four considerations in deciding which measure of central tendency is the best one to report for a given dataset. In general, researchers tend to prefer the mean as the measure of central tendency to report. There are reasons why this is the case, one of which is that it does require all scores in a dataset to be used in its calculation. Another will become evident in Chapter 5 as the mean, in combination with a measure of variability we'll learn about later in this chapter, can provide us with a great deal of information. In this section, however, we will discuss four considerations before defaulting to the mean as our measure of central tendency.

Consideration 1: Outliers in the data

The town I live in has a population of about 8,000 people, and the mean per capita (per person) income is $15,000. Now, suppose that Donald Trump moves to my town.

Although it is difficult to pinpoint precisely, let's say that Trump's income in 2014 was $380 million (Snyder, 2015). With 8,001 people now living my town, what is the mean income of my town? $62,492. Does this information indicate that the 8,000 people in my town who are not Donald Trump suddenly made $47,492 more this year? Of course not; in this example, we need to calculate the median because Trump's income is an outlier. I know it will take a lot of time to arrange 8,001 income scores from lowest to highest, so let's assume that 2,000 people in my town earn $5,000 annually, 4,000 people earn $12,000 annually, 2,000 people earn $31,000 annually, and Trump earns $380,000,000 annually. That gives us a mean of $62,492. However, the median will be $12,000, which is obviously quite a bit lower and a more valid indicator of incomes than the mean.

Unlike the mean, the median is not as influenced by a few outliers (extreme scores) in a dataset. If you really want a valid picture of incomes (in my town or throughout the United States), the median tends to be a better measure of central tendency to report than the mean because one extreme income (outlier), such as Donald Trump's, won't blur what is the case for the majority of the dataset.

Consideration 2: Skewed data distributions

You may have noticed in the previous section that in our example data, the mean, median, and mode were almost identical. Indeed, when we have a large dataset, this will normally be the case. In fact, because it is "normally" the case, we say that such distributions of data are **normally distributed**. That is, the three measures of central tendency are all approximately equal. This is important because when we have a normal distribution of data, there is a great deal of information available to us. We will reveal that information in the next chapter.

> **Normal distribution**: dataset in which the measures of central tendency are approximately equal to each other, thus creating a symmetrical, bell-shaped distribution of scores.

Figure 4.2 contains a normal distribution. The mode should be pretty easy to figure out as it is the highest point of the curve, which is where most scores cluster. As we move farther away from the highest point of the curve, we see that there are progressively fewer scores. Knowing that the measures of central tendency are approximately equal, it becomes easy to locate the mean and median.

Of course, as we just discussed in our example with the variable of income, not all distributions are normal. In the case of the distribution of income, the distribution is skewed. By "skewed," we mean that the distribution of scores is not symmetrical as it is in a normal distribution. That is, outliers affect the shape of the distribution and make it non-normal. Income distribution is a powerful example of a **positively skewed distribution**. Take a look at Figure 4.3, which contains a positively skewed distribution. Let's walk through Figure 4.3. The mode tends to be toward the low end of the distribution. If most scores are low, then why is it called a positively skewed distribution? That's because a few high (positive) scores are skewing the distribution. The high scores are "weird" or unusual in the distribution, hence, the term *positively skewed*.

Positively skewed distribution: data distribution in which there are a few unusually high (skewed) scores; most scores tend to be toward the low end of the distribution.

In a positively skewed distribution, as we saw with the example of incomes in my town, the mean will tend be higher than the mode, with median falling between these two statistics.

If we have positively skewed distributions, we must of course have **negatively skewed distributions**. Recall that in a positively skewed distribution, the mean is greater than the mode. So in a negatively skewed distribution, guess which is greater, the mean or the mode? If you don't know or don't feel like guessing, take a look at Figure 4.4. Here, the mode is toward the high end of the range of scores. However, the mean is less than the mode because a few low scores are dragging it down from the mode. The low scores are "weird" or skewed, so that's why it is called a negatively skewed distribution. For example, most college seniors apply for a lot of jobs during their last year of school. That's a normal behavior to engage in when looking for your first "post–college-degree" job. Therefore, the mode is fairly high. A few seniors may complete only a few or no job applications. These few seniors are outliers that drag down the value of the mean of this distribution. Therefore, it is negatively skewed.

Negatively skewed distribution: data distribution in which there are a few unusually low (skewed) scores; most scores tend to be toward the high end of the distribution.

When looking at a skewed distribution, you should "follow the tail." That is, look for the end of the distribution that contains very few scores. If there are few scores on the high

[PLEASE SEE NOTE FROM AUTHOR: "I SENT MORGAN HAND-DRAWING OF THIS AND OTHER REQUESTED "STOCK" FIGURES; HOWEVER, I DON'T KNOW IF THEY'VE BEEN PULLED/ FINALIZED OR NOT. THAT WAS ON APRIL 1 THAT I SENT HER THAT INFORMATION, SO I DOUBT THEY'D BE READY SO QUICKLY, BUT I CANNOT SAY FOR SURE.]

Figure 4.2 A Normal Distribution of Scores

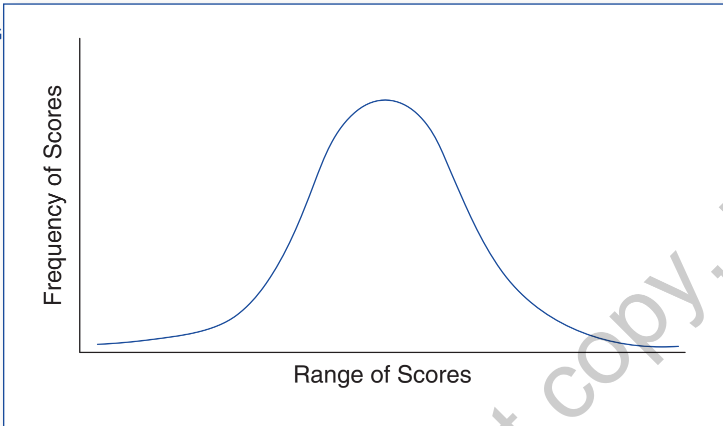


Figure 4.3 A Positively Skewed Distribution of Scores

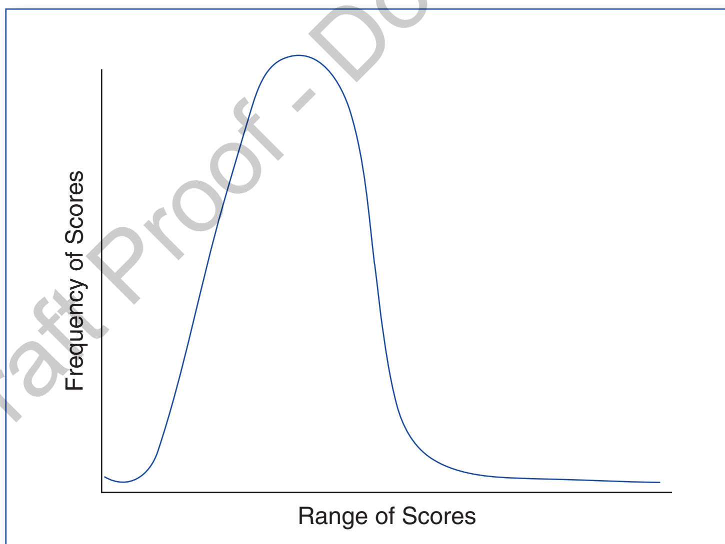
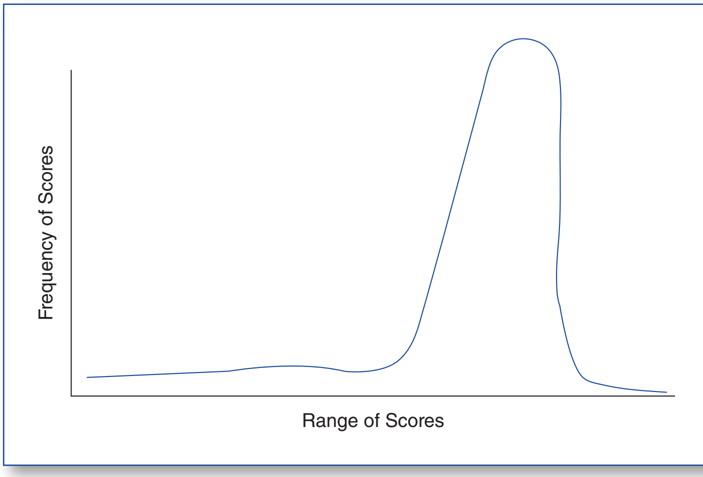


Figure 4.4 A Negatively Skewed Distribution of Scores



end of the distribution, it is positively skewed. If there are few scores on the low end of the distribution, it is negatively skewed.

Consideration 3: A variable's scale of measurement

To calculate a mean, we need all scores in a dataset. That's one critical reason why researchers tend to report the mean more than the other two measures of central tendency. *To calculate a mean, data need to be measured with a scale measurement*. Suppose, for instance, we had ordinal (rank-ordered) data. If we have data that are rankings, there is no way to calculate an interpretable mean. Likewise, with nominal (categorical) data, such as a person's sex, the mean is an impossible statistic to calculate.

For scale data, it is not a crime to report all three measures of central tendency as Wendt (2013) did. Again, researchers tend to prefer the mean and often report it without mentioning a median or mode. In reading published research studies, this is generally not a problem. However, be careful when you read or hear news stories that report a "mean" or "average." Always think about what these stories are telling you and whether outliers in the data may have skewed the results in the story being reported.

As we discussed in the two previous considerations, the median is a good alternative to the mean when we have outliers that may skew the picture of the data that the mean will paint for us. To calculate a median, we need scale data or ordinal data. In fact, *for ordinal data, the median is typically the most preferred measure of central tendency*. For example, suppose we look at the *U.S. News & World Reports* rankings of colleges and various graduate programs of study. These data are, of course, ordinal. We learned in Chapter 2 that we cannot tell how much one ranked score differs from the next ranked score. If University A is ranked 23rd and University B is ranked 24th, how much better, according to this ranking system, is University A than University B? We have no idea. Therefore, calculating a mean is meaningless. We use the median for ordinal data because it gives us an idea, albeit an imperfect one, of an entity's standing in relation to some standard.

What college or university do you attend? What is your class standing at your college? Data collected from these sorts of questions are all measured with a nominal measurement scale. *When using a nominal scale of measurement, we have to report the mode for that variable*. It would make no sense to report a mean or median. For instance, suppose we collected data on class standing (first-year, sophomore, junior, or senior). There is no practical way to calculate a mean or median on such data. However, we can calculate a mode simply by seeing how many students of each class participated in our research. If there were more sophomores than any other class standing, the mode would be sophomores. The mode is also a good choice to report when the data are bimodal or multimodal, especially in a large dataset.

Consideration 4: Open-ended response ranges

In many psychological studies, there are predetermined upper and lower limits on a response range. For instance, in question 2 in the previous Learning Check, scores on a midterm could fall only between 0% and 100%. But what happens when there is no limit on scores in a distribution? For example, I like to complete customer satisfaction surveys not only because I get a chance to win stuff, but also because as a social scientist,

Table 4.4 Typical Items on Customer Satisfaction Surveys

How many people were in line ahead of you at checkout?
_____ None
_____ One person
_____ Two people
_____ Three people
_____ Four people
_____ Five or more people
Including this visit, how many times have you visited this store in the past 30 days?
_____ Once
_____ Twice
_____ Three times
_____ Four times
_____ Five or more times

I think I can provide useful feedback that other researchers can make use of. On these sorts of surveys, we might see items such as the ones in Table 4.4.

What does "five or more" mean in terms of quantitative information? It could mean 5, 6, 9, 27, or some other number. We don't know. In this case, the researchers doing the survey should rely on the median response.

How might we calculate a median in this circumstance? Let's consider the top portion of Table 4.5, which contains a small dataset of customer responses. In these data, we see four respondents with no one in front of them at checkout, four respondents with one person in front of them at checkout, and so on. To calculate a median, we arrange the numbers from lowest to highest. Doing so gives us the following:

0, 0, 0, 0, 1, 1, 1, 1, 2, 2, 3, 3, 4, 5, 5

To find a median with an odd number of scores, we take the two numbers in the middle of the distribution and find the mean of those two numbers:

0, 0, 0, 0, 1, 1, 1, 1, 2, 2, 3, 3, 4, 5, 5

The mean of 1 and 1, is, of course 1. Thus, our median is 1. Although not a perfect representation of the dataset, using the median with an open-ended response range paints as valid of a picture the data as possible.

Table 4.5 Frequencies on Typical Items on Customer Satisfaction Surveys

How many people were in line ahead of you at checkout?	
	f **(frequency)**
_____ None	4
_____ One person	4
_____ Two people	2
_____ Three people	2
_____ Four people	1
_____ Five or more people	2

Including this visit, how many times have you visited this store in the past 30 days?	
	f (frequency)
_____ Once	6
_____ Twice	2
_____ Three times	1
_____ Four times	3
_____ Five or more times	2

LEARNING CHECK

1. A distribution of scores has a mean = 30, median = 20, and mode = 10. The distribution
   a) has a positive skew.
   b) has a negative skew.
   c) is normal.
   d) is bimodal.

   A: a
2. Explain your response to the previous question.

   A: The mean is greater than the other two measures of central tendency because of a few high scores. These high scores are the unusual scores in the dataset, which makes the distribution *positively* skewed.
3. Seven friends have a mean income of $300/week, and their median income is $270/week. Rich, the lowest paid, gets fired from his $200/week job and now has an income of $0/week. What is the median weekly income of the seven friends after Rich lost his job?

   A: The median is still $270/week.
4. Explain your response to question 3.

   A: Let's arrange these data (before Rich got fired) so that we have a mean of $300 and a median of $270. Here is one way to do so (but if you can do it another way, go for it; it won't change the logic behind this question):

   $500, $350, $300, $270, $240, $240, $200

   The $270 is the median because it divides the dataset in half, with three incomes above it and three incomes below it.
   Now, with Rich being fired, here are the new incomes:

   $500, $350, $300, $270, $240, $240, $0

Here, the median is $270 because it still divides the dataset in half.

(Continued)

(Continued)

5. A survey asks whether people think a politician is innocent or guilty of embezzlement. Which would be the best measure of central tendency to describe this dataset?
   a) mean
   b) median
   c) mode

   A: c
6. Explain your response to the previous question.

   A: The data are nominal; that is, people believe the politician is innocent or guilty. Thus, with categorical data as in this example, the mode is the best measure of central tendency to report.
7. Look again at Table 4.5. Look at the second question, in the lower portion of that table.

   What is the correct measure of central tendency to report?

   A: median

   Why is it the correct measure of central tendency to report?

   A: It has an open-ended response range; that is, "five or more times" is an unknown quantity.

   What is the value of this measure of central tendency?

   A: 2.5

Measures of Central Tendency and SPSS

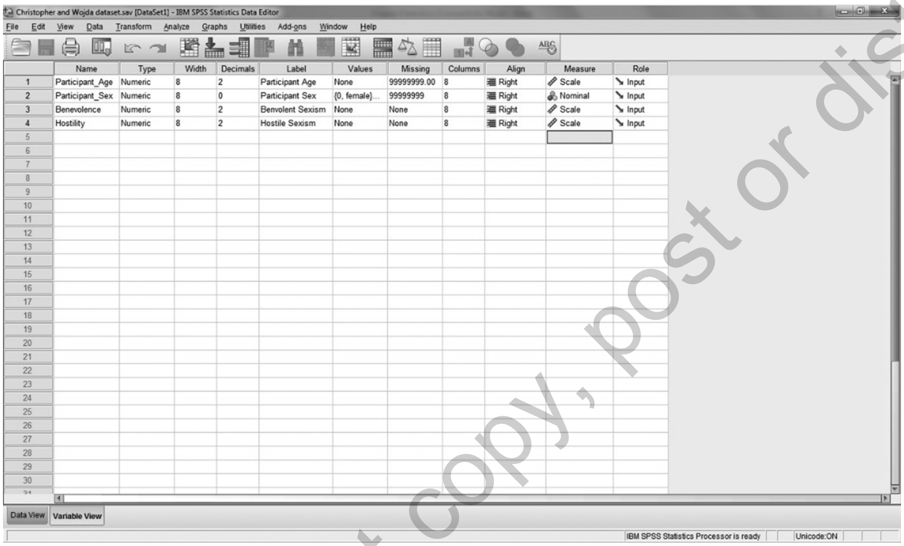
As we learn to generate measures of central tendency using SPSS, we will discuss a new study and use its data. Then in the Learning Check, we'll return to Wendt's (2013) data and test our understanding of using SPSS with the measures of central tendency.

We will use data from a study by Andrew Christopher and Mark Wojda (2008), who examined sources of sexist attitudes toward women. In their research, Christopher and Wojda studied two types of sexist attitudes toward women: hostile sexism and benevolent sexism. Hostile sexism is a type of sexism/prejudice in which people view women in a blatantly negative and mean way. For example, hiring a less qualified man over a more qualified woman or making openly disparaging comments about women would be two instances of hostile sexism. Benevolent sexism is a type of sexism/prejudice in which people view women with subjectively positive feelings (e.g., more sensitive to other people's feelings) but that place women in roles restrictive relative to men. Let's take an example of benevolent sexism. If a person thinks women are indeed more sensitive to other people's feelings, then that person might believe that women won't make good corporate leaders because they cannot make tough decisions for fear of upsetting other people.

In their research, Christopher and Wojda (2008) used an online survey to collect data from 349 U.S. citizens. The sample consisted of 182 women and 167 men, some as young as 18, some as old as 82. Included in a set of questionnaires were measures of hostile and benevolent sexism (Glick & Fiske, 1996). The measure of hostile sexism contained 11 items, such as "Most women fail to appreciate fully all that men do for them" and "When women lose to men in a fair competition, they typically complain about being discriminated against." The measure of benevolent sexism contained 11 items, such as "Women should be protected and cherished by men" and "Women, when compared to men, tend to have a superior moral sensibility." Participants used a 0 (*strongly disagree*) to 5 (*strongly agree*) response range to indicate the extent to which they agreed with each statement.

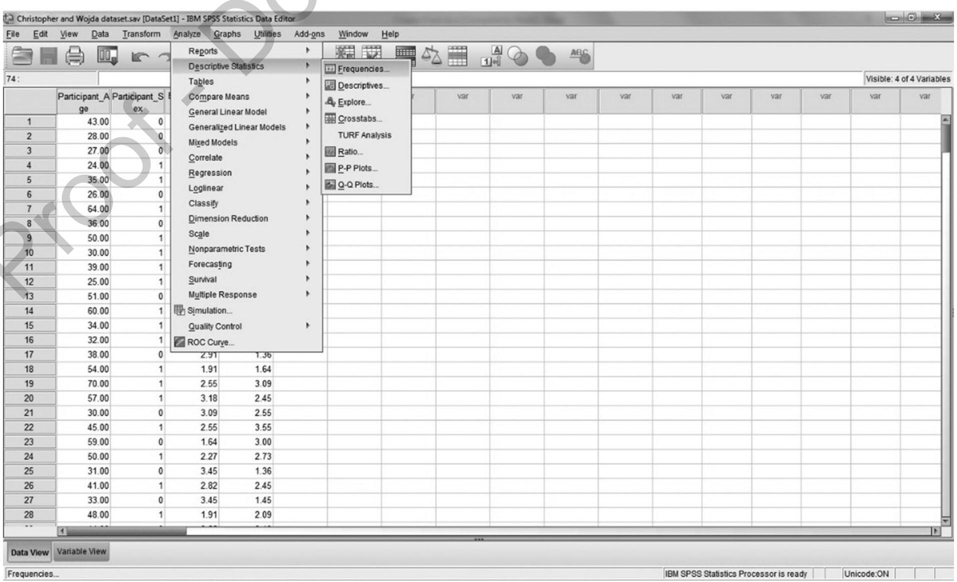
You have access to Christopher and Wojda's (2008) data in the SPSS file "Christopher and Wojda dataset.sav". Open that SPSS file now so that we can learn how to use SPSS to generate the measures of central tendency.

Before we generate our measures of central tendency, please humor me and click on the *Variable View* in the bottom left portion of the spreadsheet. Before we deal with our data, we must understand the nature of our data, and that is what *Variable View* does for us. For instance, notice the variable `Participant Sex`. Why do we need values for this variable? Why is its measure nominal? Why are benevolent and hostile sexism both scale measures? Understanding these sorts of issues is essential to generating output and to making sense of that output.

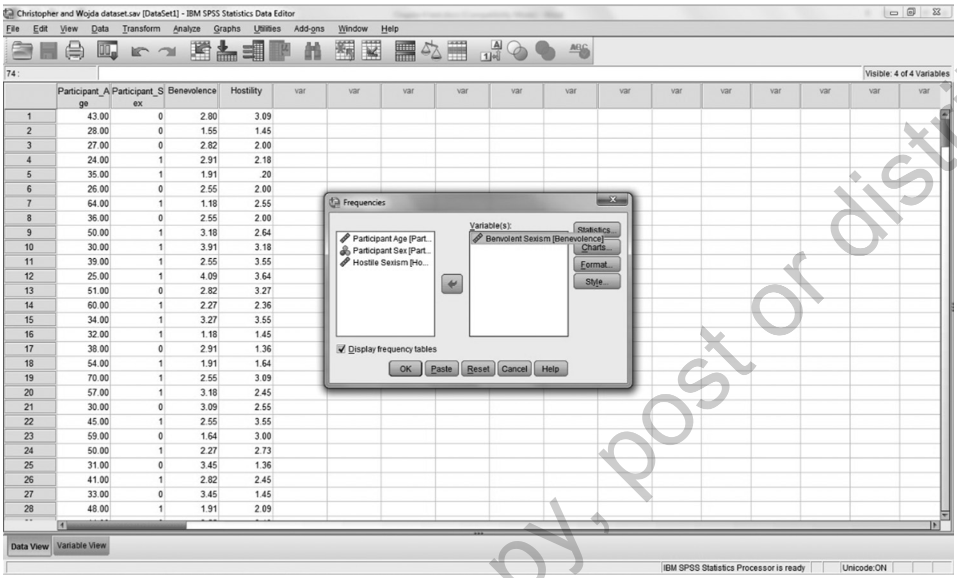


Click on *Data View* to see the data we are working with. To generate the measures of central tendency, here's what you do:

1. Click on *Analyze* → *Descriptive Statistics* → *Frequencies*.

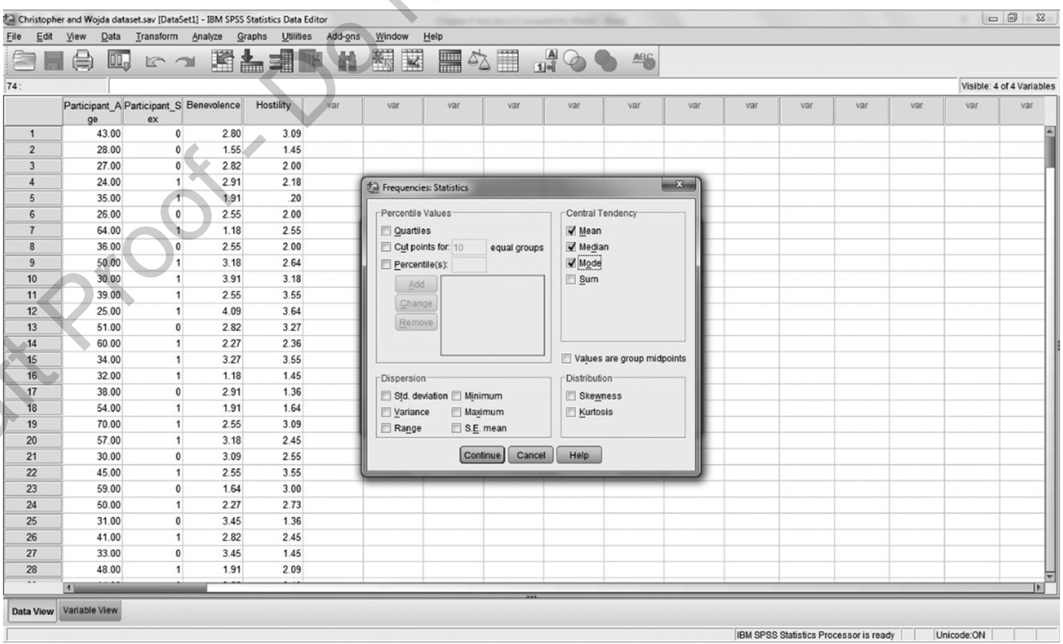


2. In the *Frequencies* window, we click into the *Variable(s)* box to find the variables we want descriptive statistics for. Here, let's get them for the variable `Benevolent Sexism`.



3. Now click on the *Statistics* box in the upper right part of the *Frequencies* window. Here is where we tell SPSS what information to give us. Right now, we're after the three measures of central tendency. These appear in the upper right part of the *Frequencies: Statistics* box.

4. Click *Continue* to return to the *Frequencies* window.



5. Notice how the *Display frequencies table* box is checked in the lower left corner of the window? Please keep that checked as it might come in handy in a moment.
6. Click *OK*, and your measures of central tendency will appear.

Let's start at the top of the SPSS output with the *Statistics* box. First, notice at the top of the box is `Benevolent Sexism`. This is our variable for which we generated this output. In the box is information about our sample size (*N*). We have 349 valid responses, and no missing data for this variable.

Next, we have listed each one of our three measures of central tendency. As we can see, our mean is 2.47; our median is 2.45; and our mode is 2.45. However, do you see the footnote after the mode? Multiple modes exist, but SPSS gives us only the smallest of the modes (how inconsiderate, I know).

Statistics

Benvolent Sexism

N	Valid	349
	Missing	0
Mean		2.4706
Median		2.4545
Mode		2.45[a]

a. Multiple modes exist. The smallest value is shown

Benvolent Sexism

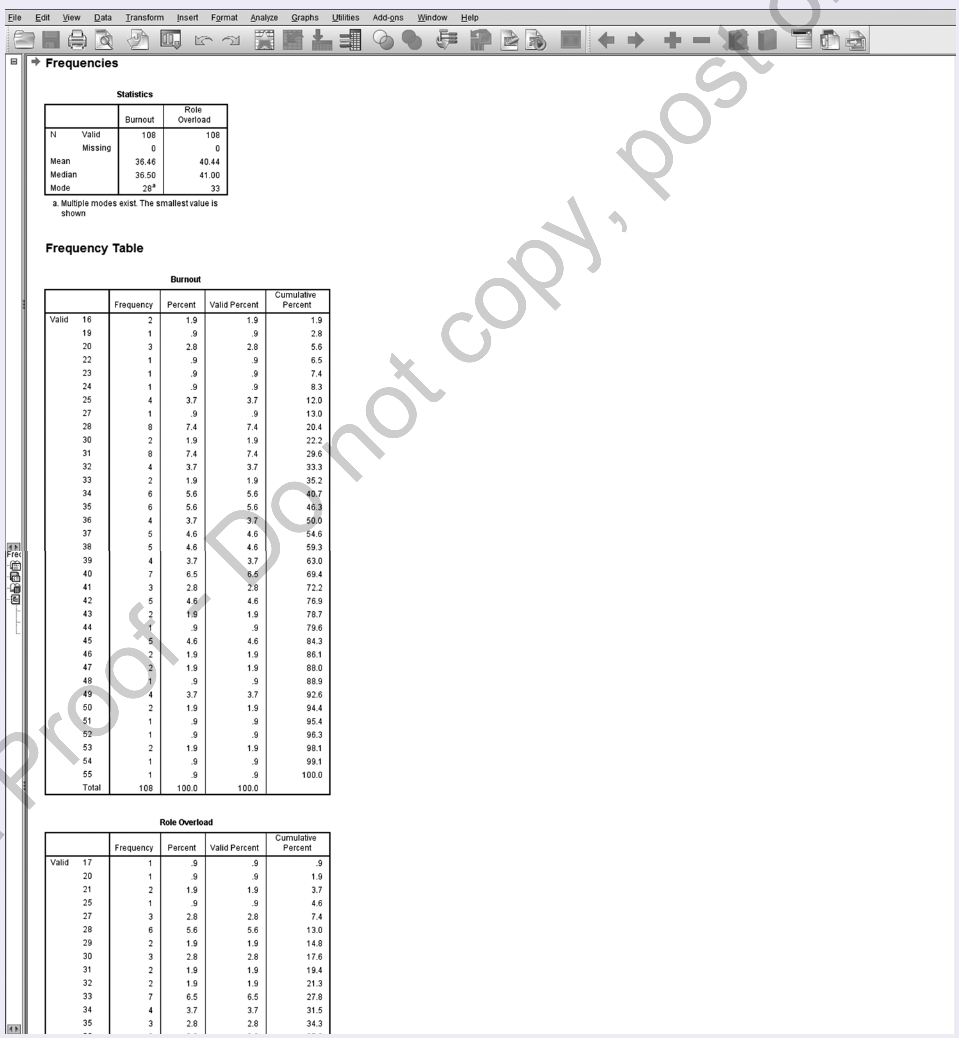
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.27	2	.6	.6	.6
	.36	1	.3	.3	.9
	.45	1	.3	.3	1.1
	.64	1	.3	.3	1.4
	.73	4	1.1	1.1	2.6
	.82	1	.3	.3	2.9
	.91	4	1.1	1.1	4.0
	1.00	2	.6	.6	4.6
	1.09	1	.3	.3	4.9
	1.18	8	2.3	2.3	7.2
	1.27	4	1.1	1.1	8.3
	1.36	2	.6	.6	8.9
	1.45	9	2.6	2.6	11.5
	1.55	6	1.7	1.7	13.2
	1.64	7	2.0	2.0	15.2
	1.70	1	.3	.3	15.5
	1.73	11	3.2	3.2	18.6
	1.82	10	2.9	2.9	21.5
	1.90	1	.3	.3	21.8
	1.91	13	3.7	3.7	25.5
	2.00	9	2.6	2.6	28.1
	2.09	15	4.3	4.3	32.4
	2.10	1	.3	.3	32.7
	2.18	12	3.4	3.4	36.1
	2.27	12	3.4	3.4	39.5
	2.30	1	.3	.3	39.8
	2.36	16	4.6	4.6	44.4
	2.40	1	.3	.3	44.7
	2.45	19	5.4	5.4	50.1
	2.55	17	4.9	4.9	55.0
	2.60	1	.3	.3	55.3
	2.64	18	5.2	5.2	60.5
	2.73	19	5.4	5.4	65.9
	2.80	1	.3	.3	66.2
	2.82	14	4.0	4.0	70.2
	2.89	1	.3	.3	70.5
	2.91	11	3.2	3.2	73.6
	3.00	17	4.9	4.9	78.5
	3.09	8	2.3	2.3	80.8
	3.18	6	1.7	1.7	82.5
	3.27	10	2.9	2.9	85.4
	3.36	12	3.4	3.4	88.8
	3.45	7	2.0	2.0	90.8
	3.55	6	1.7	1.7	92.6
	3.64	4	1.1	1.1	93.7
	3.73	4	1.1	1.1	94.8
	3.82	4	1.1	1.1	96.0
	3.91	2	.6	.6	96.6
	4.00	2	.6	.6	97.1
	4.09	4	1.1	1.1	98.3
	4.27	2	.6	.6	98.9
	4.36	2	.6	.6	99.4
	4.45	1	.3	.3	99.7
	4.73	1	.3	.3	100.0
	Total	349	100.0	100.0	

So, to determine all of the modes, we need our frequency table that appears next. Let's look at the mode we do know, which is 2.45. We see that its frequency is 19, meaning the score of 2.45 occurred 19 times in the dataset. So, to determine all of the modes, we need to find that other score or scores that occurred 19 times. When we look carefully at the frequency of each score, we see that the score of 2.73 also occurred 19 times. Therefore, our modes are 2.45 and 2.73.

LEARNING CHECK

1. From the previous chapter, return to the file "Wendt's data.sav". From this file, use SPSS to generate the measures of central tendency for the variables of `Burnout` and `Role Overload`.

   A: Here is what your output should look like:



Frequencies

Statistics

		Burnout	Role Overload
N	Valid	108	108
	Missing	0	0
Mean		36.46	40.44
Median		36.50	41.00
Mode		28[a]	33

a. Multiple modes exist. The smallest value is shown

Frequency Table

Burnout

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	16	2	1.9	1.9	1.9
	19	1	.9	.9	2.8
	20	3	2.8	2.8	5.6
	22	1	.9	.9	6.5
	23	1	.9	.9	7.4
	24	1	.9	.9	8.3
	25	4	3.7	3.7	12.0
	27	1	.9	.9	13.0
	28	8	7.4	7.4	20.4
	30	2	1.9	1.9	22.2
	31	8	7.4	7.4	29.6
	32	4	3.7	3.7	33.3
	33	2	1.9	1.9	35.2
	34	6	5.6	5.6	40.7
	35	6	5.6	5.6	46.3
	36	4	3.7	3.7	50.0
	37	5	4.6	4.6	54.6
	38	5	4.6	4.6	59.3
	39	4	3.7	3.7	63.0
	40	7	6.5	6.5	69.4
	41	3	2.8	2.8	72.2
	42	5	4.6	4.6	76.9
	43	2	1.9	1.9	78.7
	44	1	.9	.9	79.6
	45	5	4.6	4.6	84.3
	46	2	1.9	1.9	86.1
	47	2	1.9	1.9	88.0
	48	1	.9	.9	88.9
	49	4	3.7	3.7	92.6
	50	2	1.9	1.9	94.4
	51	1	.9	.9	95.4
	52	1	.9	.9	96.3
	53	2	1.9	1.9	98.1
	54	1	.9	.9	99.1
	55	1	.9	.9	100.0
	Total	108	100.0	100.0	

Role Overload

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	17	1	.9	.9	.9
	20	1	.9	.9	1.9
	21	2	1.9	1.9	3.7
	25	1	.9	.9	4.6
	27	3	2.8	2.8	7.4
	28	6	5.6	5.6	13.0
	29	2	1.9	1.9	14.8
	30	3	2.8	2.8	17.6
	31	2	1.9	1.9	19.4
	32	2	1.9	1.9	21.3
	33	7	6.5	6.5	27.8
	34	4	3.7	3.7	31.5
	35	3	2.8	2.8	34.3

36	3	2.8	2.8	37.0
37	2	1.9	1.9	38.9
38	1	.9	.9	39.8
39	3	2.8	2.8	42.6
40	3	2.8	2.8	45.4
41	6	5.6	5.6	50.9
42	5	4.6	4.6	55.6
43	6	5.6	5.6	61.1
44	6	5.6	5.6	66.7
45	2	1.9	1.9	68.5
46	6	5.6	5.6	74.1
47	4	3.7	3.7	77.8
48	1	.9	.9	78.7
49	2	1.9	1.9	80.6
50	6	5.6	5.6	86.1
51	1	.9	.9	87.0
52	3	2.8	2.8	89.8
53	3	2.8	2.8	92.6
54	2	1.9	1.9	94.4
57	1	.9	.9	95.4
59	1	.9	.9	96.3
60	1	.9	.9	97.2
61	2	1.9	1.9	99.1
62	1	.9	.9	100.0
Total	108	100.0	100.0	

2. Now use this output and answer the following questions about it:
   a) What was the mean `Burnout` score?

      A: 36.46
   b) What was the mode for the `Burnout` variable?

      A: 28 and 31
   c) Interpret what the number(s) in response to question 2b mean(s) in plain English.

      A: The most frequently occurring scores on the `Burnout` variable were 28 and 31.
   d) What was the sample size for the `Role Overload` variable?

      A: 108 respondents
   e) What was the median `Role Overload` score?

      A: 41.0
   f) Interpret what the number in question 2e means in plain English.

      A: 50% of the scores on the `Role Overload` variable were greater than 41.0, and 50% of the scores on the `Role Overload` variable were less than 41.0.
3. Use Christopher and Wojda's (2008) dataset to generate the measures of central tendency for the variable of hostile sexism. After doing so, present them in a table (see Table 4.3 for an example of such a table).

   A: Here is what your output should look like:

File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help

Frequencies

Statistics

Hostile Sexism

N	Valid	349
	Missing	0
Mean		2.1994
Median		2.2727
Mode		2.55

(Continued)

(Continued)

Hostile Sexism

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1	.3	.3	.3
	.09	2	.6	.6	.9
	.18	1	.3	.3	1.1
	.20	1	.3	.3	1.4
	.27	3	.9	.9	2.3
	.36	1	.3	.3	2.6
	.45	2	.6	.6	3.2
	.50	1	.3	.3	3.4
	.55	1	.3	.3	3.7
	.73	4	1.1	1.1	4.9
	.82	7	2.0	2.0	6.9
	.90	1	.3	.3	7.2
	.91	8	2.3	2.3	9.5
	1.00	7	2.0	2.0	11.5
	1.09	9	2.6	2.6	14.0
	1.18	10	2.9	2.9	16.9
	1.27	5	1.4	1.4	18.3
	1.36	7	2.0	2.0	20.3
	1.40	1	.3	.3	20.6
	1.45	7	2.0	2.0	22.6
	1.50	1	.3	.3	22.9
	1.55	10	2.9	2.9	25.8
	1.60	1	.3	.3	26.1
	1.64	15	4.3	4.3	30.4
	1.73	8	2.3	2.3	32.7
	1.82	7	2.0	2.0	34.7
	1.91	13	3.7	3.7	38.4
	2.00	15	4.3	4.3	42.7
	2.09	14	4.0	4.0	46.7
	2.18	9	2.6	2.6	49.3
	2.27	10	2.9	2.9	52.1
	2.30	1	.3	.3	52.4
	2.36	14	4.0	4.0	56.4
	2.40	1	.3	.3	56.7
	2.45	16	4.6	4.6	61.3
	2.55	17	4.9	4.9	66.2
	2.64	13	3.7	3.7	69.9
	2.73	11	3.2	3.2	73.1
	2.82	10	2.9	2.9	75.9
	2.91	16	4.6	4.6	80.5
	3.00	7	2.0	2.0	82.5
	3.09	9	2.6	2.6	85.1
	3.13	1	.3	.3	85.4
	3.18	6	1.7	1.7	87.1
	3.20	1	.3	.3	87.4
	3.27	7	2.0	2.0	89.4
	3.36	7	2.0	2.0	91.4
	3.45	3	.9	.9	92.3
	3.55	7	2.0	2.0	94.3
	3.64	8	2.3	2.3	96.6
	3.73	1	.3	.3	96.8
	3.82	1	.3	.3	97.1
	4.00	1	.3	.3	97.4
	4.09	1	.3	.3	97.7
	4.18	4	1.1	1.1	98.9
	4.27	1	.3	.3	99.1
	4.36	1	.3	.3	99.4
	4.64	1	.3	.3	99.7
	4.91	1	.3	.3	100.0
	Total	349	100.0	100.0	

Table 4.6 shows what your table should look like.

Table 4.6 Descriptive Statistics for Hostile Sexism

Variable	*M*	*Mdn*	Mode
Hostile Sexism	2.20	2.27	2.55

Note. Scores on hostile sexism could range from 0 to 5.

Abbreviations. *M* = mean; *Mdn* = median.

MEASURES OF VARIABILITY

In the first half of this chapter, we explored the three measures of central tendency. Although the mean, median, and mode are essential in helping describe a dataset concisely, they do not provide certain information that is often essential in learning about a dataset. In this section, we begin by examining what we mean by "variability" and why it is critical to any research study. We then examine three specific measures of variability and how to calculate and interpret each one. We conclude this chapter by learning how to generate the three measures of variability using SPSS and interpreting the output SPSS provides us. Throughout this half of the chapter, we will, as needed, use the two datasets we used when discussing the measures of central tendency.

What Is Variability? Why Should We Care About Variability?

On average in the United States, people eat 6,000 pieces of pizza during their lifetimes (Reiter, 2015). In addition, the average person uses 90 gallons of water in his or her daily routine (U.S. Geological Survey, 2015). Finally, the average American spends 40 minutes each day checking Facebook feeds (Brustein, 2014). These are interesting (perhaps) pieces of trivia about the typical American. Notice that each one involves the "average" American. In other words, we are talking the mean number of pieces of pizza, the mean number of gallons of water consumed, and the mean amount of time checking Facebook feeds.

You no doubt know people who will eat far less (or more) pizza than the mean of 6,000 pieces in a lifetime. In addition, some people use more or less than 90 gallons of water each day. Some Americans seem to spend all their time on Facebook, whereas others never or rarely check their accounts. In other words, there is a great deal of variability in the extent to which people engage in different behaviors. By **variability**, we are talking about the extent to which scores in a dataset tend to be similar (clustered together) or different (spread out).

Variability: extent to which scores are similar (or different) in a dataset.

In any research study, it is essential that scores on the variables contain variability. For example, suppose in Wendt's (2013) research, each of her 108 participants had the same score on the academic burnout measure. Why would this be a problem? If we want to learn about relationships between variables (such as year in college and academic burnout), there need to be differences among the scores on each measure. Let's look back at a scatterplot that we created in Chapter 3. You can see this scatterplot again in Figure 4.5 here.

It is a scatterplot between role overload and burnout. As you saw in Chapter 3 and can see again, as role overload scores increase, burnout scores tend to increase. Now, let's change it up a bit, and make the scatterplot that Wendt (2013) would have obtained had all 108 of her participants had the same score on the burnout measure.

Figure 4.5

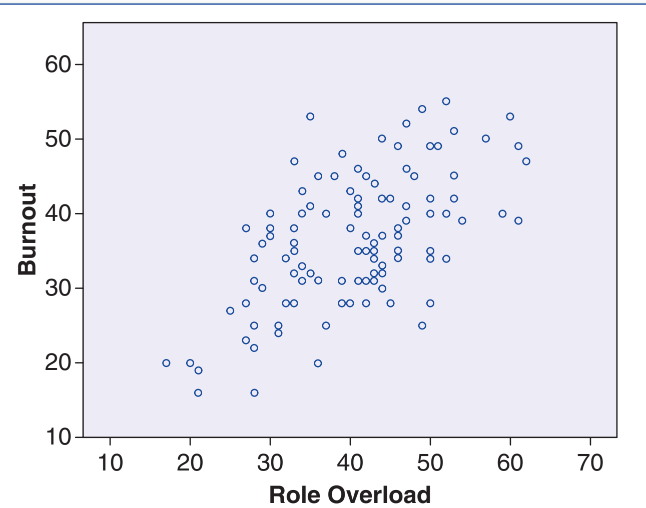
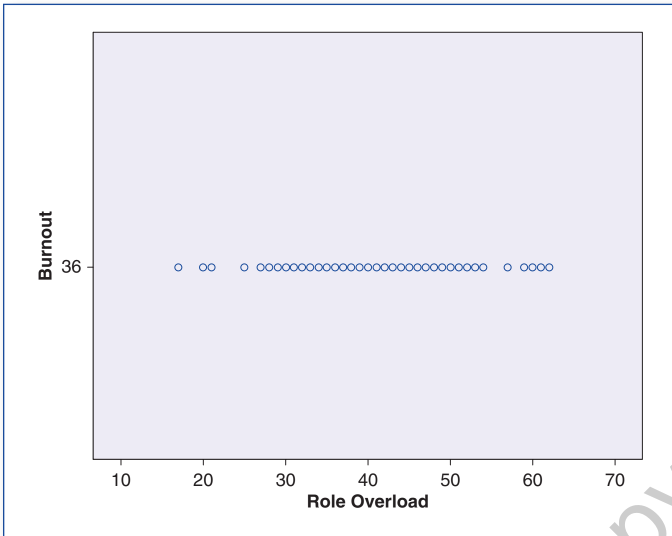


Figure 4.6 Scatterplot With Variability in Role Overload Scores but No Variability in Burnout Scores



For the sake of example, we'll set the burnout score to be 36.5. Now let's examine Figure 4.6 and compare it with Figure 4.5.

Which figure is more informative? I am not sure what we can learn from Figure 4.6 because there is no way to see how burnout is related to role overload. All burnout scores are the same. Because they are all the same, there is no way to learn anything from this research.

Researchers generally recognize the importance of variability in their work. However, there can be occasions when a study lacks variability in one or more of the variables it is studying. For instance, suppose you are taking an advanced college calculus class. Suppose your teacher asks you the following question on a test in this class:

$$1 + 1 = ?$$

I bet you don't need to have taken an advanced calculus class to answer this question correctly. As students, we all enjoy an easy question once in a while. However, if the purpose of a test is to determine who understands the material and who needs help, such a question is not good because everyone is going to get it correct. This is an example of a **ceiling effect**. That is, if everyone is getting 100% correct on such questions, there is no variability to determine mastery of course content. All scores are toward (or, in this case, at) the high end of the possible distribution of scores.

Ceiling effect: lack of variability in a dataset that occurs because scores are clustered together at the top end of the range of possible scores.

Of course, if there are ceiling effects, there must also be the opposite, which indeed are called **floor effects**. Let's discuss an example of a floor effect. Suppose you are still in your advanced college calculus class, and on your test, you see the following question:

How many counties are there in the state of Texas?

As you know from Chapter 2, this is probably not a valid question for a calculus class even though it asks for a quantitative response. Of course, I bet most, if not all, students in this class would answer this question incorrectly. If most or all of the responses are incorrect, it is just as big a problem as questions that everyone answers correctly. Once again, a lack of variability renders this question worthless in determining who does and does not understand the material in this class.

Floor effect: lack of variability in a dataset that occurs because scores are clustered together at the low end of the range of possible scores.

The bottom line is that when measuring behavior and cognition, whether with a test, survey, behavioral observation, or any other method, it is important that such behaviors and cognitions be given a chance to vary. That does not guarantee that they will vary; however, the researcher must not do things that introduce potential ceiling and floor effects into the dataset. The next time you are taking a test and encounter a particularly difficult question, just remember your teacher is trying to avoid ceiling effects. Of course, if the question is too difficult, your teacher runs the risk of introducing a floor effect into the test results.

Three Measures of Variability

Recall from the first half of this chapter that the purpose of the measures of central tendency was to describe a large set of data with one number (realizing that in the case of mode, there could be more than one number). Researchers also need to give readers an idea of how much variability there is in a dataset. Just as there were three measures of central tendency, we have three ways to measure variability. In this section, we will discuss these measures of variability and learn how to calculate each one. To do so, we will return to the small dataset we used in the first half of this chapter to calculate the three measures of central tendency. Here again is this dataset:

6, 5, 4, 6, 7, 8, 3, 5, 9, 2

Range

The simplest measure of variability is the **range**. To calculate the range, we simply take the highest score in a dataset and subtract from it the lowest score in a dataset. Stated differently, we subtract the minimum score from the maximum score. In our dataset, we have the highest scoring being 9 and the lowest score being 2. Thus, the range is 9 – 2 = 7.

Range: the difference between the highest and the lowest score in a dataset.

The range is a nice first step in painting a picture of the variability in a dataset. However, how many numbers are involved in its calculation? Two: the highest score and the lowest score. With only two numbers being used, the range may not be terribly informative, especially if there is an outlier in the dataset. Instead, wouldn't it be great if we could paint a picture of variability based on all numbers in a dataset? As luck would have it, we can do so.

Recall that we said the mean is the most commonly reported measure of central tendency. In part this is because the mean is calculated using all numbers in a dataset. Another reason is because we need the mean to calculate the other two measures of variability, which we will discuss now.

Variance

The second measure of variability is called the **variance**. It is the average squared deviation from the dataset's mean. That definition probably makes little sense right now. It will make sense after we calculate a variance. Formulaically, we have

$$\frac{\Sigma(x-M)^2}{N-1}$$

where

Σ means "summation"

x is an individual score

M is the sample mean

N is the total number of scores in the dataset

Here again is our small dataset:

6, 5, 4, 6, 7, 8, 3, 5, 9, 2

Variance: average squared deviation from the mean.

We know from earlier in this chapter that the mean (*M*) of this dataset is 5.50. This is the starting point for calculating a variance (and the third measure of variability that we'll discuss a little later). Indeed, to say there is variability, it must be in reference to some benchmark. That benchmark is the mean.

Therefore, we need to take each score and subtract from it the mean. Here are those calculations:

Score	Score − Mean
6	0.5
5	−0.5
4	−1.50
6	0.50
7	1.50
8	2.50
3	−2.50
5	−0.50
9	3.50
2	−3.50

To get a number that represents the variability in this dataset, it might be tempting to sum the differences between the scores and the mean. Go ahead and do so now. What did you get? If you got zero, you added correctly. Congratulations. However, there is a problem. There is not zero variability in this dataset. Just look at the numbers in the dataset; obviously, these are not all the same. When we sum the differences between individual scores and their mean, the answer will always be zero. So, we are not done calculating the variance.

What we need to do is get rid of those negative numbers. To do so, we will multiply each difference score by itself. That is, we square each difference score and, thus, eliminate the negative numbers. Let's do that now:

Score	Score − Mean	(Score − Mean)^2
6	0.5	0.25
5	−0.5	0.25
4	−1.50	2.25
6	0.50	0.25
7	1.50	2.25
8	2.50	6.25

Score	Score – Mean	(Score – Mean)^2
3	−2.50	6.25
5	−0.50	0.25
9	3.50	12.25
2	−3.50	12.25
		42.50

Now we sum the squared difference scores. This will give us something called the **sum of squares**, the sum of each score's squared deviation from the mean. In this example, the sum of squares is 42.50.

With our sum of squares calculated, we can now finish computing the variance. To do so, we need the mean sum of squares but with one qualification. Specifically, rather than divide the sum of squares by the sample size of 10, we need to subtract 1 from the sample size. Why do we need to subtract 1 from the sample size? There is likely less variability in a sample than in that sample's corresponding population. Therefore, we try to obtain an "unbiased estimate" of the population variance based on our sample data. By making the variability greater (by decreasing the denominator), we are "correcting" for the fact that there is more variability in a population than in a sample because the population is larger than the sample.

As an expert on the mean from the previous chapter, you know what to do at this point. Take the sum of squares and divide it by the number of scores in the dataset minus 1.[1] Doing so gives us

$$\text{Variance} = \frac{42.50}{(10-1)}$$

$$\text{Variance} = \frac{42.50}{9}$$

$$\text{Variance} = 4.72$$

At this point, I have some good news and some bad news. Let's get the bad news out of the way. Variance is typically not reported in research. Frankly, it is difficult to interpret in any meaningful way. This 4.72 we just calculated does not have any useful meaning to us. You might be asking why we bothered to calculate it. That leads to the good news. We need the variance to calculate the standard deviation, our last measure of variability. It is the standard deviation that researchers use most often to describe the variability of a dataset.

Sum of squares: sum of each score's squared deviation from the mean.

Standard deviation

Conceptually, the **standard deviation** is the average amount by which scores in a dataset tend to vary around the mean. You might be thinking that it sounds a lot like the variance, and it is. Recall that when computing the variance, we needed to square the difference scores to remove the negative difference scores. Now we need to "undo" this procedure. To do so, we take the square root of the variance. Thus, we have

$$\text{Standard deviation} = \sqrt{4.72}$$
$$= 2.17$$

That's it. Formulaically, we have

$$\sqrt{\frac{S(x - M)^2}{N - 1}}$$

Standard deviation: measure of the extent to which scores in a dataset tend to vary around the mean.

Now, we said that the variance is not useful to us (other than as a step to calculate the standard deviation) because it cannot be meaningfully interpreted. How then do we interpret our standard deviation of 2.17? We first need to consider the possible range of scores. Here again is our small dataset:

6, 5, 4, 6, 7, 8, 3, 5, 9, 2

We can see there is no score higher than 9 and no score lower than 2. Given these are only "example data," we don't know the possible range of scores. However, with a given range of 7, 2.17 would be a fairly large standard deviation compared with a dataset with a range of 14. A small value indicates there is little variability among scores in the dataset, whereas a large value indicates there is a great deal of variability among scores in the dataset. In the next chapter, we will learn how to take the descriptive statistics we are learning about now to uncover particular information about almost any dataset in psychology. At that time, we will learn how to use and interpret the standard deviation more precisely.

Reporting variability in research

When reporting measures of variability, like reporting the measures of variability, researchers tend to do so either in the text of an article or in a table. Typically, researchers do not report a variance for two reasons. First, as we said, it is difficult to interpret because we had to square all of the difference scores to calculate it. Second, we can take the standard deviation, which is typically reported, and square it to obtain the variance.

By using scores on her burnout measure, Wendt (2013) could have reported the measure of variability in the text as such:

For the range, there are no symbols. It would simply be "The range was 39."

For the standard deviation, it would be reported as "$SD = 8.75$."

Table 4.7 contains the range and standard deviation for the burnout measure in Wendt's (2013) research.

Table 4.7 Descriptive Statistics for Variables in Wendt's (2013) Research

Variable	Range	*SD*
Participant Age	23 – 15 = 8 years	1.73 years
Burnout	55 – 16 = 39	8.75
Dysfunctional Perfectionism	54 – 11 = 43	8.24
Role Overload	62 – 17 = 45	9.68

Notes. For burnout, scores could range from 15 to 75. For dysfunctional perfectionism, scores could range from 11 to 55. For role overload, scores could range from 13 to 65.

Abbreviation. *SD* = standard deviation.

LEARNING CHECK

1. In this chapter's first Learning Check, we used data in which a researcher surveyed 10 undergraduate psychology majors about their study behaviors. Here again is that list of the number of hours they spent studying on the weekend:

6	5	3	4	9
7	3	7	8	3

   a) Calculate the range.

      A: Highest score − lowest score = 9 − 3 = 6

   b) Calculate the variance.

      A: 4.94

   c) Calculate the standard deviation.

      A: 2.22

2. Why is "sample size − 1" used when calculating the sample standard deviation?

   A: We almost always work with sample data. Because a sample is smaller than its population, the sample will have less variability than its population. When calculating a sample standard deviation, we must correct for this state of affairs by "inflating" our calculation to try and reflect the population. With sample data, we subtract 1 from the sample size, thus, decreasing the denominator of the calculation. With a smaller denominator, we are increasing the quotient (increasing variability in describing the dataset).

3. The value of one score in a dataset is changed from 20 to 30. Which measure(s) of variability is (are) certain to be changed?

   a) the range.

   b) the variance.

   c) the range and standard deviation.

   d) the variance and standard deviation.

      A: d

4. Explain your answer to question 3.

   A: The range needs only the highest and lowest scores in the dataset for its calculation. Therefore, it may not change if that one score being changed isn't the highest or lowest one. However, the variance uses all numbers in a dataset in its calculation. Therefore, it is bound to change when one score is changed. The standard deviation is the square root of the variance, so it too must change when one score is changed.

5. Which measure of variability would be more affected by an outlier: the range or the standard deviation?

   A: the range

6. Explain your response to question 5.

   A: Because of how the range is calculated, an outlier will affect it more than the outlier would affect the standard deviation. An extreme score will affect the standard deviation but just not as much it will affect the range because the standard deviation uses all numbers in a dataset in its calculation.

(Continued)

(Continued)

7. You are reading a research article and you see Table 4.8.

Table 4.8 Descriptive Statistics for Variables in This Research

Variable	Range	*SD*
Participant Age	67 – 25 = 42 years	11.73 years
Benevolent Sexism	4.96 – 0.97 = 3.99	1.09
Hostile Sexism	4.79 – 0.50 = 4.29	1.89

Note. For both benevolent sexism and hostile sexism, scores could range from 0 to 5.

Abbreviation. *SD* = standard deviation.

Use this table to answer the following questions:

a) What was the range of participant ages in this research?

A: 25 years old to 67 years old, for a range of 42 years

b) What is the standard deviation for the measure of benevolent sexism?

A: 1.09

c) Interpret the difference in standard deviations for the measures of benevolent sexism and hostile sexism.

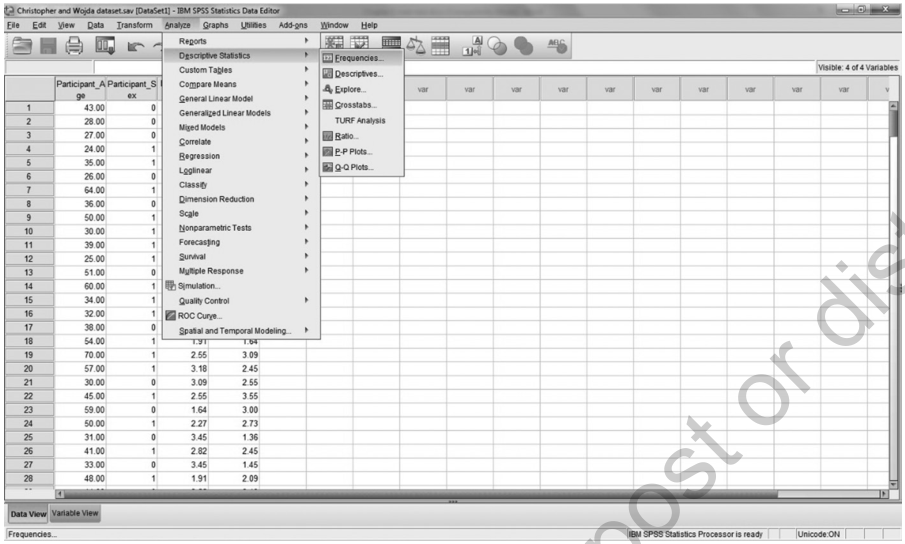
A: The standard deviation for the measure of hostile sexism is higher than it is for benevolent sexism. This difference indicates more variability in hostile sexism scores than in benevolent sexism scores.

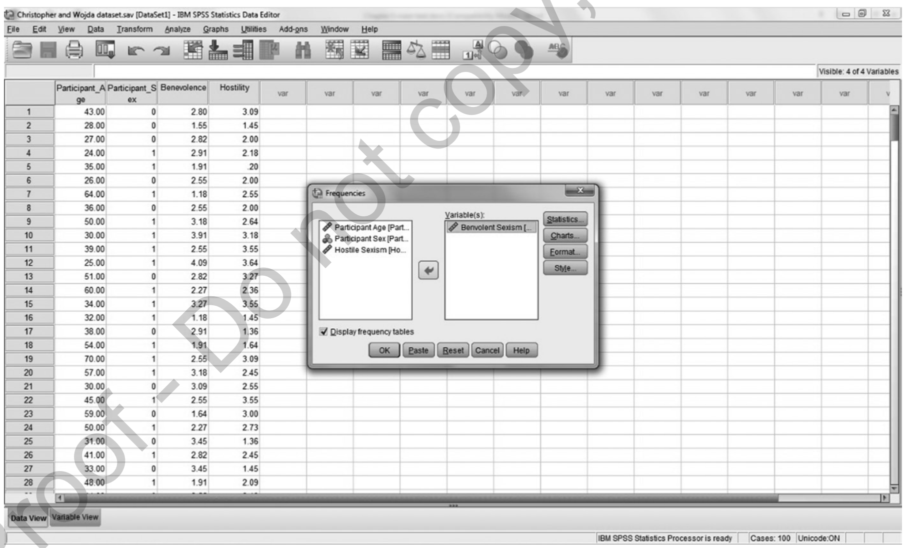
Measures of Variability and SPSS

Earlier in this chapter, we used a dataset that Christopher and Wojda (2008) gathered in which they examined hostile sexism and benevolent sexism toward women. We are going to use their data again here to learn how to generate and interpret the measure of variability using SPSS. If you want to refresh yourself on this research study, please reread the "Measures of Central Tendency and SPSS" section.

Once again, use the dataset titled "Christopher and Wojda dataset.sav". Open that SPSS file now, and let's get to work. To generate the measures of variability we have discussed,

1. Click on *Analyze* → *Descriptive Statistics* → *Frequencies*.
2. In the *Frequencies* window, we click on the variable in the *Variable(s)* box for which we want statistics (in this case, measures of variability). Let's get them for the variable `Benevolent Sexism`.
3. Now click on the *Statistics* box in the upper right part of the *Frequencies* window. Here is where we tell SPSS what information to give us. Right now, we're concentrating on the measures of variability, which SPSS calls measures of "dispersion" in the lower left corner of this window. There you will find the range,

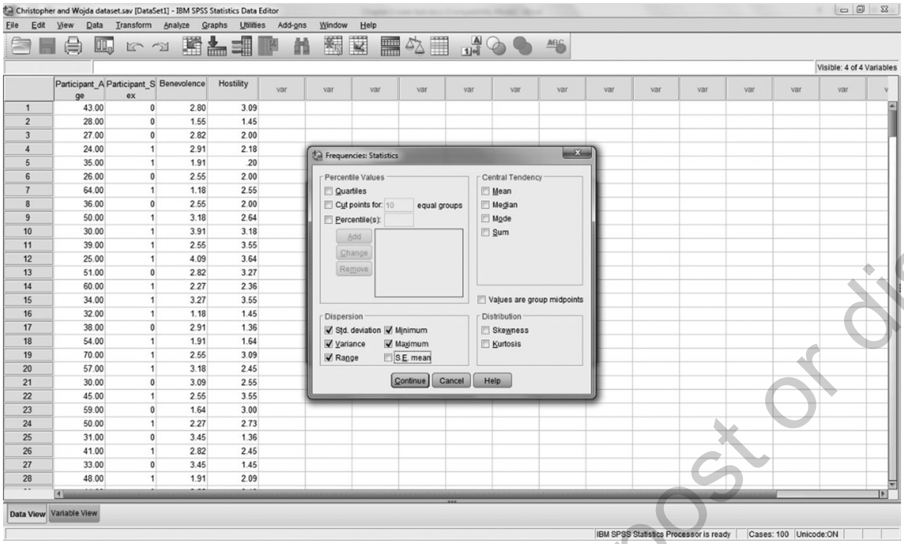


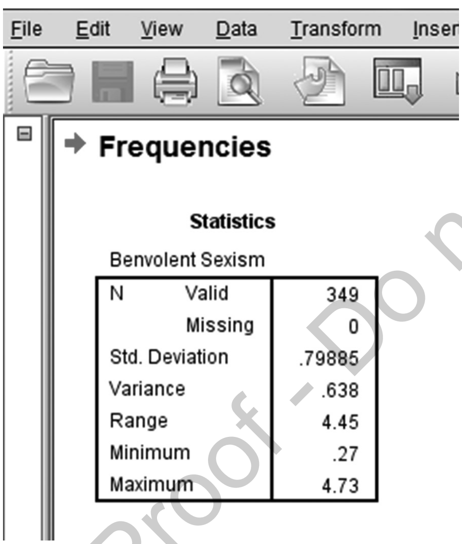


variance, and standard deviation. Select these measures, and for the sake of completeness, select *Minimum* and *Maximum* as well.

4. Click *Continue* to return to the *Frequencies* window.
5. Click *OK*, and your output will soon appear.

(In the output displayed in this screenshot, I deleted the frequency distribution table.)





Statistics

Benvolent Sexism

N	Valid	349
	Missing	0
Std. Deviation		.79885
Variance		.638
Range		4.45
Minimum		.27
Maximum		4.73

As we did when learning to interpret SPSS output for the measures of central tendency, let's start with the *Statistics* box, which contains the measures of variability statistics. First, we have our *N* (sample size), which is 349 with no missing data.

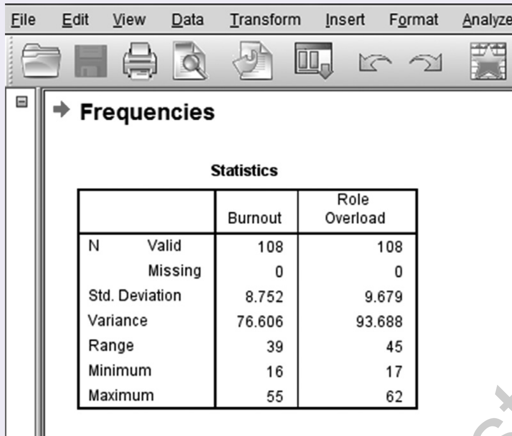
Next are the statistics for each of the three measures of central tendency. Remember that the standard deviation is the square root of the variance, and as we said previously, the variance is rarely reported in research because it is almost redundant with the standard deviation. The standard deviation is more interpretable than the variance because, just like the raw data in the SPSS spreadsheet, the standard deviation does not contain any numbers that have been squared.

It is also important when interpreting a standard deviation to consider the possible range of scores on a measure. Recall that mean scores on the measure of benevolent sexism could range from 0 to 5 and, in fact, did range from 0.27 to 4.73 (a range of 4.45). In the next chapter, we will examine how the measures of central tendency and variability can be used together to allow us to get even more information from a dataset than we already can get.

LEARNING CHECK

1. Return to the SPSS file "Wendt's data.sav". From this file, use SPSS to generate the measures of variability for the variables `Burnout` and `Role Overload`.

   A: Here is what your SPSS output should look like:



Statistics

		Burnout	Role Overload
N	Valid	108	108
	Missing	0	0
Std. Deviation		8.752	9.679
Variance		76.606	93.688
Range		39	45
Minimum		16	17
Maximum		55	62

Now, use this output and answer the following questions about it:

a) What was the range of scores for the `Burnout` measure?

A: 55 − 16 = 39

b) For the `Burnout` measure, explain how the standard deviation was calculated.

A: It is the square root of the variance; that is, $8.752 = \sqrt{76.606}$.

c) What was the sample size for the `Burnout` measure?

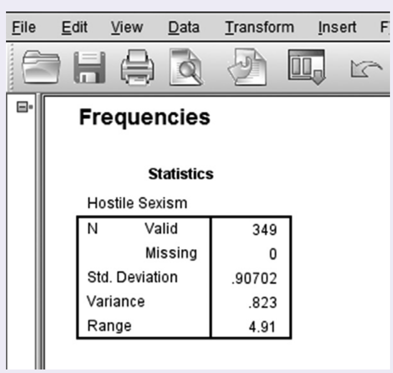
A: 108 participants

d) What was the standard deviation for the `Role Overload` measure?

A: 9.68

2. Use Christopher and Wojda's (2008) dataset to generate the measures of variability for the variable `Hostile Sexism`. After doing so, present them in a table (see Table 4.8 for an example of such a table).

A: Here is what your output should look like:



Statistics

Hostile Sexism

N	Valid	349
	Missing	0
Std. Deviation		.90702
Variance		.823
Range		4.91

(Continued)

(Continued)

Table 4.9 shows what your table should look like:

Table 4.9 Descriptive Statistics for Hostile Sexism

Variable	Range	*SD*
Hostile Sexism	4.91 – 0 = 4.91	0.91

Notes. Scores on hostile sexism could range from 0 to 5.

NOTE

1. Recalculate the variance by taking 42.50 and dividing by 10 instead of 10 – 1. Do you see how the variance is now smaller than without subtracting 1 from the sample size? Without subtracting 1, we are likely underestimating how much variability is in the population because of the fact we are, necessarily, using sample data. There is more variability in the population than in a sample from that population because, as you already know, populations are larger than any of their corresponding samples.

CHAPTER APPLICATION QUESTIONS

1. Why can't we calculate a mean when we have nominal data?

   A: Nominal data are categorical data. Therefore, any numerical value we place on these categories is arbitrary and has no interpretable meaning. Without scale data, we cannot calculate a mean.

3. The only measure of central tendency we are certain to observe as a value in our dataset is:

   a) the mean.

   b) the median.

   c) the mode.

   d) all measures of central tendency must be actual values in the distribution.

   A: c

4. Explain why "c" is the correct answer to question 3.

   A: The mean does not need to lie in our dataset. Suppose we have the following dataset:

   5, 5, 5, 5, 7, 7, 7, 7

   The mean is 6, and the median is 6. However, the number 6 is not in the dataset. The only measure of central tendency guaranteed to be in the dataset is the mode (which, in this example, is 5 and 7).

6. A survey asked students which pizza place they preferred. The results are as follows:

Pizza Place	*f* (frequency)
Late Night Pizza®	5
Papa John's®	6
Pizza Hut®	3
Little Caesars®	5

a) What is the best measure of central tendency for this data?

A: Mode

b) Explain your answer to the previous question.

A: These are nominal data, so the only appropriate measure of central tendency is the mode. It would make no sense to calculate a mean or median in this example.

c) What is the mode of this distribution?

A: Papa John's® is the mode because it has the greatest frequency of student preference.

10. Not everything naturally follows a normal distribution, such as incomes in the United States. The distribution of salaries in the United States is:

a) negatively skewed because poor people represent outliers who earn significantly less than most people.
b) positively skewed because poor people represent outliers who earn significantly less than most people.
c) negatively skewed because rich people represent outliers who earn significantly more than most people.
d) positively skewed because rich people represent outliers who earn significantly more than most people.

A: d

13. In a positively skewed distribution, Betty scored the mean, Barney scored the median, and Fred scored the mode. Who had the highest score?

a) Betty

b) Barney

c) Fred

d) They all scored approximately the same.

A: a

14. Explain your response to the previous question.

A: In a positively skewed distribution, a few unusually high scores "drag up" the mean so that it is higher than most of the other scores. In a positively skewed distribution, most scores are toward the low end of the distribution; the mean is larger than the median or mode in a positively skewed distribution.

16. Explain the difference between a ceiling effect and a floor effect.

A: A ceiling effect occurs when a dataset has a lack of variability because scores tend to cluster together at the high end of the possible range of scores. A floor effect also involves a lack of variability in a dataset; however, with a floor effect, scores tend to cluster together at the low end of the possible range of scores.

17. Explain the relationship between the variance and the standard deviation.

    A: We need to calculate the variance as the preliminary step in calculating the standard deviation. Once we calculate the variance, we take the square root of it and we have our standard deviation.

18. Use the following sample data for questions a–e:

    7, 8, 0, 2, 6, 4, 5, 7, 6

    a) Calculate the range.

    A: 8

    b) Calculate the mean.

    A: 5.0

    c) Calculate the sum of squares

    A: 29

    d) Calculate the variance.

    A: 6.75

    e) Calculate the standard deviation.

    A: 2.60

19. Which measure of variability is **MOST** affected by extreme scores (i.e., by outliers)?

    a) the range
    b) the median
    c) the standard deviation
    d) the mean

    A: a

20. Dr. Hill noticed that the distribution of students' scores on his last math test had an extremely large standard deviation. This fact indicates that the:

    a) test was a good measure of students' knowledge.
    b) students' scores tended to vary quite a bit.
    c) students generally performed well on the test.
    d) test was given to a large number of students.

    A: b

21. Gunnar wants to know how consistent his bowling scores have been during the past seasons. Which of the following measures would provide the most appropriate answer to his question?

    a) the mean
    b) the median
    c) the standard deviation
    d) the range

    A: c

22. Explain why choice "c" is the best answer to the previous question.

    A: The mean and median are measures of central tendency and do not provide information about consistency (variability). The range tends to be affected by outliers in the data and requires only two datapoints in its calculation. Because the standard deviation uses all datapoints in its calculation, it is the best measure of consistency (variability) for Gunnar to calculate.

24. In a set of five scores, suppose all scores are 8. What is the value of (a) the median, (b) the mode, and (c) the standard deviation?

    A: (a) median = 8; (b) mode = 8; (c) standard deviation = 0

QUESTIONS FOR CLASS DISCUSSION

2. Why can't we calculate a mean when we have ordinal data?

5. Use the following distribution to answer the three questions that follow it:

Score	*f* (frequency)
12	7
11	7
10	6
9	4
8	0
7	0
6	0
5	1

   a) The above distribution:
      a) has a positive skew.
      b) has a negative skew.
      c) is normal.

   b) The mode for the above distribution is:
      a) 7.
      b) 0 and 7.
      c) 11 and 12.
      d) 6, 7, and 8.

   c) Which of the following numbers would be considered an outlier in the above distribution?
      a) 0
      b) 1
      c) 5
      d) 7

7. How does it affect the mean when you add a constant to every score? That is, if an instructor adds 5 points to everyone's test score, how will the mean change?
   a) The new mean and the old mean will be the same.
   b) The new mean will be 5 points higher than the old mean.
   c) The new mean will be 5 points less than the old mean.
   d) There is not enough information to answer this question.

8. Let's return to an example from a prior Learning Check. Seven friends have a mean income of $300/week, and their median income is $270/week. Rich, the lowest paid, gets fired from his $200/week job and now has an income of $0/week.

   What is the mean weekly income of the seven friends after Rich lost his job?

9. Students voted for their preferred professors by ranking them. To describe this dataset, which measure(s) of central tendency should be reported? Explain your response.

11. A sample of scores on a stress questionnaire is 2, 3, 5, 1, 6, 4, and 77. Which measure of central tendency is the most appropriate to describe this distribution?

12. Explain your response to the previous question.

15. Why is variability an absolute necessity in a research study?

23. Here are two sets of data with the same means:

    Dataset A: 14, 15, 15, 16, 16, 16, 17, 17

    Dataset B: 0, 2, 9, 15, 18, 22, 27, 33

   Without doing any calculations, explain which dataset has a larger standard deviation.

25. Why can't we calculate any of the measures of variability for nominal data?