LEARNING OBJECTIVES

After reading this chapter, you should be able to:

1. Distinguish between a population mean and a sample mean.
2. Calculate and interpret the mean, the median, and the mode.
3. Calculate and interpret the weighted mean for two or more samples with unequal sample sizes.
4. Identify the characteristics of the mean.
5. Identify an appropriate measure of central tendency for different distributions and scales of measurement.
6. Compute the mean, the median, and the mode using SPSS.
3.1 Introduction to Central Tendency

A *measure of central tendency* is a single score that tends to be located at or near the center of a distribution and is used to represent or describe a data set. Three common measures of central tendency are the *mean*, the *median*, and the *mode*.

- We use different notation for a population size and sample size.
  - Population size = $N$ (all individuals belonging to a group of interest).
  - Sample size = $n$ (a subset of all individuals belonging to a group of interest).

3.2 Measures of Central Tendency

*Mean:* The mean is the balance point or the average value in a data distribution.

- The mean is calculated by taking the sum of the scores and dividing them by the population or sample size.
  - For a population: $\mu = \frac{\sum x}{N}$.
  - For a sample: $M = \frac{\sum x}{n}$.

*Weighted mean:* The weighted mean is the combined mean of two or more groups of scores, where the number of scores in each group is disproportionate or unequal.

- The weighted mean formula is $M_w = \frac{\sum (M \times n)}{\sum n}$.
  - $M =$ the mean for each group.
  - $n =$ the sample size per group.

*Median:* The median is the midpoint of a data distribution such that half of the scores in a distribution occur above it and half occur below it. The median is also located at the 50th percentile in a cumulative percent distribution.

- When there are an odd number of scores in the distribution (or an odd $n$), calculate the median by ordering the scores from least to greatest and counting in to the middle score. The number of scores above and below the median will be the same.
o When \( n \) is an odd number, we locate the position of the middle score using \( \frac{n+1}{2} \) and then count in by that many scores.

- When there are an even number of scores in the distribution (or an even \( n \)), calculate the median by ordering the scores from least to greatest and average the two middle scores.

o When \( n \) is an even number, we use \( \frac{n+1}{2} \) to find the median position. The solution will be a .5 decimal. The middle value is the average of the two middle scores positioned above and below the median position.

**Mode:** The mode is the score that occurs most frequently or often in a data set. It is most commonly used with the mean or median to describe central tendency.

### 3.3 Characteristics of the Mean

The mean is calculated using every score in the distribution. For this reason, changing scores in a data set in any way can influence the value of the mean for that data. In all, there are five characteristics of the mean.

**Changing an existing score:** Because every score in a distribution influences the value of the mean, changing one score into another score will change the mean.

- Increasing the value of an existing score will increase the value of the mean.
- Decreasing the value of an existing score will decrease the value of the mean.

**Adding a new score or removing an existing score:** Adding a new score to the distribution without changing an existing score or deleting a score entirely from the distribution will change the mean as well. Depending on the value of the score that is added or removed, the mean will be affected differently.

- Adding a new score with a value greater than the mean will increase the mean.
- Adding a new score with a value less than the mean will decrease the mean.
- Removing a score with a value greater than the mean will decrease the mean.
- Removing a score with a value less than the mean will increase the mean.
- Adding or removing a value that is equal to the mean will not change the value of the mean.

**Adding, subtracting, multiplying, or dividing each score by a constant:** When every score in a distribution is changed by the same constant, the mean will change by that constant.

- For example, if we add a constant of 5 to each score in a distribution, then the mean will increase by 5. If we subtract this constant from all scores, then the mean will decrease by 5. Likewise, multiplying and dividing a constant will change the mean by that constant.
Summing the differences of scores from their mean: The sum of the differences of scores from their mean is zero. The mean is the balance point of a distribution, which is why the differences of scores above and below the mean cancel out (i.e., their sum is 0).

- The notation for calculating the sum of the differences of the scores from the mean is \( \sum (x - M) \).

Summing the squared differences of scores from their mean: Summing the squared differences of scores from their mean will produce a minimal solution that can’t be achieved by summing the squared difference of the scores from any other value (other than the value of the mean).

- The notation for summing the squared differences of scores from their mean is \( \sum (x - M)^2 \).

3.4 Choosing an Appropriate Measure of Central Tendency

The choice of which measure of central tendency to use to summarize a data set largely depends on the shape of the distribution and the measurement scale of the data.

The mean is used to describe normal distributions and for data on an interval or ratio scale.

- A normal distribution is a distribution of data that are symmetrically distributed around the mean, the median, and the mode.
- The mean is used to describe data measured on the interval or ratio scale as these are the only two scales in which differences between scores and the mean are informative.

The median is used to describe skewed distributions and for data measured on an ordinal scale.

- A skewed distribution is a distribution of data with outliers or scores that fall substantially above or below most other scores in a data set.
  - Positively skewed: A distribution where a few scores fall substantially above most other scores in a data set.
  - Negatively skewed: A distribution where a few scores fall substantially below most other scores in a data set.
- The median is used to describe ordinal data because values on this scale indicate direction only.

The mode is used to describe modal distributions and for data measured on a nominal scale.

- A modal distribution is a data distribution where one or more scores occur the most frequently or often.
  - Unimodal distributions have one mode.
  - Bimodal distributions have two modes.
Multimodal distributions have more than two modes.

Nonmodal distributions have no mode; instead, all scores occur at the same frequency, thus appearing graphically as a straight line.

- The mode is used to describe nominal data because numbers on this scale are not measured in terms of quantity or amount.

3.5 Research in Focus: Describing Central Tendency

Gulledge, Stahmann, and Wilson (2004) used the mean, median, and mode when describing data regarding types of nonsexual romantic physical contact among college students in relationships at Brigham Young University. Identifying all three measures allowed us to identify, describe, and interpret how participants in this study reported their modes of affection.

3.6 SPSS in Focus: Mean, Median, and Mode

SPSS can be used to compute the mean, the median, and the mode. Each measure of central tendency is computed using the Analyze, then Descriptive Statistics and Frequencies options in the menu bar.
Adding or removing scores to or from the mean: When adding or removing a score, the mean will change. When recalculating the new mean, don’t forget to not only add or subtract the difference of the score from the $\sum x$ but also to add or subtract 1 from the sample size because a new score was added, or an existing score deleted, from the original sample size.

Changing values and the mean: When a value is greater than the mean and it is added to the $\sum x$, the value of the mean will increase. However, when a value that is greater than the mean is removed from the $\sum x$, the value of the mean will decrease. Think of individual scores as pulling the mean up or down in their direction. Thus, when a high score is added, it is pulling up the mean. Yet when a high score is removed, it can no longer pull up the mean, so the mean falls back down toward the lower scores. Thus, when a lower score is added to the $\sum x$, it pulls the mean down (the mean decreases). When a lower score is subtracted, the mean is no longer being pulled down, and so the mean increases.

Skewed distributions: A positively skewed distribution will have a few values that pull the tail of a distribution toward larger values, but the majority of the scores in the body will be in the middle or lower portion of the distribution. Thus, a positively skewed distribution will have one tail that is skewed toward the larger values. A negatively skewed distribution has a few values that pull the distribution toward smaller values, but the majority of the scores in the body will be in the middle or upper portion of the distribution. Thus, a distribution with a negative skew will have one tail that is skewed toward the smaller values.
<table>
<thead>
<tr>
<th>KEY TERM WORD SEARCHES</th>
</tr>
</thead>
<tbody>
<tr>
<td>POPULATION MEAN</td>
</tr>
<tr>
<td>MODE</td>
</tr>
<tr>
<td>SAMPLE MEAN</td>
</tr>
<tr>
<td>SYMMETRICAL DISTRIBUTION</td>
</tr>
<tr>
<td>NORMAL DISTRIBUTION</td>
</tr>
<tr>
<td>CENTRAL TENDENCY</td>
</tr>
<tr>
<td>SKEWED DISTRIBUTION</td>
</tr>
<tr>
<td>MEAN</td>
</tr>
<tr>
<td>POSITIVELY SKewed</td>
</tr>
<tr>
<td>ARITHMETIC MEAN</td>
</tr>
<tr>
<td>NEGATIVELY SKewed</td>
</tr>
<tr>
<td>AVERAGE</td>
</tr>
<tr>
<td>MODAL DISTRIBUTION</td>
</tr>
<tr>
<td>MULTIMODAL DISTRIBUTION</td>
</tr>
<tr>
<td>UNIMODAL DISTRIBUTION</td>
</tr>
<tr>
<td>WEIGHTED MEAN</td>
</tr>
<tr>
<td>BIMODAL DISTRIBUTION</td>
</tr>
<tr>
<td>SAMPLE SIZE</td>
</tr>
<tr>
<td>MEDIAN</td>
</tr>
<tr>
<td>RECTANGULAR DISTRIBUTION</td>
</tr>
<tr>
<td>POPULATION SIZE</td>
</tr>
<tr>
<td>GAUSSIAN DISTRIBUTION</td>
</tr>
<tr>
<td>BELL-SHAPED DISTRIBUTION</td>
</tr>
</tbody>
</table>
ACROSS
1 A distribution of scores, where one score occurs most often or most frequently.
3 The number of individuals that constitute an entire group or population.
5 A distribution of scores where a few outliers are substantially larger (toward the right tail of a graph) than most other scores.
6 A distribution of scores that includes outliers or scores that fall substantially above or below most other scores in a data set.
9 The value in a data set that occurs most often or most frequently.
10 Refers to statistical measures for locating a single score that is most representative or descriptive of all scores in a distribution.
11 The middle value of a distribution of data listed in numeric order.
13 The combined mean of two or more groups of scores, where the number of scores in each group is disproportionate or unequal.
15 The mean for a set of scores in an entire population.
16 A distribution of scores, where one or more scores occur most often or most frequently.
17 A distribution of scores, where more than two scores occur most often or most frequently.
18 A theoretical distribution with data that are symmetrically distributed around the mean, the median, and the mode.

DOWN
2 A distribution of scores where a few outliers are substantially smaller (toward the left tail of a graph) than most other scores.
4 The number of individuals that constitute a subset of those selected from a larger population.
7 A distribution of scores, where two scores occur most often or most frequently.
8 The sum of a set of scores in a distribution, divided by the total number of scores summed.
12 A distribution of scores, where all scores occur at the same frequency is called a ________ modal distribution (fill in the blank).
14 The mean for a sample or a subset of scores from a population.
LO 1: Distinguish between a population mean and sample mean.

1. Although different symbols are used to represent the number of scores in a sample and population, the computation of central tendency is ________ for samples and populations.
   a. different  
   b. the same  
   c. minimal  
   d. summed

2. The ________ is the sum of $N$ scores divided by $N$, whereas the ________ is the sum of $n$ scores divided by $n$.
   a. population mean; sample mean  
   b. sample mean; population mean

3. The population mean is identified by the Greek letter ________; the sample mean is identified by an italicized ________.
   a. $\mu$; $\Sigma$  
   b. $M$; $\mu$  
   c. $\mu$; $M$  
   d. $\Sigma$; $M$

LO 2: Calculate and interpret the mean, the median, and the mode.

4. The mean is the balance point of a distribution, which means that it is:
   a. always equal to 0  
   b. not always located at the center of a distribution of scores  
   c. equal to the score in the middle of a distribution of scores  
   d. used to measure the extent to which objects balance

5. A researcher records the following number of mistakes made during a sports broadcast: 0, 2, 0, 5, 2, 3, 0, 8, 1, and 4. What is the mean for these data?
   a. 1.5  
   b. 2.0  
   c. 2.5  
   d. 3.0
6. The median is:
   a. the preferred measure of central tendency when outliers are in a data set
   b. the middle value in a distribution listed in numerical order
   c. at the 50th percentile of a cumulative percent distribution
   d. all of the above

7. A researcher records the following time (in seconds) that children in a sample play with an unfamiliar toy: 12, 14, 10, 6, 8, 10, 13, 12, 4, 12, and 6. Which measure of central tendency (the mean, the median, or the mode) is largest?
   a. mean
   b. median
   c. mode
   d. all measures for central tendency are equal

8. The ______ is the value in a data set that occurs most often or most frequently.
   a. mode
   b. median
   c. mean

LO 3: Calculate and interpret the weighted mean for two or more samples with unequal sample sizes.

9. When a weighted mean is computed for two or more samples with unequal sample sizes, which value is used as the weight in the formula?
   a. \( M \)
   b. \( n \)
   c. \( \Sigma x \)
   d. \( \mu \)

10. A researcher records \( M = 12 \) in a sample of men \((n = 10)\) and \( M = 8 \) in a sample of women \((n = 20)\). What is the weighted mean for these samples?
    a. 12.0
    b. 10.33
    c. 10.0
    d. 9.33

11. The weighted mean is used to compute the mean for samples with ______ sizes.
    a. unequal
    b. equal
    c. proportional
    d. fractional
LO 4: Identify the characteristics of the mean.

12. The mean for a distribution of scores is $M = 6$. If we measure a new score equal to 10, then what will happen to the mean?
   a. the mean will increase
   b. the mean will decrease
   c. the mean will remain the same

13. The mean for a distribution of scores is $M = 3.2$. If we remove an existing score equal to 3.2, then what will happen to the mean?
   a. the mean will increase
   b. the mean will decrease
   c. the mean will remain the same

14. Which of the following will decrease the value of the mean for a distribution of scores?
   a. adding a new score above the mean
   b. deleting an existing score above the mean
   c. adding a new score equal to the mean
   d. deleting an existing score below to the mean

15. A researcher measures a sample mean for five samples. The sample mean in each sample is $M = 2$, $M = 4$, $M = 5$, $M = 7$, and $M = 9$. What will the sum of the differences of scores from the mean equal in each sample?
   a. minimal
   b. equal to 0
   c. larger as the mean in the sample gets larger
   d. not enough information because the actual scores in each sample are not given

16. The sum of the squared differences of scores from their mean:
   a. is minimal
   b. is equal to 0
   c. can be any positive or negative number

LO 5: Identify an appropriate measure of central tendency for different distributions and scales of measurement.

17. The mean is used to describe:
   a. data that are normally distributed
   b. interval scale data
   c. ratio scale data
   d. all of the above
18. Data are distributed such that the mean, median, and mode are equal. Which measure of central tendency is most appropriate for these data?
   a. the mean
   b. the median
   c. the mode

19. When outliers exist in a data set, which measure of central tendency is most appropriate?
   a. the mean
   b. the median
   c. the mode

20. The median can be used for skewed distributions and ordinal data. Why is the median appropriate for describing ordinal data?
   a. ordinal data are the only type of data that have a median
   b. ordinal data are always positively or negatively skewed
   c. ordinal data convey direction only (more or less than)
   d. ordinal data convey the same information as ratio data

21. The mode is the primary measure of central tendency to describe data on which scale of measurement?
   a. nominal
   b. ordinal
   c. interval
   d. ratio
SPSS IN FOCUS

Mean, Median, and Mode

Follow the General Instructions Guide to complete this exercise. Also, an example for following these steps is provided in the SPSS in Focus section (Section 3.6) of the book. Complete and submit the SPSS grading template and a printout of the output file.

Exercise 3.1: Alcohol Consumption at Local Parties I

A researcher wants to study alcohol consumption at local parties. The researcher is particularly interested in the amount of alcohol consumed at parties where untrained friends or acquaintances mix the drinks (instead of trained bartenders). For this study, a standard drink was defined as 12 oz of beer (5% vol.), 5 oz of wine (12% vol.), 1.5 oz of liquor (40% vol.), or 1.5 oz of liquor in a mixed drink (National Institute of Alcohol Abuse and Alcoholism, 2000). Below is the number of standard drinks consumed by 30 people at one local party. Compute the mean, the median, and the mode for these data.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4</td>
<td>6.6</td>
<td>8.2</td>
</tr>
<tr>
<td>2.3</td>
<td>0.5</td>
<td>3.2</td>
</tr>
<tr>
<td>1.4</td>
<td>14.2</td>
<td>3.5</td>
</tr>
<tr>
<td>5.4</td>
<td>9.3</td>
<td>6.4</td>
</tr>
<tr>
<td>7.5</td>
<td>2.0</td>
<td>7.9</td>
</tr>
<tr>
<td>8.2</td>
<td>4.4</td>
<td>0.9</td>
</tr>
<tr>
<td>12.1</td>
<td>10.0</td>
<td>5.4</td>
</tr>
<tr>
<td>3.2</td>
<td>2.8</td>
<td>5.4</td>
</tr>
<tr>
<td>1.1</td>
<td>5.6</td>
<td>6.3</td>
</tr>
<tr>
<td>2.0</td>
<td>5.0</td>
<td>8.8</td>
</tr>
</tbody>
</table>

Chapter 3  Summarizing Data: Central Tendency

With regard to the SPSS exercise, answer the following questions:

State the dependent variable: ________________________

State the following values (from the SPSS output table):

Mean ____________________
Median ____________________
Mode(s) ____________________

Is there more than one mode in this distribution? How does SPSS display or indicate that the distribution of scores has more than one mode?

Interpret the data displayed in the SPSS output table. Which measure of central tendency is the most appropriate statistic to summarize these data?
CHAPTER SUMMARY ORGANIZED BY LEARNING OBJECTIVE

LO 1: Distinguish between a population mean and a sample mean.

- The mean for a set of scores in an entire population is called a **population mean**; the mean for a sample (or subset of scores from a population) is called a **sample mean**. Each mean is computed the same but with different notation used to identify the sample size ($n$) and population size ($N$).

LO 2: Calculate and interpret the mean, the median, and the mode.

- Measures of **central tendency** are statistical measures for locating a single score that is most representative or descriptive of all scores in a distribution. Three measures of central tendency are the mean, the median, and the mode.

  - The **mean** is the sum of a set of scores divided by the total number of scores summed:

    \[ \mu = \frac{\sum x}{N} \]

    The population mean is the sum of $N$ scores divided by $N$.

    \[ M = \frac{\sum x}{n} \]

    The sample mean is the sum of $n$ scores divided by $n$.

  - The **median** is the middle score in a data set that is listed in numerical order, where half of all scores fall above and half fall below its value. Unlike the mean, the value of the median is not shifted in the direction of outliers—the median is always at the center or midpoint of a data set. To find the median position, list scores in numerical order and apply this formula:

    \[ \text{Median position} = \frac{n + 1}{2} \]

  - The **mode** is the value in a data set that occurs most often or most frequently. The mode is often reported with the mean or the median.

LO 3: Calculate and interpret the weighted mean for two or more samples with unequal sample sizes.

- The **weighted mean** is the combined mean of two or more groups of scores, where the number of scores in each group is disproportionate or unequal. The formula for the weighted mean of two or more samples with unequal sample sizes is

  \[ M_w = \frac{\sum (M \times n)}{\sum n} \]

LO 4: Identify the characteristics of the mean.

- The mean has the following characteristics:

  a. Changing an existing score will change the mean.

  b. Adding a new score or completely removing an existing score will change the mean, unless that value equals the mean.

  c. Adding, subtracting, multiplying, or dividing each score in a distribution by a constant will cause the mean to change by that constant.

  d. The sum of the differences of scores from their mean is zero.
e. The sum of the squared differences of scores from their mean is minimal.

LO 5: Identify an appropriate measure of central tendency for different distributions and scales of measurement.

- The mean is used to describe (1) data that are normally distributed and (2) interval and ratio scale data.
- The median is used to describe (1) data in a skewed distribution and (2) ordinal scale data.
- The mode is used to describe (1) any type of data with a value that occurs the most, although it is typically used together with the mean or the median, and (2) nominal scale data.

SPSS LO 6: Compute the mean, the median, and the mode using SPSS.

- SPSS can be used to compute the mean, median, and mode. Each measure of central tendency is computed using the Analyze, then Descriptive Statistics and Frequencies options in the menu bar. These actions will bring up a dialog box that will allow you to identify the variable and select the Statistics option to select and compute the mean, the median, and the mode (for more details, see Section 3.6).