



Figure 4.1.7

4.1 Exercises

Basic Concepts

1. Describe the difference between statistics and parameters.
2. Describe three major attributes used in summarizing a data set.
3. What are numerical descriptive statistics and why are they important?
4. Identify and describe five measures of location. List the advantages and disadvantages of each.
5. What is a resistant measure?
6. Describe a situation in which using the weighted mean as a measure of location would be appropriate.
7. What does it mean if we say that a data set is positively skewed? Negatively skewed?
8. Explain why the mean should not be calculated for a nonstationary time series.
9. What is a moving average? When is it useful?
10. When you compute the mean of a data set are you thinking inductively or deductively? How about the median, mode, and trimmed mean?
11. How does determining a statistic relate to empiricism?

Exercises

12. Calculate the mean, median, 10% trimmed mean, and the mode for the following data.

25	29	31	38	39	40	41	44	46	47
50	53	55	57	62	62	67	71	74	78

13. Using the US Violent Crime data set found on the companion website, determine the mean, median, mode, and 20% trimmed mean for the year 2019. Round your answers to 2 decimal places.
14. Use the US Violent Crime rate data set from the previous problem.
- Remove Washington, D.C. (District of Columbia) from the data and then determine the mean, median, mode, and the 20% trimmed mean for the year 2019. Round your answers to 2 decimal places.
 - Compare your results with those in exercise #13. Order the four measures of center from the measure with the least amount of change to the one with the greatest amount of change. Which measure of center was the most sensitive to the outlier (Washington, D.C.)?
15. Determine the mean, median, the 10% trimmed mean, and the mode for the following data.

2	22	6	18	10	14	12	12	16	8
---	----	---	----	----	----	----	----	----	---

16. Discuss the usefulness of each of the measures of central tendency with respect to the following situations.
- A company is considering a move into a regional market for specialty soft drinks. In analyzing the size of the containers that its competitors are currently offering, would the company be more interested in the mean, median, or mode of their containers?
 - The creative director for an advertising agency is trying to target an ad campaign that will be shown in one city only. Would he be more interested in the mean or median family income in the city?
 - A young economist was assigned the task of comparing the interest rates on ninety-day certificates of deposit (CDs) in three major cities. Should she compare the mean, median, or modal interest for the banks in the three cities?
 - A telephone company is interested in knowing how customers rate their service: excellent, good, average, or poor. Would the company be more interested in studying the mean, median, or mode of the customer service ratings?

Data

stat.hawkeslearning.com

Discovering Statistics and Data,
Fourth Edition > Data Sets > US
Violent Crime 2020

17. Discuss the usefulness of each of the measures of central tendency with respect to the following situations.
- A doctor is interested in analyzing the increase in systolic blood pressure caused by a certain antibiotic. Would the doctor be more interested in studying the mean, median or mode of the systolic blood pressures?
 - A car manufacturer is trying to decide in what colors it should offer its new sports coupe. In analyzing the preferred colors of other sports coupes, would the manufacturer be more interested in the mean, median, or mode of the colors?
 - A manufacturer of chocolate bars is interested in knowing how people rate its chocolate: the best, above average, average, below average, or the worst. Would the company be more interested in the mean, median, or mode of the ratings?
 - A realtor is interested in studying the prices of recent home sales in an area which has many diverse neighborhoods. Would the mean, median, or mode of the prices of recent home sales be the best measure of central tendency?
18. Using the Amazon Stock Price data set from the companion website, perform the following calculations. Round your answers to 2 decimal places.
- Compute and graph a 3-day moving average of the daily closing stock price for the years 1997 to 2017.
 - Describe any trends that you see in the data from the graph in part **a**.
19. A stroke patient has been monitoring her blood pressure daily for 20 days since she was discharged from the hospital. The systolic blood pressure readings are recorded below.

Systolic Blood Pressure Readings									
100	104	117	95	98	111	105	106	115	101
101	102	115	116	113	103	104	119	127	132

- What level of measurement does the data possess?
 - Compute the mean, 10% trimmed mean, and the 20% trimmed mean.
 - Which measure computed in part **b**. best describes the typical systolic blood pressure of the patient?
 - Would conclusions drawn from this data be an example of empiricism or rationalism?
20. Using the CO₂ Emissions data set from the companion website, answer the following questions.
- What level of measurement does the data possess?
 - For the United States, create a time series plot of CO₂ emissions with year on the horizontal axis. Does the data have trend?
 - How should we summarize this data, as a time series or with traditional summary measures of central tendency? Explain.
 - Would conclusions drawn from this data be an example of empiricism or rationalism?

Data

stat.hawkeslearning.com
Discovering Statistics and Data, Fourth Edition > Data Sets > Amazon Stock Price

Technology

To learn how to filter data, please visit stat.hawkeslearning.com and navigate to **Discovering Statistics and Data, Fourth Edition > Technology Instructions > Data Manipulation > Filtering**.

Data

stat.hawkeslearning.com
Discovering Statistics and Data, Fourth Edition > Data Sets > CO₂ Emissions



21. Consider the following monthly sales for a small clothing store in a resort community.

Monthly Sales			
Month	Sales (\$)	Month	Sales (\$)
January	100,500	July	200,000
February	120,000	August	185,000
March	133,000	September	175,000
April	145,000	October	120,000
May	160,000	November	180,000
June	180,000	December	330,000

- Draw a line graph of the data.
 - Compute the two-period moving averages for the data.
 - Compute the three-period moving averages for the data.
 - Add line graphs for the two-period moving averages and three-period moving averages to the graph which you constructed in part **a**.
 - Which series of data (the original sales data, the two-period moving averages, or the three-period moving averages) do you think best represents sales for the year? Why?
22. Consider the following average monthly balances for one bank customer for January through March. Compute the weighted average balance for the three-month period. Note that each average monthly balance must be weighted by the number of days in that month on a non-leap year. Round to the nearest cent.

Average Monthly Balances for a Bank Customer (January through March)	
Month	Average Monthly Balance
January	\$1885.67
February	\$1312.92
March	\$2001.53

23. Using the US County Data set on the companion website, determine the percentage of the total US population that has at least a high school diploma by utilizing the weighted mean. Use the *At.Least.High.School.Diploma* variable, which is the percentage of a county population with at least a high school diploma, as the data values, and use the *Total.Population* variable as the weights.

Data

stat.hawkeslearning.com
Discovering Statistics and Data,
Fourth Edition > Data Sets > US
County Data

24. The following data represent the blood types of 20 patients on the surgical ward at a local hospital.

Blood Type of Surgical Patients			
Patient	Blood Type	Patient	Blood Type
1	O+	11	A+
2	O+	12	O-
3	A+	13	A-
4	B+	14	O+
5	O+	15	A+
6	A+	16	B-
7	B+	17	O+
8	A+	18	A+
9	AB+	19	O+
10	O+	20	O+

- What level of measurement does the data possess?
- What is the modal blood type of patients on the surgical ward?

4.2 Measures of Dispersion

Suppose all people looked alike, all cars looked alike, everyone wore the same kind of clothes, and there was only one kind of hamburger (plain). Without diversity it would be a boring world and a world in which statistics would be of little value. Since much of statistics is devoted to describing, analyzing, and explaining variability, understanding how variability is measured is essential to understanding statistics.

The concept of **variability** (also referred to as **dispersion** or **spread**) is as vague as the concept of central tendency. And vague concepts lead to different measurement ideas. The same issues that are important in evaluating location measures are meaningful in evaluating measures of dispersion.

Many of the good measures of dispersion use the concept of deviation from the mean. If the mean is a focal point or base, use it as a common basis from which to measure variation. The distance that a point is from its mean is called a **deviation from the mean**. A data set and its deviations from the mean are calculated in Table 4.2.1.

Table 4.2.1 - Calculating Deviations from the Mean	
Mean = 10	
Data Values	Deviations from the Mean (Data - Mean = Deviation)
3	$3 - 10 = -7$
12	$12 - 10 = 2$
20	$20 - 10 = 10$
15	$15 - 10 = 5$
0	$0 - 10 = -10$