Chapter 2: Descriptive Statistics

1. _____ provide facts and figures that can be used for analysis and interpretation of a population of interest.
   a. Data
   b. Variables
   c. Range
   d. Query

   Answer: A
   Difficulty: Easy
   LO: 2.1, Page 16
   Bloom's: Knowledge
   BUSPROG: Analytic Skills
   DISC: Descriptive Statistics
   Feedback: Data are the facts and figures collected, analyzed, and summarized for presentation and interpretation.

2. A variable is defined as a
   a. quantity of interest that can take on same values.
   b. set of values.
   c. quantity of interest that can take on different values.
   d. characteristic that takes on same values from a set of values.

   Answer: C
   Difficulty: Easy
   LO: 2.1, Page 16
   Bloom's: Knowledge
   BUSPROG: Analytic Skills
   DISC: Descriptive Statistics
   Feedback: A characteristic or a quantity of interest that can take on different values is known as a variable.

3. A set of values corresponding to a set of variables is defined as a(n) _____.
   a. quantity
   b. event
   c. factor
   d. observation

   Answer: D
   Difficulty: Easy
   LO: 2.1, Page 16
   Bloom's: Knowledge
   BUSPROG: Analytic Skills
   DISC: Descriptive Statistics

Feedback: An observation is a set of values corresponding to a set of variables.

4. The difference in a variable measured over observations (time, customers, items, etc.) is called as _____.
   a. observed differences
   b. variation
   c. variable change
   d. descriptive analytics

   Answer: B
   Difficulty: Moderate
   LO: 2.1, Page 16
   Bloom's: Knowledge
   BUSPROG: Analytic Skills
   DISC: Descriptive Statistics
   Feedback: Variation is the difference in a variable measured over observations (time, customers, items, etc.).

5. A variable whose values are not known with certainty is called a _____.
   a. certain variable
   b. random variable
   c. constant variable
   d. decision variable

   Answer: B
   Difficulty: Moderate
   LO: 2.1, Page 17
   Bloom's: Knowledge
   BUSPROG: Analytic Skills
   DISC: Descriptive Statistics
   Feedback: A quantity whose values are not known with certainty is called a random variable, or uncertain variable.

6. _____ act(s) as a representative of the population.
   a. The analytics
   b. The variance
   c. A sample
   d. The random variables

   Answer: C
   Difficulty: Easy
   LO: 2.2, Page 17
   Bloom's: Knowledge
   BUSPROG: Analytic Skills
   DISC: Descriptive Statistics

Feedback: A subset of the population is known as a sample, and it acts as a representative of the population.

7. The act of collecting data that are representative of the population data is called
    a. random sampling.
    b. sample data.
    c. population sampling.
    d. applications of business analytics.

    Answer: A
    Difficulty: Easy
    LO: 2.2, Page 18
    Bloom's: Knowledge
    BUSPROG: Analytic Skills
    DISC: Descriptive Statistics
    Feedback: A representative sample can be gathered by random sampling of the population data.

8. The data on grades (A, B, C, and D) scored by all students in a test is an example of
    a. quantitative data.
    b. sample data.
    c. categorical data.
    d. analytical data.

    Answer: C
    Difficulty: Easy
    LO: 2.2, Page 18
    Bloom's: Knowledge
    BUSPROG: Analytic Skills
    DISC: Descriptive Statistics
    Feedback: If arithmetic operations cannot be performed on the data, they are considered categorical data.

9. The data on the time taken by 10 students in a class to answer a test is an example of
    a. population data.
    b. categorical data.
    c. time series data.
    d. quantitative data.

    Answer: D
    Difficulty: Easy
    LO: 2.2, Page 18
    Bloom's: Knowledge
    BUSPROG: Analytic Skills
    DISC: Descriptive Statistics

Feedback: Data are considered quantitative data if numeric and arithmetic operations, such as addition, subtraction, multiplication, and division, can be performed on them.

10. _____ are collected from several entities at the same point in time.
    a. Time series data
    b. Categorical and quantitative data
    c. Cross-sectional data
    d. Random data

    Answer: C
    Difficulty: Moderate
    LO: 2.2, Page 18
    Bloom's: Knowledge
    BUSPROG: Analytic Skills
    DISC: Descriptive Statistics
    Feedback: Cross-sectional data are collected from several entities at the same, or approximately the same, point in time.

11. Data collected from several entities over several time periods is
    a. categorical and quantitative data.
    b. time series data.
    c. source data.
    d. cross-sectional data.

    Answer: B
    Difficulty: Easy
    LO: 2.2, Page 18
    Bloom's: Knowledge
    BUSPROG: Analytic Skills
    DISC: Descriptive Statistics
    Feedback: Time series data are collected over several time periods.

12. In a(n) _____, one or more variables are identified and controlled or manipulated so that data can be obtained about how they influence the variable of interest identified first.
    a. experimental study
    b. observational study
    c. categorical study
    d. variable study

    Answer: A
    Difficulty: Easy
    LO: 2.2, Page 18
    Bloom's: Knowledge
    BUSPROG: Analytic Skills
    DISC: Descriptive Statistics

Feedback: In an experimental study, a variable of interest is first identified. Then one or more other variables are identified and controlled or manipulated so that data can be obtained about how they influence the variable of interest.

13. The data collected from the customers in restaurants about the quality of food is an example of
    a. variable study.
    b. cross-sectional study.
    c. experimental study.
    d. observational study.

    Answer: D
    Difficulty: Moderate
    LO: 2.2, Page 19
    Bloom's: Knowledge
    BUSPROG: Analytic Skills
    DISC: Descriptive Statistics
    Feedback: Nonexperimental, or observational, studies make no attempt to control the variables of interest. Some restaurants use observational studies to obtain data about customer opinions on the quality of food, quality of service, atmosphere, and so on.

14. When the data are large and when it is difficult to analyze all at once, which of the following feature in Excel is used to make the data more manageable and to develop insights?
    a. Frequency table
    b. Sorting and filtering
    c. Fill color
    d. Charts

    Answer: B
    Difficulty: Easy
    LO: 2.3, Page 21
    Bloom's: Knowledge
    BUSPROG: Analytic Skills
    DISC: Descriptive Statistics
    Feedback: Excel contains option to sort and filter data so that one can identify patterns of the data more easily.

15. A summary of data that shows the number of observations in each of several nonoverlapping bins is called
    a. a frequency distribution.
    b. a sample summary.
    c. a bin distribution.
    d. an observed distribution.

    Answer: A
    Difficulty: Easy
    LO: 2.4, Page 25

Bloom's: Knowledge
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: A frequency distribution is a summary of data that shows the number (frequency) of observations in each of several nonoverlapping classes, typically referred to as bins, when dealing with distributions.

16. Which of the following gives the proportion of items in each bin?
    a. Frequency
    b. Percent frequency
    c. Relative frequency
    d. Bin proportion

Answer: C
Difficulty: Easy
LO: 2.4, Page 27
Bloom's: Knowledge
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: The relative frequency of a bin equals the fraction or proportion of items belonging to a class.

17. Compute the relative frequencies for the data given in the table below:

Grades	Number of students
A	16
B	28
C	33
D	13
Total	90

    a. 0.31, 0.14, 0.37, 0.18
    b. 0.37, 0.14, 0.31, 0.18
    c. 0.14, 0.31, 0.37, 0.18
    d. 0.18, 0.31, 0.37, 0.14

Answer: D
Difficulty: Moderate
LO: 2.4, Page 27
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: The relative frequency of a bin equals the fraction or proportion of items belonging to a class. Relative frequency of a bin = Frequency of the bin /n.

18. Consider the data below. What percentage of students scored grade C?

Grades	Number of students
A	16
B	28
C	33
D	13
Total	90

a. 33%

b. 31%

c. 37%

d. 28%

Answer: C
Difficulty: Moderate
LO: 2.4, Page 27
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: A percent frequency distribution summarizes the percent frequency of the data for each bin. The percent frequency of a bin is the relative frequency multiplied by 100.

19. Which of the following are necessary to be determined to define the classes for a frequency distribution with quantitative data?
   a. Number of nonoverlapping bins, width of each bin, and bin limits
   b. Width of each bin and bin lower limits
   c. Number of overlapping bins, width of each bin, and bin upper limits
   d. Width of each bin and number of bins

Answer: A
Difficulty: Moderate
LO: 2.4, Page 28
Bloom's: Knowledge
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: The three steps necessary to define the classes for a frequency distribution with quantitative data are: determine the number of nonoverlapping bins, determine the width of each bin, and determine the bin limits.

20. The purpose of using enough bins is to show the
   a. number of observations.
   b. number of variables.
   c. variation in the data.

d. correlation in the data.

21. _____ is a graphical summary of data previously summarized in a frequency distribution.
    a. Box plot
    b. Histogram
    c. Line chart
    d. Scatter chart

22. Identify the shape of the distribution in the below figure.
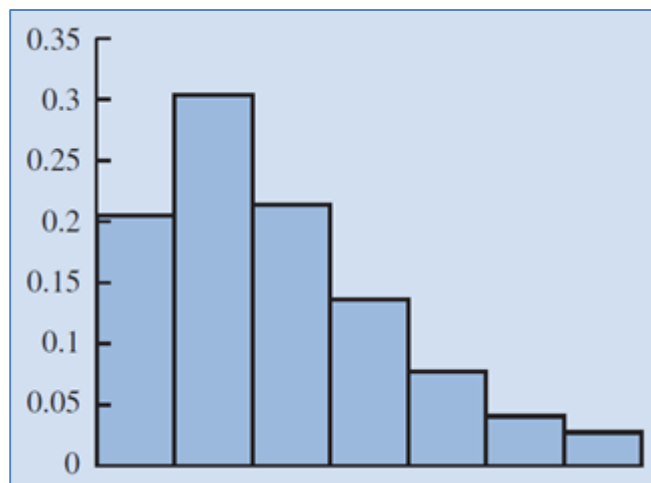
    a. Moderately skewed left
    b. Symmetric

c. Highly skewed right

d. Moderately skewed right

Answer: D
Difficulty: Moderate
LO: 2.4, Pages 33-34
Bloom's: Knowledge
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: A histogram is said to be skewed to the right if its tail extends farther to the right than to the left. The given histogram is, therefore, moderately skewed to the right.

23. The _____ shows the number of data items with values less than or equal to the upper class limit of each class.

a. cumulative frequency distribution

b. frequency distribution

c. percent frequency distribution

d. relative frequency distribution

Answer: A
Difficulty: Easy
LO: 2.4, Page 34
Bloom's: Knowledge
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: The cumulative frequency distribution shows the number of data items with values less than or equal to the upper class limit of each class.

24. The _____ is a point estimate of the population mean for the variable of interest.

a. sample mean

b. median

c. Sample

d. geometric mean

Answer: A
Difficulty: Moderate
LO: 2.5, Page 35
Bloom's: Knowledge
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: The sample mean is a point estimate of the (typically unknown) population mean for the variable of interest.

25. Compute the mean of the following data:

| 56 | 42 | 37 | 29 | 45 | 51 | 30 | 25 | 34 | 57 |

a. 42.8

b. 52.1

c. 40.6

d. 39.4

Answer: C
Difficulty: Moderate
LO: 2.5, Page 35
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: The mean provides a measure of central location for the data. It is computed as:

Mean = $\frac{56+42+37+29+45+51+30+25+34+57}{10} = \frac{406}{10} = 40.6$.

26. Compute the median of the following data:

| 32 | 41 | 36 | 24 | 29 | 30 | 40 | 22 | 25 | 37 |

a. 28

b. 31

c. 40

d. 34

Answer: B
Difficulty: Moderate
LO: 2.5, Pages 36-37
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: The median is the value in the middle when the data are arranged in ascending order

(smallest to largest value). Computed as: Median = average of middle two values = $\frac{30+32}{2}$ = 31.

27. Compute the mode for the following data:

| 12 | 16 | 19 | 10 | 12 | 11 | 21 | 12 | 21 | 10 |

a. 21

b. 11

c. 12

d. 10

Answer: C
Difficulty: Moderate
LO: 2.5, Page 37
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: The mode is the value that occurs most frequently in a data set. The value 12 occurs

with the greatest frequency of 3. Therefore, the mode is 12.

28. Compute the geometric mean for the following data on growth factors of an investment for 10 years:

1.10   0.50   0.70   1.21   1.25   1.12   1.16   1.11   1.13   1.22

a. 1.0221
b. 1.0148
c. 1.0363
d. 1.1475

Answer: B
Difficulty: Moderate
LO: 2.5, Pages 38-39
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: The geometric mean is a measure of location that is calculated by finding the nth root of the product of n values. Geometric mean =
$\sqrt[10]{(1.1)(0.5)(0.7)(1.21)(1.25)(1.12)(1.16)(1.11)(1.13)(1.22)}$ = 1.0148.

29. The simplest measure of variability is the

a. variance.
b. standard deviation.
c. coefficient of variation.
d. range.

Answer: D
Difficulty: Easy
LO: 2.6, Page 41
Bloom's: Knowledge
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: The simplest measure of variability is the range.

30. The variance is based on the

a. deviation about the median.
b. number of variables.
c. deviation about the mean.
d. correlation in the data.

Answer: C
Difficulty: Easy
LO: 2.6, Page 41
Bloom's: Knowledge
BUSPROG: Analytic Skills
DISC: Descriptive Statistics

Feedback: The variance is based on the deviation about the mean, which is the difference between the value of each observation (x_i) and the mean.

31. For the following sample data, compute the variance.

| 32 | 41 | 36 | 24 | 29 | 30 | 40 | 22 | 25 | 37 |

    a. 45.6
    b. 35.5
    c. 41.04
    d. 29.4

    Answer: A
    Difficulty: Moderate
    LO: 2.6, Page 42
    Bloom's: Application
    BUSPROG: Analytic Skills
    DISC: Descriptive Statistics
    Feedback: The variance is based on the deviation about the mean, which is the difference between the value of each observation (x_i) and the mean.

    It is computed as, $s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$ = 410.4/9 = 45.6.

32. Compute the standard deviation for the following sample data.

| 32 | 41 | 36 | 24 | 29 | 30 | 40 | 22 | 25 | 37 |

    a. 5.96
    b. 6.41
    c. 5.42
    d. 6.75

    Answer: D
    Difficulty: Moderate
    LO: 2.6, Page 43
    Bloom's: Application
    BUSPROG: Analytic Skills
    DISC: Descriptive Statistics
    Feedback: The standard deviation is defined to be the positive square root of the variance.

    It is computed as $s = \sqrt{s^2} = \sqrt{45.6}$ = 6.75.

33. Compute the coefficient of variation for the following sample data.

| 32 | 41 | 36 | 24 | 29 | 30 | 40 | 22 | 25 | 37 |

    a. 18.64 percent
    b. 21.36 percent
    c. 20.28 percent
    d. 21.67 percent

    Answer: B

34. Compute the 50th percentile for the following data:

| 10 | 15 | 17 | 21 | 25 | 12 | 16 | 11 | 13 | 22 |

   a. 18.6
   b. 13.3
   c. 15.5
   d. 17.7

35. Compute the third quartile for the following data.

| 10 | 15 | 17 | 21 | 25 | 12 | 16 | 11 | 13 | 22 |

   a. 21.25
   b. 15.5
   c. 21.5
   d. 11.75

36. Compute IQR for the following data.

| 10 | 15 | 17 | 21 | 25 | 12 | 16 | 11 | 13 | 22 |

   a. 6.25
   b. 7.75
   c. 5.14

d. 9.50

Answer: D
Difficulty: Moderate
LO: 2.7, Page 46
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: The difference between the third and first quartiles is often referred to as the interquartile range, or IQR. IQR = 21.25 − 11.75 = 9.50.

37. A _____ determines how far a particular value is from the mean relative to the data set's standard deviation.

    a. coefficient of variation
    b. *z*-score
    c. variance
    d. percentile

    Answer: B
    Difficulty: Moderate
    LO: 2.7, Page 46
    Bloom's: Application
    BUSPROG: Analytic Skills
    DISC: Descriptive Statistics
    Feedback: A *z*-score helps us determine how far a particular value is from the mean relative to the data set's standard deviation.

38. For data having a bell-shaped distribution, approximately _____ percent of the data values will be within one standard deviation of the mean.

    a. 95
    b. 66
    c. 68
    d. 97

    Answer: C
    Difficulty: Easy
    LO: 2.7, Page 48
    Bloom's: Knowledge
    BUSPROG: Analytic Skills
    DISC: Descriptive Statistics
    Feedback: Approximately 68 percent of the data values will be within one standard deviation of the mean for data having a bell-shaped distribution.

39. Any data value with a *z*-score less than –3 or greater than +3 is treated as a(n)
    a. outlier.
    b. usual value.

c. whisker.

d. *z*-score value.

Answer: A
Difficulty: Easy
LO: 2.7, Page 49
Bloom's: Knowledge
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: Any data value with a *z*-score less than −3 or greater than +3 is treated as an outlier.

40. Which of the following graphs provide information on outliers and IQR of a data set?

a. Histogram

b. Line chart

c. Scatter chart

d. Box plot

Answer: D
Difficulty: Easy
LO: 2.7, Page 49
Bloom's: Knowledge
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: A box plot is a graphical summary of the distribution of data and it is developed from the quartiles for a data set. Therefore, the information on the outliers and IQR can be obtained from a box plot.

41. If covariance between two variables is near 0, then it implies that

a. there exists a positive relationship between the variables.

b. the variables are not linearly related.

c. the variables are negatively related.

d. the variables are strongly related.

Answer: B
Difficulty: Easy
LO: 2.8, Page 53
Bloom's: Knowledge
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: If the covariance between two variables is near 0, then the variables are not linearly related.

42. The correlation coefficient will always take values

a. greater than 0.

b. between −1 and 0.

c. between −1 and +1.

d. less than −1.

Answer: C
Difficulty: Easy
LO: 2.8, Page 55
Bloom's: Knowledge
BUSPROG: Analytic Skills
DISC: Descriptive Statistics
Feedback: The correlation coefficient will always take values between –1 and +1.

Problems

1. A student willing to participate in a debate competition required to fill a registration form. State whether each of the following information about the participant provides categorical or quantitative data.
   a. What is your date of birth?
   b. Have you participated in any debate competition previously?
   c. If yes, how many debate competitions have you participated so far?
   d. Have you won any of the competitions?
   e. If yes, how many have you won?

   Answer:
   a. Quantitative.
   b. Categorical.
   c. Quantitative.
   d. Categorical.
   e. Quantitative.

   Difficulty: Easy
   LO: 2.2, Page 18
   Bloom's: Application
   BUSPROG: Analytic Skills
   DISC: Descriptive Statistics

2. The following table provides information on the number of billionaires in a country and the continents on which these countries are located.

Nationality	Continent	Number of Billionaires
United States	North America	426
Brazil	South America	38
Russia	Europe	105
Mexico	North America	37
India	Asia	54
Turkey	Europe	40
United Kingdom	Europe	31
Hong Kong	Asia	39

Germany	Europe	57
Canada	North America	28
China	Asia	120

a. Sort the countries from largest to smallest based on the number of billionaires. What are the top 5 countries according to the number of billionaires?

b. Filter the countries to display only the countries located in North America.

Answer:

a.

Nationality	Continent	Number of Billionaires
United States	North America	426
China	Asia	120
Russia	Europe	105
Germany	Europe	57
India	Asia	54
Turkey	Europe	40
Hong Kong	Asia	39
Brazil	South America	38
Mexico	North America	37
United Kingdom	Europe	31
Canada	North America	28

The top five countries with more number of billionaires are United States, China, Russia, Germany, and India.

b.

Nationality	Continent	Number of Billionaires
United States	North America	426
Mexico	North America	37
Canada	North America	28

Difficulty: Moderate
LO: 2.3, Pages 21-23
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics

3. The data on the percentage of visitors in the previous and current years at 12 well-known national parks of Unites States are given below:

National Parks	Percentage of visitors previous year	Percentage of visitors current year
The Smokies	78.2%	84.2%
The Grand Canyon	83.5%	81.6%
Theodore Roosevelt	81.6%	84.8%
Yosemite	74.2%	78.4%
Yellowstone	77.9%	76.2%
Olympic	86.4%	88.6%
The Colorado Rockies	84.3%	85.4%
Zion	76.7%	78.9%
The Grand Tetons	84.6%	87.8%
Cuyahoga Valley	85.1%	86.7%
Acadia	79.2%	82.6%
Shenandoah	72.9%	79.2%

a. Sort the parks in descending order by their current year's visitor percentage. Which park has the highest number of visitors in the current year? Which park has the lowest number of visitors in the current year?

b. Calculate the change in visitor percentage from the previous to the current year for each park. Use Excel's conditional formatting to highlight the park whose visitor percentage decreased from the previous year to the current year.

c. Use Excel's conditional formatting tool to create data bars for the change in visitor percentage from the previous year to the current year for each park calculated in part b.

Answer:

a. The sorted list of parks for the current year appears as below:

National Parks	Percentage of visitors previous year	Percentage of visitors current year
Olympic	86.4%	88.6%
The Grand Tetons	84.6%	87.8%
Cuyahoga Valley	85.1%	86.7%
The Colorado Rockies	84.3%	85.4%
Theodore Roosevelt	81.6%	84.8%
The Smokies	78.2%	84.2%
Acadia	79.2%	82.6%
The Grand Canyon	83.5%	81.6%
Shenandoah	72.9%	79.2%
Zion	76.7%	78.9%
Yosemite	74.2%	78.4%
Yellowstone	77.9%	76.2%

Olympic has the highest number of visitor's in the current year and Yellowstone has the lowest number of visitors in the current year.

b.

National Parks	Percentage of visitors previous year	Percentage of visitors current year	Change in visitor percentage
The Smokies	78.2%	84.2%	6.00%
The Grand Canyon	83.5%	81.6%	-1.90%
Theodore Roosevelt	81.6%	84.8%	3.20%
Yosemite	74.2%	78.4%	4.20%
Yellowstone	77.9%	76.2%	-1.70%
Olympic	86.4%	88.6%	2.20%
The Colorado Rockies	84.3%	85.4%	1.10%
Zion	76.7%	78.9%	2.20%
The Grand Tetons	84.6%	87.8%	3.20%
Cuyahoga Valley	85.1%	86.7%	1.60%
Acadia	79.2%	82.6%	3.40%
Shenandoah	72.9%	79.2%	6.30%

c. The output using Excel's conditional formatting tool that created data bars for the change in visitor percentage from the previous year to the current year for each park appears as below.

National Parks	Percentage of visitor's previous year	Percentage of visitor's current year	Change in visitor's percentage
The Smokies	78.2%	84.2%	6.00%
The Grand Canyon	83.5%	81.6%	-1.90%
Theodore Roosevelt	81.6%	84.8%	3.20%
Yosemite	74.2%	78.4%	4.20%
Yellowstone	77.9%	76.2%	-1.70%
Olympic	86.4%	88.6%	2.20%
The Colorado Rockies	84.3%	85.4%	1.10%
Zion	76.7%	78.9%	2.20%
The Grand Tetons	84.6%	87.8%	3.20%
Cuyahoga Valley	85.1%	86.7%	1.60%
Acadia	79.2%	82.6%	3.40%
Shenandoah	72.9%	79.2%	6.30%

Difficulty: Moderate
LO: 2.3, Pages 21-25
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics

4. The partial relative frequency distribution is given below:

Group	Relative Frequency
1	0.15

2	0.32
3	0.29
4	

a. What is the relative frequency of group 4?

b. The total sample size is 400. What is the frequency of group 4?

c. Show the frequency distribution.

d. Show the percent frequency distribution.

Answer:

a. The relative frequency of group 4 is obtained as 1.00 – 0.15 – 0.32 – 0.29 = 0.24.

b. If the total sample size is 400, the frequency of group 4 is obtained as 0.24 × 400 = 96.

c.

Group	Relative Frequency	Frequency
1	0.15	60
2	0.32	128
3	0.29	116
4	0.24	96
Total	1.00	400

d.

Group	Relative Frequency	% Frequency
1	0.15	15
2	0.32	32
3	0.29	29
4	0.24	24
Total	1.00	100

Difficulty: Moderate
LO: 2.4, Pages 25-28
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics

5. A survey on the most preferred newspaper in USA listed The New York Times(TNYT), Washington Post(WP), Daily News(DN), New York Post(NYP), and Los Angeles Times (LAT) as the top five most preferred newspapers. The table below shows the preferences of 50 citizens.

TNYT	WP	NYP	WP	TNYT
DN	TNYT	LAT	WP	WP

DN	LAT	TNYT	TNYT	NYP
NYP	TNYT	WP	LAT	NYP
LAT	WP	DN	WP	LAT
WP	DN	TNYT	DN	DN
TNYT	TNYT	LAT	TNYT	NYP
LAT	LAT	NYP	WP	DN
WP	WP	TNYT	DN	TNYT
TNYT	DN	NYP	TNYT	WP

a. Are these data categorical or quantitative?

b. Provide frequency and percent frequency distributions.

c. On the basis of the sample, which newspaper is preferred the most?

Answer:

a. The given data are categorical.

b.

Newspapers	Frequency	% Frequency
TNYT	14	28
WP	12	24
DN	9	18
NYP	7	14
LAT	8	16
Total	50	100

c. The most preferred newspaper is The New York Times.

Difficulty: Moderate
LO: 2.4, Pages 25-28
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics

6. The mentor of a class researched on the number of hours spent on study in a week by each student of the class, to analyze the correlation between the study hours and the marks obtained by each student. The data on the hours spent per week by 25 students are listed below:

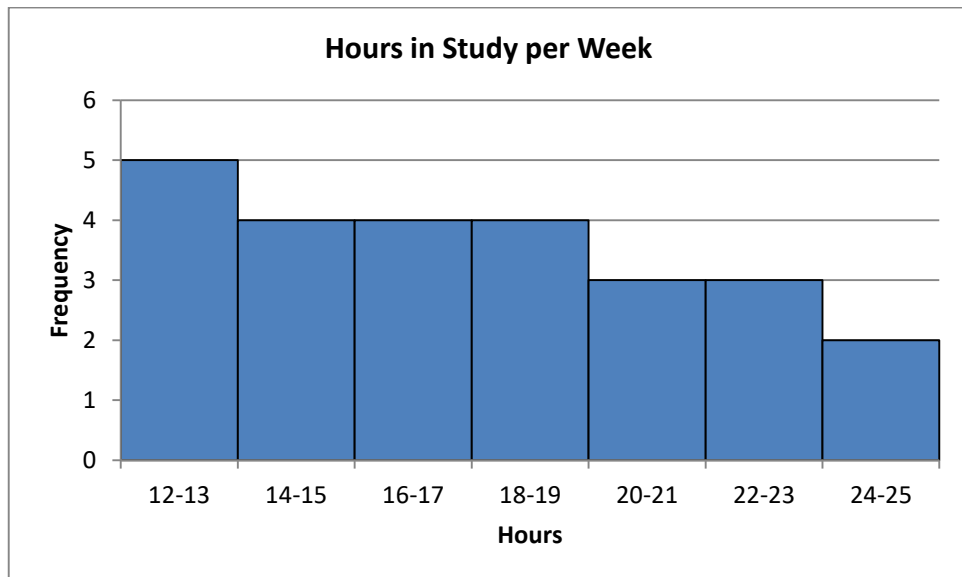|				
13	14	16	15	12
12	19	21	22	19
13	16	18	25	21
17	18	23	16	12
24	20	14	22	15

a. What is the least amount of time a student spent per week on studying after school hours in this sample? The highest?

b. Use a class width of 2 hours to prepare a frequency distribution, a relative frequency distribution, and a percent frequency distribution for the data.

c. Prepare a histogram and comment on the shape of the distribution.

Answer:

a. The least time a student spends is 12 hours, and the highest is 25 hours.

b.

Hours in Study per Week	Frequency	Relative Frequency	% Frequency
12-13	5	0.2	20
14-15	4	0.16	16
16-17	4	0.16	16
18-19	4	0.16	16
20-21	3	0.12	12
22-23	3	0.12	12
24-25	2	0.08	8
Total	25	1	100

c.

The distribution is skewed to the right.

DISC: Descriptive Statistics

7. The manager of an automobile showroom studied the time spent by each salesman interacting with the customer in a month apart from the other jobs assigned to them. The data in hours are given below.

|	
17	13
18	16
20	24
15	19
19	12
10	16
26	27
13	23
17	15
24	20
14	21
26	24

Using classes 10−13, 14−17, and so on, show:
a. The frequency distribution.
b. The relative frequency distribution.
c. The cumulative frequency distribution.
d. The cumulative relative frequency distribution.
e. The proportion of salesmen who spend 13 hours of time or less with the customers.
f. Prepare a histogram and comment on the shape of the distribution.

Answer:

Class	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
10-13	4	0.17	4	0.17
14-17	7	0.29	11	0.46
18-21	6	0.25	17	0.71
22-25	4	0.17	21	0.88
26-29	3	0.13	24	1.00 (approx.)
Total	24	≈1		

e. From the cumulative relative frequency distribution, 17% of the salesmen spend 13 hours of time or less time with the customers.

f.
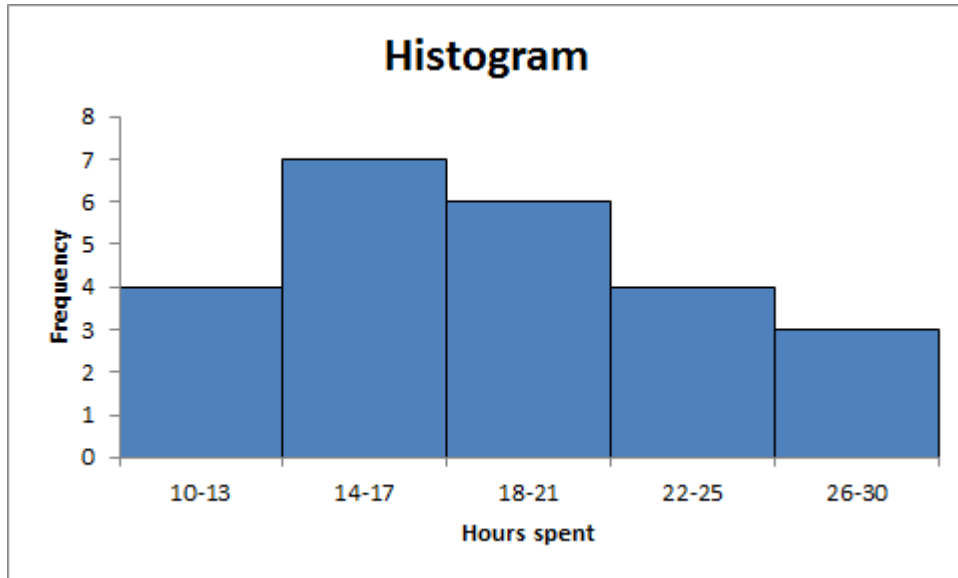
The distribution is skewed to the right.

Difficulty: Challenging
LO: 2.4, Pages 28-35
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics

8. The scores of a sample of students in a Math test are 20, 15, 19, 21, 22, 12, 17, 14, 24, 16 and in a Stat test are 16, 12, 19, 17, 22, 14, 20, 21, 24, 15, 13.
   a. Compute the mean and median scores for both the Math and the Stat tests.
   b. Compare the mean and median scores computed in part a. Comment.

   Answer:
   a. For Math test:
      Mean = 18.
      Median = 18.

      For Stat test:
      Mean = 17.5.
      Median = 17.

   b. The mean and the median scores for statistics are lower than that for mathematics. These lower values are because of an additional score 13 for statistics which is lower than the mean and the median scores for mathematics.

   Difficulty: Moderate
   LO: 2.5, Pages 35-37
   Bloom's: Application

9. Consider a sample on the waiting times (in minutes), at the billing counter in a grocery store, to be 15, 24, 18, 15, 21, 20, 15, 22, 19, 16, 15, 22, 20, 15, and 21. Compute the mean, median, and mode.

Answer: Mean = 18.53.
                 Median = 19.
                 Mode = 15.

10. Suppose that you make a fixed deposit of $1,000 in Bank X, and $500 in Bank Y. The value of each investment at the end of each subsequent year is provided in the table:

Year	Bank X ($)	Bank Y ($)
1	1,320	560
2	1,510	620
3	1,750	680
4	2,090	740
5	2,240	790
6	2,470	820
7	2,830	870
8	3,220	910
9	3,450	950
10	3,690	990

Which of the two banks provide a better return over this time period?

Answer:

a.

Year	Bank X	Growth Factor	Bank Y	Growth Factor
	1,000		500	
1	1,320	1.32	560	1.12
2	1,510	1.14	620	1.11
3	1,750	1.16	680	1.10

4	2,090	1.19	740	1.09
5	2,240	1.07	790	1.07
6	2,470	1.10	820	1.04
7	2,830	1.15	870	1.06
8	3,220	1.14	910	1.05
9	3,450	1.07	950	1.04
10	3,690	1.07	990	1.04

Geometric Mean	1.1395	**Geometric Mean**	1.0707
% of return	13.95%	**% of return**	7.07%

We observe that Bank X provides better return when compared to Bank Y.

Difficulty: Challenging
LO: 2.5, Pages 38-40
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics

11. Consider a sample on the waiting times (in minutes) at the billing counter in a grocery store to be 15, 24, 18, 15, 21, 20, 15, 22, 19, 16, 15, 22, 20, 15, and 21. Compute the 25th, 50th, and 75th percentiles.

Answer: 25th percentile = 15.
50th percentile = 19.
75th percentile = 21.

Difficulty: Moderate
LO: 2.7, Pages 44-45
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics

12. Suppose that the average time an employee takes to reach the office is 35 minutes. To address the issue of late comers, the mode of transport chosen by the employee is tracked: private transport (two-wheelers and four-wheelers) and public transport. The data on the average time (in minutes) taken using both a private transportation system and a public transportation system for a sample of employees are given below:

Private Transport	Public Transport
27	30

33	29
28	25
32	20
20	27
34	32
30	37
28	38
18	21
29	35

a. What are the mean and median travel times for employees using a private transport? What are the mean and median travel times for employees using a public transport?
b. What are the variance and standard deviation of travel times for employees using a private transport? What are the variance and standard deviation of travel times for employees using a public transport?
c. Comment.

Answer: Travel times (in minutes)
a. Using private transport:
   Mean = 27.9.
   Median = 28.5.

   Using public transport:
   Mean = 29.4.
   Median = 29.5.

b. Using private transport:
   Variance= 27.43.
   Standard deviation = 5.24.

   Using public transport:
   Variance = 39.38.
   Standard deviation = 6.28.
c. The travel times of employees using a private transport are less than that when using a public transport.

Difficulty: Moderate
LO: 2.5 and 2.6, Pages 35-37 and 41-43
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics

13. The average time a customer service executive takes to resolve an issue on a mobile handset is 26.4 minutes. The average time taken to resolve the issue by a sample of 15 such executives are shown below:

Name	Time (in minutes)
Jack	25.3
Sam	28.2
Richard	26.8
Steve	29.5
Mc Cathay	22.4
Sergio	21.7
John	24.3
Mike	22.4
Lewis	26.8
Mark	29.4
Matt	23.6
Peter	26.4
Shaggy	23.5
Jeff	26.8
Gerald	28.1

a. What is the mean resolution time?
b. What is the median resolution time?
c. What is the mode for these 15 executives?
d. What is the variance and standard deviation?
e. What is the third quartile?

Answer:

a. Mean = 25.68.
b. Median = 26.4.
c. Mode = 26.8.
d. Variance = 6.67; Standard deviation = 2.58.
e. Third Quartile = 28.1.

Difficulty: Moderate
LOs: 2.5, 2.6, 2.7, Pages 35-38, 41-43, 45-46
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics

14. Suppose that the average time an employee takes to reach the office is 35 minutes. To address the issue of late comers, the mode of transport chosen by the employee is tracked: private transport (two-wheelers and four-wheelers) and public transport. The data on the average time (in minutes) taken using both a private transportation system and a public transportation system for a sample of employees are given below:

Private Transport	Public Transport

27	30
33	29
28	25
32	20
20	27
34	32
30	37
28	38
18	21
29	35

a. Considering the travel times (in minutes) of employees using private transport. Compute the z-score for the tenth employee with travel time of 29 minutes.
b. Considering the travel times (in minutes) of employees using public transport. Compute the z-score for the second employee with travel time of 29 minutes. How does this z-score compare with the z-score you calculated for part a?
c. Based on z-scores, do the data for employees using private transport and public transport contain any outliers?

Answer:
a. For tenth employee using private transport:

The z-score is obtained as, $z = \frac{(29-27.9)}{5.24} = 0.21$.

b. For second employee using public transport:

The z-score is obtained as, $z = \frac{(29-29.4)}{6.28} = -0.06$.

Even though the employees had the same travel time, the z-score for the tenth employee in the sample who used a private transport is much larger because that employee is part of a sample with a smaller mean and a smaller standard deviation.

c.

Travel Times using Private Transport	z-score	Travel Times using Public Transport	z-score
27	-0.17	30	0.10
33	0.97	29	-0.06
28	0.02	25	-0.70
32	0.78	20	-1.50
20	-1.51	27	-0.38
34	1.16	32	0.41
30	0.40	37	1.21
28	0.02	38	1.37
18	-1.89	21	-1.34
29	0.21	35	0.89

No *z*-score is less than –3.0 or above +3.0; therefore, the *z*-scores do not indicate the existence of any outliers in either sample.

15. The results of a survey showed that on average, children spend 5.6 hours at PlayStation per week. Suppose that the standard deviation is 1.7 hours and that the number of hours at PlayStation follows a bell-shaped distribution.
    a. Use the empirical rule to calculate the percentage of children who spend between 2.2 and 9 hours at PlayStation per week.
    b. What is the *z*-value for a child who spends 7.5 hours at PlayStation per week?
    c. What is the *z*-value for a child who spends 4.5 hours at PlayStation per week?

Answer:
    a. According to the empirical rule, approximately 95% of data values will be within two standard deviations of the mean.
       2.2 is two standard deviations less than the mean and 9 is two standard deviations greater than the mean. Therefore, approximately 95% of children spend between 2.2 and 9 hours at PlayStation per week.
    b. $z = \frac{(7.5-5.6)}{1.7} = 1.12$.
    c. $z = \frac{(4.5-5.6)}{1.7} = -0.65$.

16. A study on the average minutes spent by students on internet usage is 300 with a standard deviation of 102. Answer the following questions assuming a bell-shaped distribution and using the empirical rule.
    a. What percentage of students use internet for more than 402 minutes?
    b. What percentage of students use internet for more than 504 minutes?
    c. What percentage of students use internet between 198 minutes and 300 minutes?

Answer:

    a. 402 is one standard deviation above the mean. The empirical rule states that 68% of data values will be within one standard deviation of the mean. Because a bell-shaped distribution is symmetric, 0.5×(1-68%) = 16% of the data values will be greater than (mean + 1×standard deviation) 402. 16% of students use internet for more than 402 minutes.

b. 504 is two standard deviations above the mean. The empirical rule states that 95% of data values will be within two standard deviations of the mean. Because a bell-shaped distribution is symmetric, 0.5×(1-95%) = 2.5% of the data values will be greater than (mean + 2×standard deviation) 504. 2.5% of students use internet for more than 504 minutes.

c. 198 is one standard deviation below the mean. The empirical rule states that 68% of data values will be within one standard deviation of the mean, and we expect that 0.5×(1 - 68%) = 16% of data values will be below one standard deviation below the mean. 300 is the mean, so we expect that 50% of the data values will be below the mean. Therefore, we expect 50% - 16% = 34% of the data values will be between the mean 300 and one standard deviation below the mean 198. 34% of students use internet between 198 minutes and 300 minutes.

Difficulty: Challenging
LO: 2.7, Page 48
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics

17. Eight observations taken for two variables are as follows:

x_i	y_i
11	35
13	32
17	26
18	25
22	20
24	17
26	11
28	10

a. Develop a scatter diagram with x on the horizontal axis.
b. What does the scatter diagram developed in part a indicate about the relationship between the two variables?
c. Compute and interpret the sample covariance.
d. Compute and interpret the sample correlation coefficient.

Answer:
a.
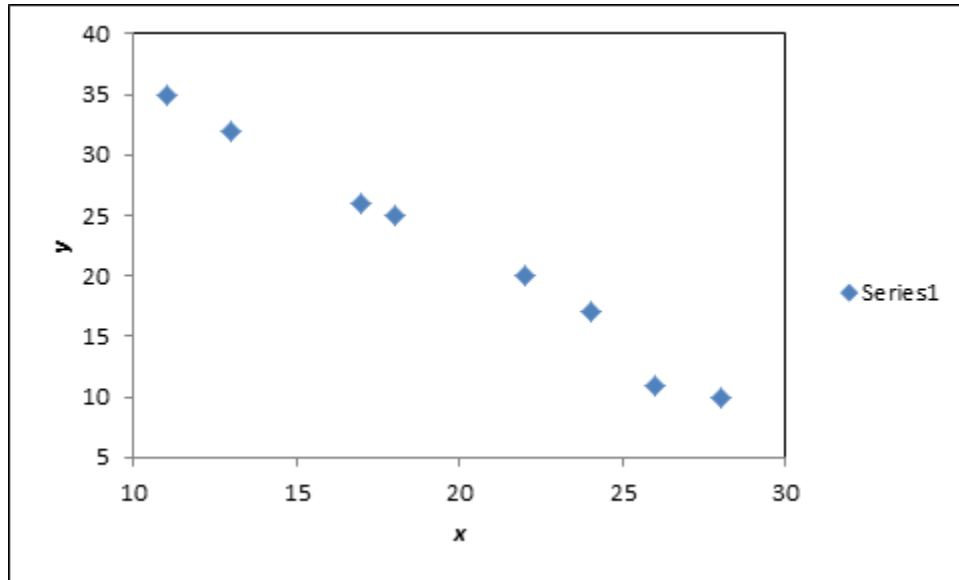
b. There appears to be a negative linear relationship between the x and y variables.

c.

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
11	35	-8.88	13	-115.38
13	32	-6.88	10	-68.75
17	26	-2.88	4	-11.50
18	25	-1.88	3	-5.63
22	20	2.13	-2	-4.25
24	17	4.13	-5	-20.63
26	11	6.13	-11	-67.38
28	10	8.13	-12	-97.50
				-391

$\bar{x}$ = 19.88

$\bar{y}$ = 22

$$s_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{-391}{7} = -55.86.$$

The negative covariance confirms that there is a negative linear relationship between the x and y variables in this data set.

d. $s_x = 6.13, s_y = 9.17$

Then the correlation coefficient is calculated as:

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{-55.86}{(6.13)(9.17)} = -0.99.$$

The correlation coefficient again confirms and indicates a strong negative linear association between the x and y variables in this data set.

Difficulty: Challenging
LO: 2.8, Page 52-56
Bloom's: Application
BUSPROG: Analytic Skills
DISC: Descriptive Statistics

18. Consider the following data on income and savings of a sample of residents in a locality:

Income ($ thousands)	Savings($ thousands)
50	10
51	11
52	13
55	14
56	15
58	15
60	16
62	16
65	17
66	17

a. Compute the correlation coefficient. Is there a positive correlation between the income and savings? What is your interpretation?
b. Show a scatter diagram of the relationship between the income and savings.

Answer:
a.

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
50	10	-7.5	-4.4	56.25	19.36	33
51	11	-6.5	-3.4	42.25	11.56	22.1
52	13	-5.5	-1.4	30.25	1.96	7.7
55	14	-2.5	-0.4	6.25	0.16	1

56	15	-1.5	0.6	2.25	0.36	-0.9
58	15	0.5	0.6	0.25	0.36	0.3
60	16	2.5	1.6	6.25	2.56	4
62	16	4.5	1.6	20.25	2.56	7.2
65	17	7.5	2.6	56.25	6.76	19.5
66	17	8.5	2.6	72.25	6.76	22.1
				292.5	52.4	116

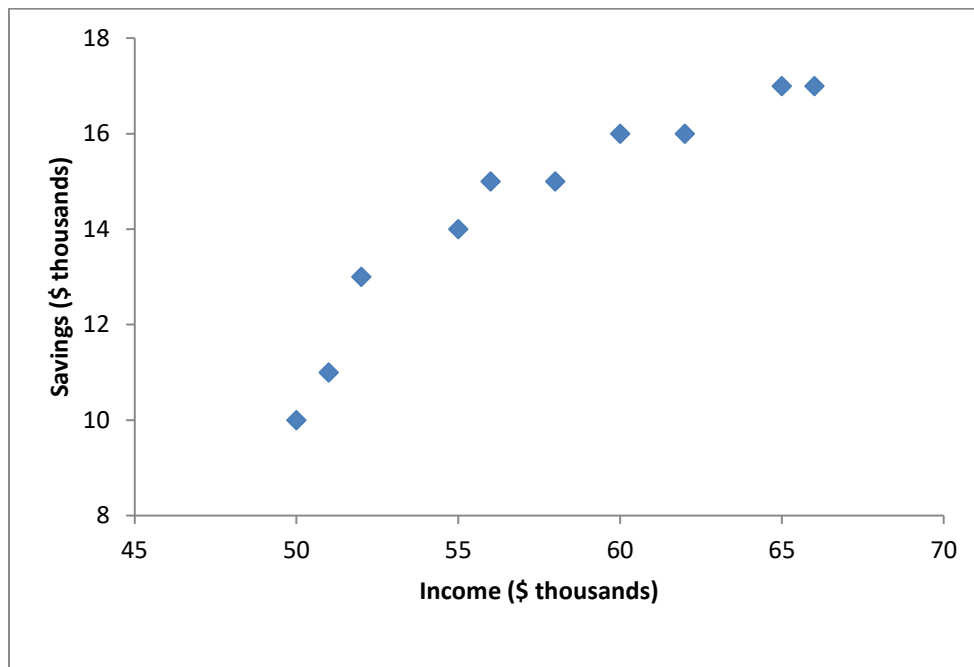$$s_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{116}{9} = 12.89.$$

$$s_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{292.5}{9}} = 5.70.$$

$$s_y = \sqrt{\frac{\sum(y - \bar{y})^2}{n-1}} = \sqrt{\frac{52.4}{9}} = 2.41.$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{12.89}{(5.70)(2.41)} = 0.938$$

This indicates that there is a strong positive relationship between income and savings.

b.

Difficulty: Challenging
LO: 2.8, Page 52-56
Bloom's: Application

BUSPROG: Analytic Skills
DISC: Descriptive Statistics