

## Chapter 02 - Descriptive Statistics

### Multiple Choice

1. A quantity of interest that can take on different values is known as a(n) \_\_\_\_\_.  
a. variable      b. parameter  
c. sample      d. observation

ANSWER: a

2. 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

3. The difference in a variable measured over observations (time, customers, items, etc.) is known as \_\_\_\_\_.  
a. observed differences      b. variation  
c. variable change      d. descriptive analytics

ANSWER: b

4. \_\_\_\_\_ acts as a representative of the population.  
a. The variable      b. The variance  
c. A sample      d. A random variable

ANSWER: c

5. The act of collecting data that are representative of the population data is called \_\_\_\_\_.  
a. random sampling      b. sample data  
c. population sampling      d. sources of data

ANSWER: a

6. The letter grades (A, B, C, D, F) of business analysis students are recorded by a professor. This variable's classification \_\_\_\_\_.  
a. is quantitative data      b. cannot be determined  
c. is categorical data      d. is time series data

ANSWER: c

7. The amount of time taken by each of 10 students in a class to complete an exam is an example of what type of data?  
a. Cannot be determined      b. Categorical data  
c. Time series data      d. Quantitative data

ANSWER: d

8. \_\_\_\_\_ 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

9. Data collected from several entities over a period of time (minutes, hours, days, etc.) are called \_\_\_\_\_.  
a. categorical and quantitative data      b. time series data

**Chapter 02 - Descriptive Statistics**

- c. source data
- d. cross-sectional data

ANSWER: b

10. 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

11. The data collected from the customers in restaurants about the quality of food is an example of a(n) \_\_\_\_\_.

- a. variable study
- b. cross-sectional study
- c. experimental study
- d. observational study

ANSWER: d

12. When working with large spreadsheets with many rows of data, it can be helpful to \_\_\_\_\_ the data to better find, view, or manage subsets of data.

- a. split
- b. sort and filter
- c. chart
- d. manipulate

ANSWER: b

13. When working with data sets in Excel, \_\_\_\_\_ can be used to automatically highlight cells that meet specified requirements.

- a. averaging
- b. conditional formatting
- c. summing
- d. sorting

ANSWER: b

14. A summary of data that shows the number of observations in each of several nonoverlapping bins is called a(n) \_\_\_\_\_.

- a. frequency distribution
- b. sample summary
- c. bin distribution
- d. observed distribution

ANSWER: a

15. Which of the following gives the proportion of items in each bin?

- a. Frequency
- b. Class size
- c. Relative frequency
- d. Bin proportion

ANSWER: c

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

<b>Grades</b>	<b>Number of students</b>
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

**Chapter 02 - Descriptive Statistics**

- c. 0.16, 0.28, 0.33, 0.13      d. 0.18, 0.31, 0.37, 0.14

ANSWER: d

17. 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

18. 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

19. The goal regarding using an appropriate number of bins is to show the \_\_\_\_\_.

- a. number of observations      b. number of variables
- c. variation in the data      d. correlation in the data

ANSWER: c

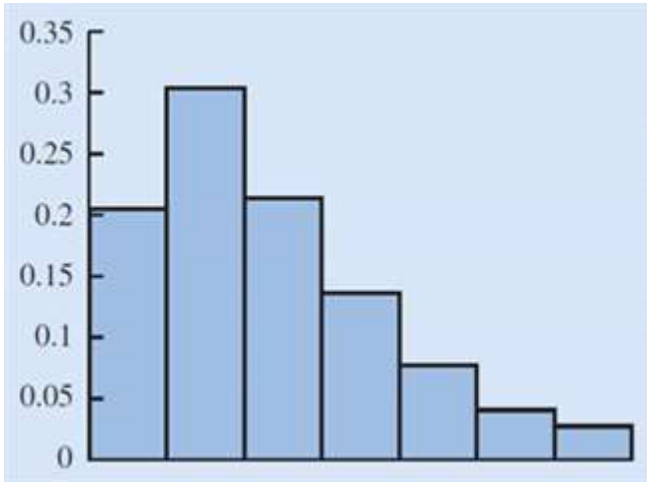
20. A \_\_\_\_\_ is a graphical summary of data previously summarized in a frequency distribution.

- a. box plot      b. histogram
- c. line chart      d. scatter chart

ANSWER: b

21. Identify the shape of the distribution in the figure below.

**Chapter 02 - Descriptive Statistics**



- a. Skewed left
- b. Symmetric
- c. Approximately bell shaped
- d. Skewed right

ANSWER: d

22. 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

23. 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

24. 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

25. 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

26. Compute the mode for the following data.

12, 16, 19, 10, 12, 11, 21, 12, 21, 10

**Chapter 02 - Descriptive Statistics**

- a. 21      b. 11
- c. 12      d. 10

ANSWER: c

27. 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

28. The simplest measure of variability is the \_\_\_\_\_.

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

ANSWER: d

29. 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

30. Use technology to 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

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

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

- a. 18.64%      b. 21.36%
- c. 20.28%      d. 21.67%

ANSWER: b

32. Compute the 50<sup>th</sup> 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

ANSWER: c

33. Compute the third quartile for the following data.

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

## Chapter 02 - Descriptive Statistics

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

ANSWER: a

34. Compute the 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

35. 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

36. 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

37. Any data value with a z-score less than  $-3$  or greater than  $+3$  is considered to be a(n) \_\_\_\_\_.

- a. outlier      b. statistic
- c. whisker     d. z-score value

ANSWER: a

38. Which of the following graphs provides information on outliers and IQR of a data set?

- a. Histogram      b. Line chart
- c. Scatter chart    d. Box plot

ANSWER: d

39. If the covariance between two variables is near 0, it implies that \_\_\_\_\_.

- a. a positive relationship exists between the variables
- b. the variables are not linearly related
- c. the variables are negatively related
- d. the variables are strongly related

ANSWER: b

40. 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

## Chapter 02 - Descriptive Statistics

41. Scores on Ms. Bond's test have a mean of 70 and a standard deviation of 11. Michelle has a score of 48. Convert Michelle's score to a  $z$ -score. (Round to two decimal places if necessary.)

- a. 2
- b. 41.64
- c.  $-2$
- d. 1.33

ANSWER: c

42. Scores on Ms. Nash's test have a mean of 64 and a standard deviation of 9. Steve has a score of 52. Convert Steve's score to a  $z$ -score. (Round to two decimal places if necessary.)

- a. 1.33
- b. 58.2
- c.  $-2$
- d.  $-1.33$

ANSWER: d

43. Scores on Ms. Bond's test have a mean of 70 and a standard deviation of 11. David has a score of 52 on Ms. Bond's test. Scores on Ms. Nash's test have a mean of 64 and a standard deviation of 6. Steven has a score of 52 on Ms. Nash's test. Which student has the higher standardized score?

- a. David's standardized score is  $-1.64$  and Steven's standardized score is  $-2.00$ . Therefore, David has the higher standardized score.
- b. David's standardized score is  $-1.64$  and Steven's standardized score is  $-2.00$ . Therefore, Steven has the higher standardized score.
- c. David's standardized score is 1.64 and Steven's standardized score is 2.00. Therefore, Steven has the higher standardized score.
- d. Cannot be determined with the information provided.

ANSWER: b

44. The College Board originally scaled SAT scores so that the scores for each section were approximately normally distributed with a mean of 500 and a standard deviation of 100. Assuming scores follow a bell-shaped distribution, use the empirical rule to find the percentage of students who scored greater than 700.

- a. 97.5%
- b. 95%
- c. 2.5%
- d. 5%

ANSWER: c

45. The College Board originally scaled SAT scores so that the scores for each section were approximately normally distributed with a mean of 500 and a standard deviation of 100. Assuming scores follow a bell-shaped distribution, use the empirical rule to find the percentage of students who scored less than 400.

- a. 16%
- b. 68%
- c. 84%
- d. 32%

ANSWER: a

**Chapter 02 - Descriptive Statistics**

46. The College Board reported that the mean Math Level 2 SAT subject test score was 686 with a standard deviation of 96. Assuming scores follow a bell-shaped distribution, use the empirical rule to find the percentage of students who scored less than 494.

- a. 97.5%
- b. 95%
- c. 2.5%
- d. 5%

ANSWER: c

47. Compute the relative frequency for students who earned an A shown in the table of grades below.

<b>Grades</b>	<b>Number of Students</b>
A	10
B	31
C	36
D	6
	<b>83</b>

- a. 0.12
- b. 0.10
- c. 0.83
- d. Not enough information

ANSWER: a

48. Compute the relative frequency for students who earned a C shown in the table of grades below.

<b>Grades</b>	<b>Number of Students</b>
A	10
B	31
C	36
D	6
	<b>83</b>

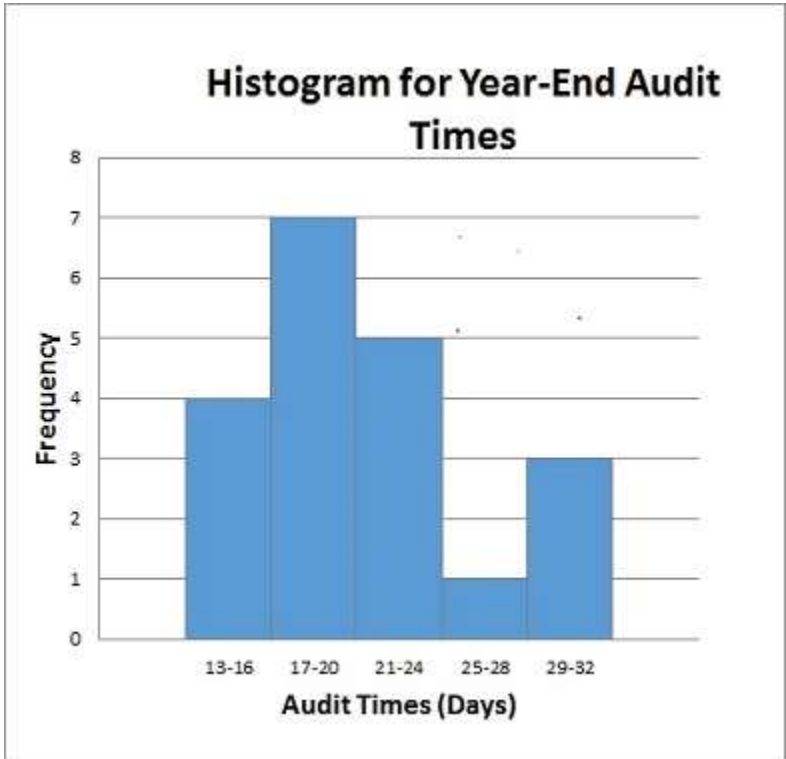
- a. 0.43
- b. 0.53
- c. 0.83
- d. Not enough information

ANSWER: a

49. Below is a histogram for the number of days that it took Wyche Accounting to perform audits in the last quarter of last year. What is the relative frequency of the 21-24 bin?



**Chapter 02 - Descriptive Statistics**

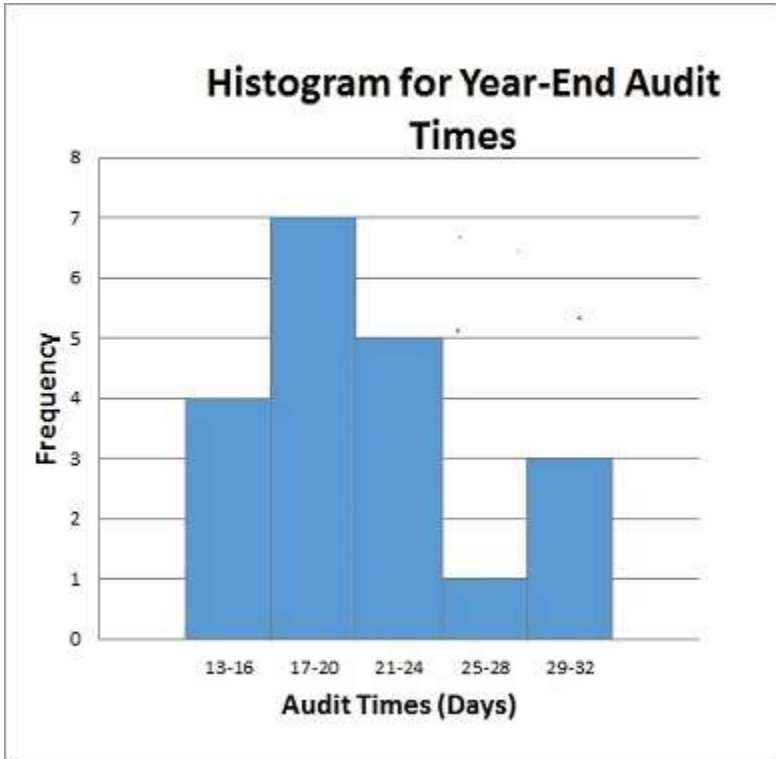


- a. 0.05
- b. 0.14
- c. 0.25
- d. 2.5

ANSWER: c

50. Below is a histogram for the number of days that it took Wyche Accounting to perform audits in the last quarter of last year. What is the frequency of the 25-28 bin?

**Chapter 02 - Descriptive Statistics**



- a. 0.05
- b. 1
- c. 0.5
- d. 4

ANSWER: b

51. What is the total relative frequency?

**20XX Contest Sales**

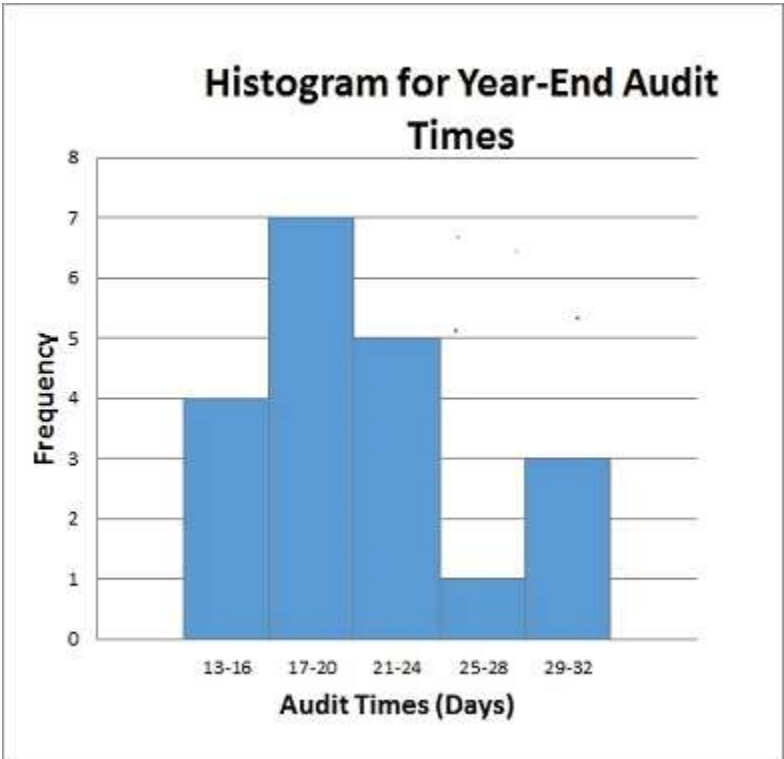
Salesman	Frequency	Relative Frequency
Frances Clonts	15	0.05
Sarah Leigh	184	0.62
Devon Pride	37	
John Townes	62	0.21
<b>Total</b>	<b>298</b>	

- a. 1
- b. 99.12
- c. 0.88
- d. Not enough information

ANSWER: a

52. Below is a histogram for the number of days that it took Wyche Accounting to perform audits in the last quarter of last year. What is the bin size for the histogram?

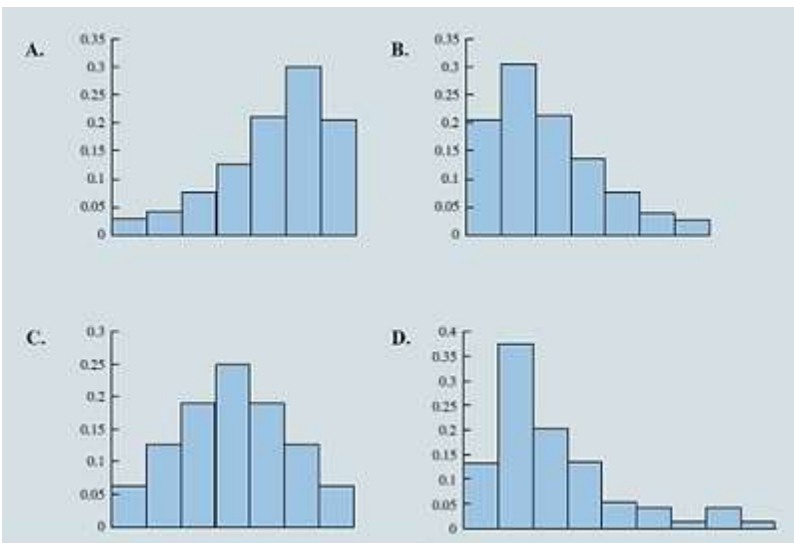
**Chapter 02 - Descriptive Statistics**



- a. 3
- b. 4
- c. 16
- d. 5

ANSWER: b

53. Select the histogram that is moderately skewed right.



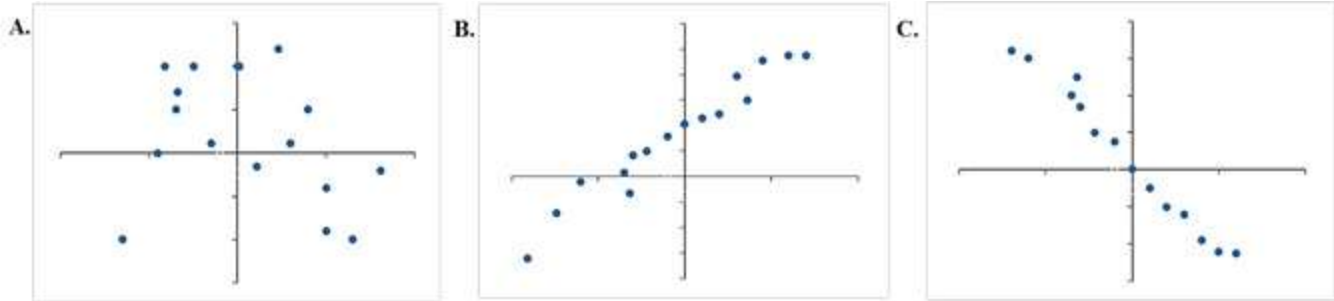
- a. A
- b. B

**Chapter 02 - Descriptive Statistics**

- c. C
- d. D

ANSWER: b

54. Which graph represents a negative linear relationship between  $x$  and  $y$ ?



- a. A
- b. B
- c. C
- d. None of the graphs display a negative linear relationship.

ANSWER: c

55. Below is the data for the number of days that it took Wyche Accounting to perform audits in the last quarter of last year.

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

What is the median number of days that it took Wyche Accounting to perform audits in the last quarter of last year?

- a. 41
- b. 40.6
- c. 39.5
- d. 42

ANSWER: c

56. What is the mode of the data set given below?

35, 47, 65, 47, 22

- a. 47.5
- b. 47
- c. 65
- d. 22

ANSWER: b

57. A sample of 13 adult males' heights are listed below.

70, 72, 71, 70, 69, 73, 69, 68, 70, 71, 67, 71, 74

**Chapter 02 - Descriptive Statistics**

Find the range of the data.

- a. 7
- b. 6.5
- c. 5
- d. 4

ANSWER: a

58. James’s manager asked him to sort the last names in the following list in descending order. What does this mean?

Customer ID	First	Last	Sales	Quantity	Discount	Profit
CG-12520	Claire	Gute	\$ 261.96	2	0	\$ 41.91
DV-13045	Darrin	VanHuff	\$ 14.62	2	0	\$ 16.87
SO-20335	Sean	O'Donnell	\$ 957.58	5	0.45	\$ (383.03)
BH-11710	Brosina	Hoffman	\$ 48.86	7	0	\$ 14.17
AA-10480	Andrew	Allen	\$ 25.55	3	0.2	\$ 5.44
IM- 15070	Irene	Maddox	\$ 407.98	3	0.2	\$ 132.59
HP-14815	Harold	Pawlan	\$ 68.81	5	0.8	\$ (123.86)
PK-19075	Pete	Kriz	\$ 665.88	6	0	\$ 13.32
AG-10270	Alejandro	Grove	\$ 55.50	2	0	\$ 9.99
ZD-21925	Zuschuss	Donatelli	\$ 8.56	2	0	\$ 2.48

- a. The last names must be sorted from A to Z.
- b. The last names must be sorted from Z to A.
- c. The last names must be sorted from the earliest to the latest that has been added to the list.
- d. James should use the Sort function to organize the data into order of sales.

ANSWER: b

59. You have been asked to reorganize the Excel table below into order of sales using the Sales column. Which option will allow you to do this quickly?

Customer ID	First	Last	Sales	Quantity	Discount	Profit
CG-12520	Claire	Gute	\$ 261.96	2	0	\$ 41.91
DV-13045	Darrin	VanHuff	\$ 14.62	2	0	\$ 16.87
SO-20335	Sean	O'Donnell	\$ 957.58	5	0.45	\$ (383.03)
BH-11710	Brosina	Hoffman	\$ 48.86	7	0	\$ 14.17
AA-10480	Andrew	Allen	\$ 25.55	3	0.2	\$ 5.44
IM- 15070	Irene	Maddox	\$ 407.98	3	0.2	\$ 132.59
HP-14815	Harold	Pawlan	\$ 68.81	5	0.8	\$ (123.86)
PK-19075	Pete	Kriz	\$ 665.88	6	0	\$ 13.32
AG-10270	Alejandro	Grove	\$ 55.50	2	0	\$ 9.99
ZD-21925	Zuschuss	Donatelli	\$ 8.56	2	0	\$ 2.48

## Chapter 02 - Descriptive Statistics

- a. Use the Cut and Paste function to reorganize the data into order of sales.
- b. Use the Filter function to organize the data into order of sales.
- c. Use the Order function to organize the data into order of sales.
- d. Use the Sort function to organize the data into order of sales.

ANSWER: d

60. Which Excel command will return all modes when more than one mode exists?

- a. MODE.MULT
- b. MODE.SNGL
- c. MODE
- d. MODES

ANSWER: a

61. In a survey of patients in a local hospital, 62.42% of the respondents indicated that the health care providers needed to spend more time with each patient. Who makes up the population?

- a. All patients in a local hospital
- b. All survey respondents
- c. Hospital patients
- d. Cannot be determined from the information given

ANSWER: a

62. In a survey of patients in a local hospital, 62.42% of the respondents indicated that the health care providers needed to spend more time with each patient. Who makes up the sample?

- a. All patients in a local hospital
- b. All survey respondents
- c. Hospital patients
- d. Cannot be determined from the information given

ANSWER: b

63. A manager of a fast food restaurant wants the drive-thru employee to ask every fifth customer if he or she is satisfied with the service. Who makes up the population?

- a. All customers who use the drive-thru window of this fast food restaurant
- b. All survey respondents
- c. All customers of this restaurant
- d. The proportion of customers who say they are satisfied with their service

ANSWER: a

64. A manager of a fast food restaurant wants the drive-thru employee to ask every fifth customer if he or she is satisfied with the service. Who makes up the sample?

- a. All customers who use the drive-thru window of this fast food restaurant
- b. All survey respondents
- c. All customers of this restaurant

## Chapter 02 - Descriptive Statistics

d. The proportion of customers who say they are satisfied with their service

ANSWER: b

65. Which of the following relationships would have a negative correlation coefficient?

- a. Supply and demand
- b. Amount of a bill at a restaurant and the amount of the tip
- c. Cost of a car and the amount of tax to be paid
- d. The square footage of a home and the price of the home

ANSWER: a

66. The distribution of hourly sales for a local family owned store is normally distributed with a mean of \$225 per hour and a standard deviation of \$75 per hour. Which of the following intervals contains the middle 95% of hourly sales?

- a. \$75 to \$375
- b. \$150 to \$300
- c. \$175 to \$275
- d. \$125 to \$325

ANSWER: a

67. Data sets commonly include observations with missing values for one or more variables. In some cases missing data naturally occur; these are called \_\_\_\_\_.

- a. legitimately missing data
- b. data cleansing
- c. illegitimate missing data
- d. missing random data

ANSWER: a

68. \_\_\_\_\_ is the process of removing variables from the analysis without losing crucial information.

- a. Data Cleansing
- b. Dimension reduction
- c. Legitimate missing data
- d. Missing random data

ANSWER: b

### Subjective Short Answer

69. A student willing to participate in a debate competition is required to fill out a registration form. State whether each of the following information about the participant provides categorical or quantitative data.

- a. What is your birth month?
- b. Have you participated in any debate competition previously?
- c. If yes, in 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. Categorical  
b. Categorical  
c. Quantitative  
d. Categorical

**Chapter 02 - Descriptive Statistics**

e. Quantitative

70. 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 five 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 the greatest number of billionaires are the 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

71. The data on the percentage of visitors in the previous and current years at 12 well-known national parks of the United 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%



**Chapter 02 - Descriptive Statistics**

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%

- 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?
- Calculate the change in visitor percentage from the previous to the current year for each park. Use Excel’s conditional formatting to highlight the parks whose visitor percentage decreased from the previous year to the current year.
- 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:

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

<b>National Parks</b>	<b>Percentage of visitors previous year</b>	<b>Percentage of visitors current year</b>
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 visitors in the current year, and Yellowstone has the lowest number of visitors in the current year.

- 

<b>National Parks</b>	<b>Percentage of visitors previous year</b>	<b>Percentage of visitors current year</b>	<b>Change in visitor percentage</b>
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%

**Chapter 02 - Descriptive Statistics**

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%

72. The partial relative frequency distribution is given below:

Group	Relative Frequency
1	0.15
2	0.32
3	0.29
4	

- What is the relative frequency of group 4?
- The total sample size is 400. What is the frequency of group 4?
- Show the frequency distribution.
- Show the percent frequency distribution.

ANSWER:

- The relative frequency of group 4 is obtained as  $1.00 - 0.15 - 0.32 - 0.29 = 0.24$ .
- If the total sample size is 400, the frequency of group 4 is obtained as  $0.24 \times 400 = 96$ .

c.

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

d.

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

73. A survey on the most preferred newspaper in the 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

**Chapter 02 - Descriptive Statistics**

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.

74. The mentor of a class researched the number of hours spent on study in a week by each student of the class in order 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 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.

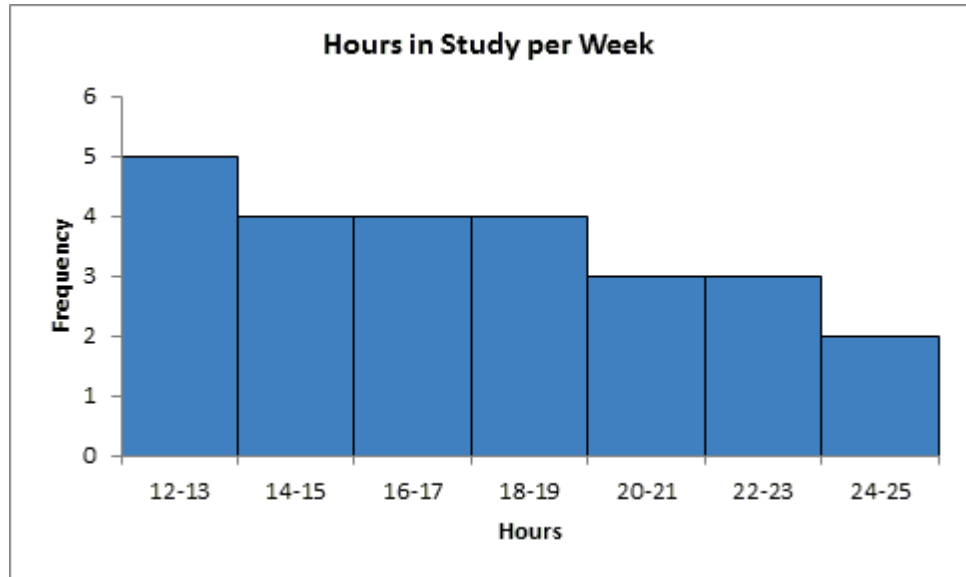
ANSWER:

- a. The least amount of time a student spent was 12 hours, and the highest was 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.

**Chapter 02 - Descriptive Statistics**



The distribution is skewed to the right.

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

- 17
- 18
- 20
- 15
- 19
- 10
- 26
- 13
- 17
- 24
- 14
- 26
- 13
- 17
- 24
- 14
- 26
- 13
- 17
- 24
- 14
- 26

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 salesperson who spent 13 hours of time or less with the customers.
- f. Prepare a histogram and comment on the shape of the distribution.

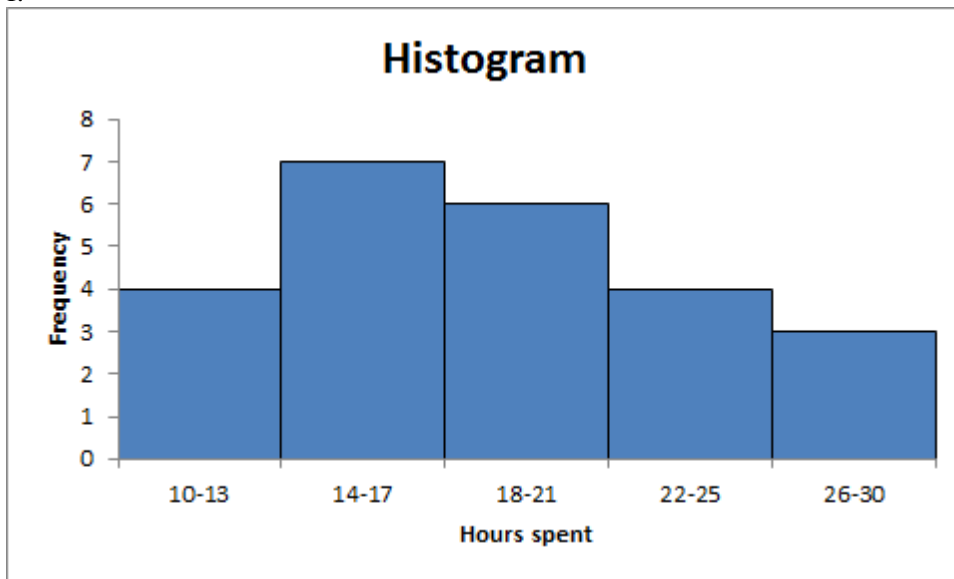
ANSWER: a. – d.

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		

**Chapter 02 - Descriptive Statistics**

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

f.



The distribution is skewed to the right.

76. 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.

- Compute the mean and median scores for both the Math and the Stat tests.
- 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.

77. 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

78. 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

**Chapter 02 - Descriptive Statistics**

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 provides 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

<b>Geometric Mean</b>	1.1395	<b>Geometric Mean</b>	1.0707
<b>% of return</b>	13.95%	<b>% of return</b>	7.07%

Bank X provides a better return when compared to Bank Y.

79. 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 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles.

ANSWER: 25<sup>th</sup> percentile = 15  
 50<sup>th</sup> percentile = 19  
 75<sup>th</sup> percentile = 21

80. 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

**Chapter 02 - Descriptive Statistics**

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 on the results.

**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 those using a public transport.

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

Name	Time (in minutes)
Jack	25.3
Samantha	28.2
Richard	26.8
Steve	29.5
Mary	22.4
Sergio	21.7
John	24.3
Michelle	22.4
Linda	26.8
Mark	29.4
Matt	23.6
Polly	26.4
Sheila	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?

**Chapter 02 - Descriptive Statistics**

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

82. 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: First, calculate the mean (AVERAGE function in Excel) and standard deviation (STDEV.S function in Excel) for private transport.

Mean = 27.9, StDev = 5.24

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

b. For second employee using public transport: First, calculate the mean (AVERAGE function in Excel) and standard deviation (STDEV.S function in Excel) for public transport.

Mean = 29.4, StDev = 6.28

The *z*-score is then 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	<i>z</i> -score	Travel Times using Public Transport	<i>z</i> -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



**Chapter 02 - Descriptive Statistics**

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.

83. The results of a survey showed that, on average, children spend 5.6 hours on PlayStation per week. Suppose that the standard deviation is 1.7 hours and that the number of hours on 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 on PlayStation per week.

b. What is the z-value for a child who spends 7.5 hours on PlayStation per week?

c. What is the z-value for a child who spends 4.5 hours on 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 on PlayStation per week.

b. 
$$z = \frac{(7.5 - 5.6)}{1.7} = 1.12.$$

c. 
$$z = \frac{(4.5 - 5.6)}{1.7} = -0.65.$$

84. 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 the Internet for more than 402 minutes?

b. What percentage of students use the Internet for more than 504 minutes?

c. What percentage of students use the 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,  $0.5 \times (1 - 68\%) = 16\%$  of the data values will be greater than (mean + 1  $\times$  standard deviation) 402. 16% of students use the 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,  $0.5 \times (1 - 95\%) = 2.5\%$  of the data values will be greater than (mean + 2  $\times$  standard deviation) 504. 2.5% of students use the 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  $0.5 \times (1 - 68\%) = 16\%$  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 the Internet between 198 minutes and 300 minutes.

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

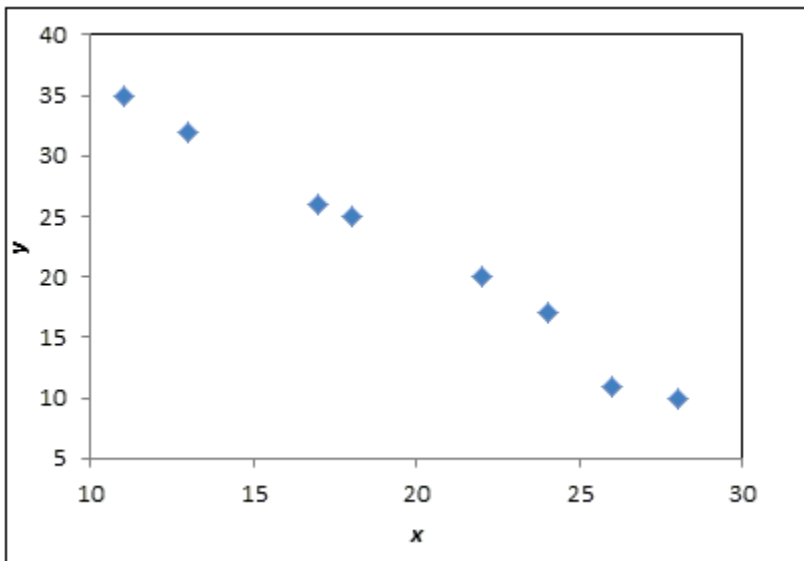
$x_i$	$y_i$

**Chapter 02 - Descriptive Statistics**

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

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

ANSWER: a.



- There appears to be a negative linear relationship between the  $x$  and  $y$  variables.

c.

$x_i$	$y_i$	$x_i - y_i$	$u - x_i$	$(x_i - \bar{x})(y_i - \bar{y})$
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

**Chapter 02 - Descriptive Statistics**

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.

86. 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

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

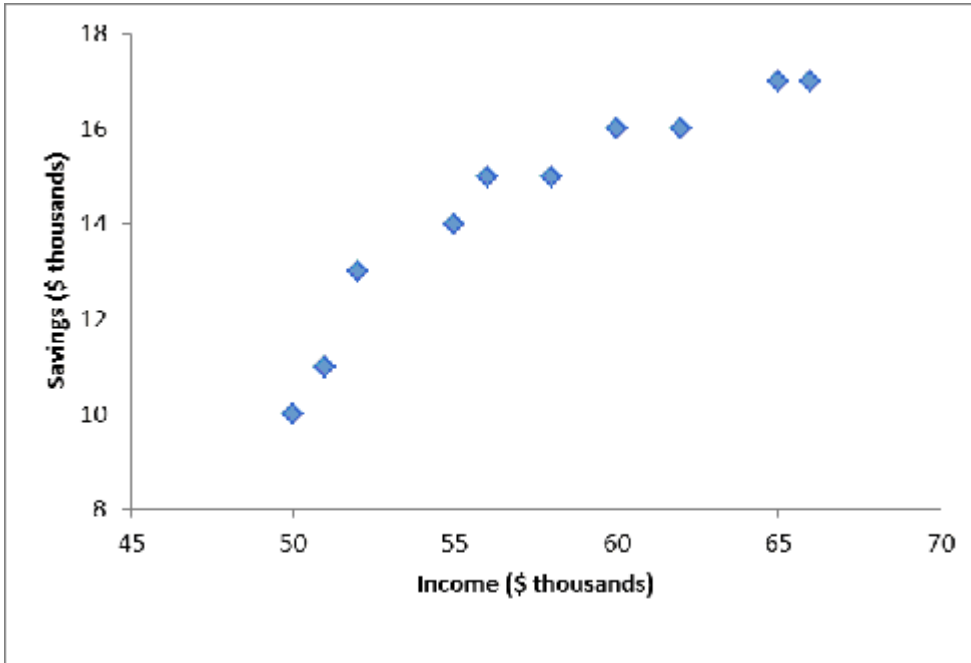
$$s_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{292.5}{9}} = 5.70.$$

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

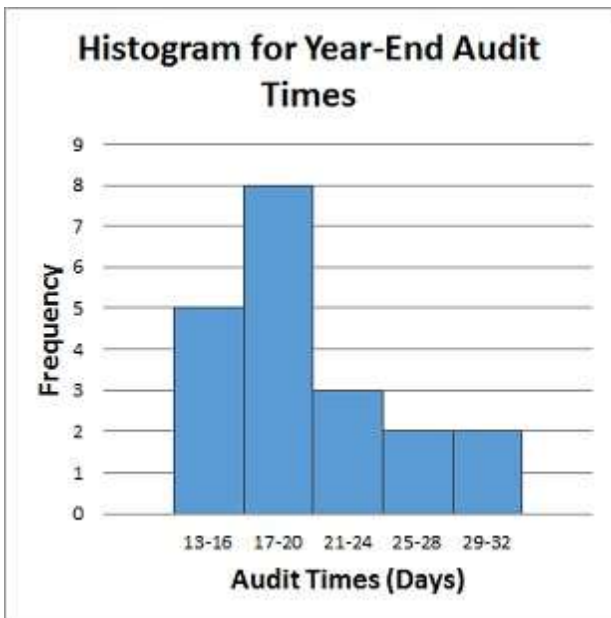
**Chapter 02 - Descriptive Statistics**

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

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



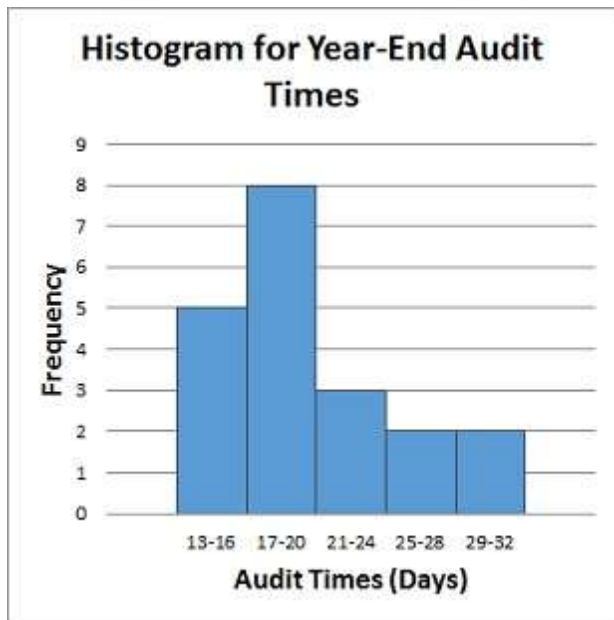
87. Below is a histogram for the number of days that it took Wyche Accounting to perform audits in the last quarter of last year. What is the relative frequency of the 21-24 bin?



ANSWER: 0.15

88. Below is a histogram for the number of days that it took Wyche Accounting to perform audits in the last quarter of last year. What is the relative frequency of the 25-28 bin?

## Chapter 02 - Descriptive Statistics



ANSWER: 0.10

89. Below are the data for the number of days that it took Wyche Accounting to perform audits in the last quarter of last year.

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

What is(are) the mode(s) number of days that it took Wyche Accounting to perform audits in the last quarter of last year?

ANSWER: None

90. What is(are) the mode(s) of the following data set?

35, 42, 65, 42, 22

ANSWER: 42

91. The difference between the largest and the smallest data values is the \_\_\_\_\_.

ANSWER: range

92. The Excel function STANDARDIZE can be used to calculate \_\_\_\_\_.

ANSWER: z-scores

93. You would \_\_\_\_\_ a table if you wanted to display only data that match specific criteria.

ANSWER: filter

94. The \_\_\_\_\_ measures the variability of the middle 50% of a data set.

ANSWER: IQR

95. Below are the data for African countries. Assess the quality of the data by identifying missing values and sort by GDP.

**Chapter 02 - Descriptive Statistics**

<b>A</b>	<b>B</b>	<b>C</b>
<b>1 Country</b>	<b>Continent</b>	<b>GDP (millions of US\$)</b>
2 Algeria	Africa	190,709
3 Angola	Africa	100,948
4 Bolivia	Africa	24,604
5 Botswana	Africa	17,570
6 Cameroon	Africa	
7 Congo, Democratic Republic of the	Africa	15,668
8 Congo, Republic of the	Africa	14,769
9 Côte d'Ivoire	Africa	24,096
10 Egypt	Africa	235,719
11 Equatorial Guinea	Africa	
12 Ethiopia	Africa	31,256
13 Gabon	Africa	16,176
14 Ghana	Africa	
15 Jordan	Africa	29,233
16 Kenya	Africa	34,796
17 Libya	Africa	36,874
18 Mali	Africa	
19 Mauritius	Africa	11,313
20 Morocco	Africa	99,241
21 Mozambique	Africa	12,827
22 Namibia	Africa	
23 Nigeria	Africa	238,920
24 Senegal	Africa	14,461
25 South Africa	Africa	408,074
26 Sudan +South Sudan	Africa	64,750
27 Tanzania	Africa	23,333
28 Uganda	Africa	16,810
29 Yemen	Africa	33,675
30 Zambia	Africa	19,206

**ANSWER:**

<b>Country</b>	<b>Continent</b>	<b>GDP (millions of US\$)</b>
Mauritius	Africa	11,313
Mozambique	Africa	12,827
Senegal	Africa	14,461
Congo, Republic of the	Africa	14,769
Congo, Democratic Republic of the	Africa	15,668
Gabon	Africa	16,176
Uganda	Africa	16,810
Botswana	Africa	17,570
Zambia	Africa	19,206

**Chapter 02 - Descriptive Statistics**

Tanzania	Africa	23,333
Côte d'Ivoire	Africa	24,096
Bolivia	Africa	24,604
Jordan	Africa	29,233
Ethiopia	Africa	31,256
Yemen	Africa	33,675
Kenya	Africa	34,796
Libya	Africa	36,874
Sudan +South Sudan	Africa	64,750
Morocco	Africa	99,241
Angola	Africa	100,948
Algeria	Africa	190,709
Egypt	Africa	235,719
Nigeria	Africa	238,920
South Africa	Africa	408,074
Cameroon	Africa	
Equatorial Guinea	Africa	
Ghana	Africa	
Mali	Africa	
Namibia	Africa	