

TEST BANK

Paul Newbold | William L. Carlson | Betty Thorne

Seventh Edition

Statistics

for Business and Economics



CHAPTER 2

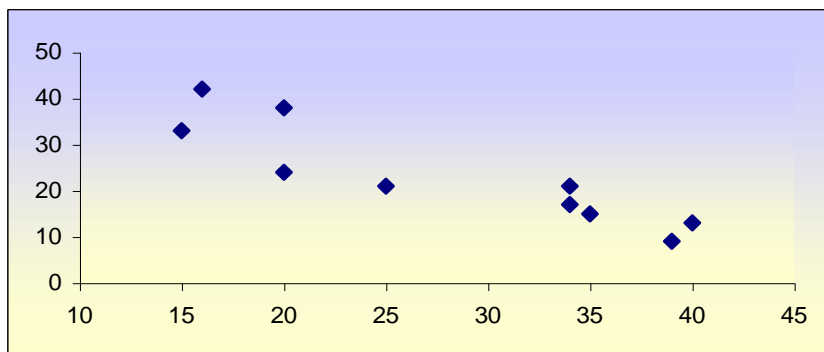
Describing Data: Numerical

Multiple-Choice Questions

1. A scatter plot can illustrate all of the following except:
 - A) the median of each of the two variables
 - B) the range of each of the two variables
 - C) an indication of extreme large or extreme small values (outliers)
 - D) patterns of values**ANSWER: A**
2. If you are interested in comparing variation in sales for small and large stores selling similar goods, which of the following is the most appropriate measure of dispersion?
 - A) The range
 - B) The interquartile range
 - C) The standard deviation
 - D) The coefficient of variation**ANSWER: D**
3. Consider the following (X, Y) data: $(53, 37)$, $(34, 26)$, $(10, 29)$, $(63, 55)$, $(28, 36)$, $(58, 48)$, $(28, 41)$, $(50, 42)$, $(39, 21)$, and $(35, 46)$. What is the correlation coefficient?
 - A) 0.710
 - B) 0.782
 - C) 0.674
 - D) 0.590**ANSWER: D**
4. Suppose you are told that the mean sample of numbers is below the median. What does this information suggest?
 - A) The distribution is symmetric.
 - B) The distribution is skewed to the right or positively skewed.
 - C) The distribution is skewed to the left or negatively skewed.
 - D) There is insufficient information to determine the shape of the distribution.**ANSWER: C**

Describing Data: Numerical

5. For the following scatter plot, what would be your best estimate of the correlation coefficient?



- A) -0.8
- B) -1.0
- C) -0.2
- D) -0.3

ANSWER: A

6. Suppose that we are interested in exploring the determinants of successful high schools. One possible measure of success might be the percentage of students who go on to college. The teachers' union argues that there should be a relationship between the average teachers' salary and high school success. After running a regression of the percentage of students going on to college and average teachers' salary, it is pointed out that one school has a large negative residual. Which of the following is true?

- A) This school has very low values for both variables.
- B) This school has very high values for both variables.
- C) This school performed much worse than expected.
- D) This school performed much better than expected.

ANSWER: C

7. Given a set of 25 observations, for what values of the correlation coefficient would we be able to say that there is evidence that a relationship exists between the two variables?

- A) $|r_{xy}| \geq 0.40$
- B) $|r_{xy}| \geq 0.35$
- C) $|r_{xy}| \geq 0.30$
- D) $|r_{xy}| \geq 0.25$

ANSWER: A

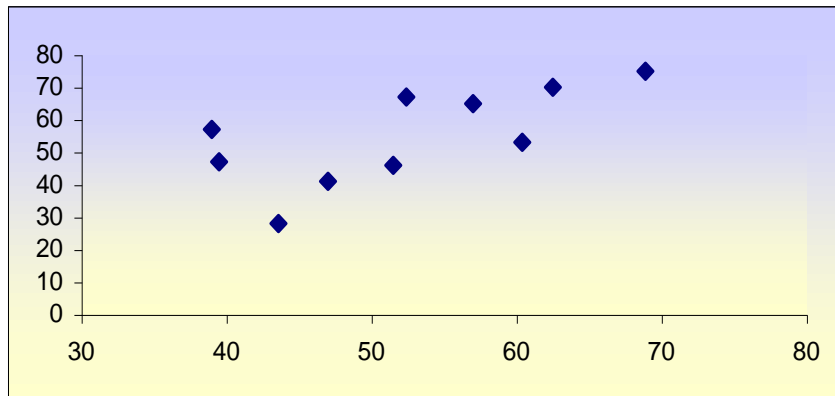
Chapter 2

8. Which of the following statements is true about correlation coefficient and covariance?

- A) The covariance is the preferred measure of the relationship between two variables since it is generally larger than the correlation coefficient.
- B) The correlation coefficient is a preferred measure of the relationship between two variables since its calculation is easier than the covariance.
- C) The covariance is a standardized measure of the relationship between variables.
- D) The correlation coefficient is the preferred measure of the relation between variables since it is a standardized measure.

ANSWER: C

9. For the following scatter plot, what would be your best estimate of the correlation coefficient?



- A) 1.0
- B) 0.7
- C) 0.3
- D) 0.1

ANSWER: B

10. Which of the following descriptive statistics is least affected by outliers?

- A) Mean
- B) Median
- C) Range
- D) Standard deviation

ANSWER: B

Describing Data: Numerical

11. Which of the following statements is true?
- A) The correlation coefficient is always greater than the covariance.
 - B) The covariance is always greater than the correlation coefficient.
 - C) The covariance may be equal to the correlation coefficient.
 - D) Neither the covariance nor the correlation coefficient can be equal to zero.
- ANSWER: C**
12. Which measures of central location are not affected by extremely small or extremely large values data values?
- A) Arithmetic mean and median
 - B) Median and mode
 - C) Mode and arithmetic mean
 - D) Geometric mean and arithmetic mean
- ANSWER: B**
13. Suppose you are told that sales this year are 30% higher than they were six years ago. What has been the average annual increase in sales over the past six years?
- A) 5.0%
 - B) 4.5%
 - C) 4%
 - D) 3.5%
- ANSWER: B.**
14. Suppose you are told that sales this year are 20% higher than they were five years ago. What has been the annual average increase in sales over the past five years?
- A) 5.2%
 - B) 4.7%
 - C) 4.2%
 - D) 3.7%
- ANSWER: B**
15. Suppose you are told that over the past four years, sales have increased at rates of 10%, 8%, 6%, and 4%. What has been the average annual increase in sales over the past four years?
- A) 7.0%
 - B) 6.7%
 - C) 6.4%
 - D) 6.5%
- ANSWER: A**

Chapter 2

16. Suppose you are told that the average return on investment for a particular class of investments was 7.8% with a standard deviation of 2.3. Furthermore, the histogram of the distribution of returns is approximately mound-shaped. We would expect that 95 percent of all of these investments had a return between what two values?

- A) 5.5% and 10.1%
- B) 0% and 15%
- C) 3.2% and 12.4%
- D) 0.9% and 14.7%

ANSWER: C

17. What is the relationship among the mean, median and mode in a positively skewed distribution?

- A) They are all equal
- B) The mean is always the smallest value
- C) The mean is always the largest value
- D) The mode is the largest value

ANSWER: C

18. The manager of a local RV sales lot has collected data on the number of RVs sold per month for the last five years. That data is summarized below

# of Sales	0	1	2	3	4	5	6
# of Months	2	6	9	13	21	7	2

What is the weighted mean number of sales per month?

- A) 3.31
- B) 3.23
- C) 3.54
- D) 3.62

ANSWER: B

19. A recent survey of Fortune 500 firms found that on average, they contribute \$332.54 per month for each salaried employee's health insurance. If you are told that almost all salaried employees at Fortune 500 firms receive a health insurance contribution between \$220.61 and \$444.47, what must the standard deviation for this data be?

- A) \$37.31
- B) \$46.65
- C) \$55.98
- D) \$74.64

ANSWER: A

Describing Data: Numerical

20. A bored carpenter counts the actual number of nails in 10 boxes of nails and records his findings as: 254, 275, 302, 287, 312, 323, 301, 319, 342, and 298. What can we say about the shape of the distribution of the number of nails?
- A) Symmetric
 - B) Skewed to the right.
 - C) Approximately mound-shaped.
 - D) Skewed to the left.

ANSWER: D

21. Which of the following statements is not true?
- A) Measures of central tendency are numbers that describe typical values in the data.
 - B) The coefficient of variation is the least used measure of central tendency.
 - C) The mean is the most widely used measure of location.
 - D) All of the above.

ANSWER: B

22. A professor collected data on the number of absences in an introductory statistics class of 100 students over the course of a semester. The data are summarized below.

# of Absences	0	1	2	3	4	5	6
# of Students	5	13	24	23	17	11	7

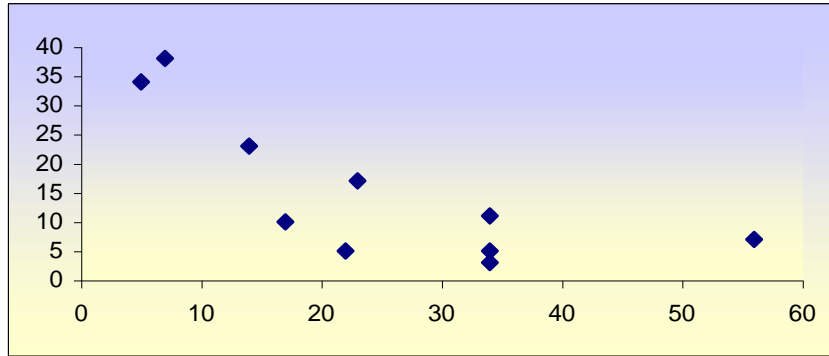
What is the weighted mean number of absences per semester?

- A) 3.14
 - B) 2.0
 - C) 2.95
 - D) 3.07
23. What is the relationship among the mean, median and mode in a negatively skewed distribution?
- A) They are all equal
 - B) The mean is always the smallest value
 - C) The mean is always the largest value
 - D) The mode is the largest value

ANSWER: B

Chapter 2

24. Looking at the scatter plot below, what value would be your best estimate for the correlation coefficient?



- A) -0.7
- B) -0.3
- C) -1.0
- D) 0.0

ANSWER: A

THE NEXT TWO QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

A recent survey asked respondents about their monthly purchases of lottery tickets. The monthly expenditures, in dollars, of ten people who play the lottery are 23, 15, 11, 20, 28, 35, 13, 10, 20, and 24.

25. What can we say about the shape of the distribution of monthly purchases of lottery tickets?
- A) Skewed to the left.
 - B) Skewed to the right.
 - C) Approximately mound-shaped.
 - D) None of the above.

ANSWER: C

26. Which of the following statements are not true?

- A) The 75th percentile is equal to 23.5.
- B) The median is equal to the mode.
- C) The mean is 19.9.
- D) The distribution is approximately symmetric.

ANSWER: A

Describing Data: Numerical

27. Over the past 10 years, the return on Stock A has averaged 8.4% with a standard deviation of 2.1%. The return on Stock B has averaged 3.6% with a standard deviation of 0.9%. Which of the following statements is true?
- A) Stock A has smaller relative variation than Stock B.
 - B) Stock B has smaller relative variation than Stock A.
 - C) Both stocks exhibit the same relative variation.
 - D) Unable to tell with the given information.

ANSWER: C

28. The median value of the data values 12, 32, 48, 8, 22, 9, 30, and 18 equals
- A) 20
 - B) 22
 - C) 24
 - D) 26

ANSWER: A

THE NEXT FIVE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

The police lieutenant in charge of the traffic division review the number of traffic citations issued by each of the police officers in his division. He finds that the mean number of citations written by each officer is 23.2 citations per day, with a standard deviation of 3.1. Assume that the distribution of the number of tickets issued is approximately mound-shaped.

29. Which of the following statements is true?
- A) Almost all of the officers wrote somewhere between 20.1 and 26.3 citations per day.
 - B) Almost all of the officers wrote more than 17 citations per day.
 - C) Almost all of the officers wrote less than 15 citations per day.
 - D) Approximately 95% of the officers wrote between 20.1 and 26.3 citations.

ANSWER: C

30. The coefficient of variation for the number of citations is:
- A) 13.36%
 - B) 7.48%
 - C) 6.68
 - D) Cannot be determined without the sample size.

ANSWER: A

Chapter 2

31. Suppose that you are also told that the median for these data was 19.3. Which of the following statements may be made about the shape of the distribution?

- A) It is skewed to the right.
- B) It is skewed to the left.
- C) It is approximately symmetric.
- D) Cannot be determined without more information.

ANSWER: A

32. What would be a reasonable estimate for the 75th percentile?

- A) Between 23.2 and 26.3
- B) Between 26.3 and 29.4
- C) Between 29.4 and 32.5
- D) Greater than 32.5

ANSWER: B

33. What would be a reasonable estimate for the 99th percentile?

- A) Between 23.2 and 26.3
- B) Between 26.3 and 29.4
- C) Between 29.4 and 32.5
- D) Greater than 32.5

ANSWER: C

34. What is the relationship among the mean, median and mode in a symmetrical distribution?

- A) They are all equal
- B) The mean is always the smallest value
- C) The mean is always the largest value
- D) The mode is the largest value

ANSWER: A

THE NEXT SIX QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

The police lieutenant in charge of the traffic division has reviewed the number of traffic citations issued per day by each of the 10 police officers in his division. The data were: 13, 21, 12, 34, 31, 13, 22, 26, 25, and 23.

35. What is the mean number of citations issued per day?

- A) 22.0
- B) 22.5
- C) 13.0
- D) 13.5

ANSWER: A

Describing Data: Numerical

36. What is the median number of citations issued per day?
- A) 22.0
 - B) 22.5
 - C) 13.0
 - D) 13.5
- ANSWER: B**
37. What is the mode of the number of citations issued per day?
- A) 22.0
 - B) 22.5
 - C) 13.0
 - D) 13.5
- ANSWER: C**
38. What is the first quartile of the number of citations issued per day?
- A) 22.0
 - B) 22.5
 - C) 13.0
 - D) 27.25
- ANSWER: C**
39. What is the third quartile of the number of citations issued per day?
- A) 22.0
 - B) 22.5
 - C) 13.0
 - D) 27.25
- ANSWER: D**
40. What would you conclude if the sample correlation coefficient is equal to -1.00?
- A) All the data points must fall exactly on a straight line with a positive slope.
 - B) All the data points must fall exactly on a straight line with a negative slope.
 - C) Most of the data points must fall exactly on a straight line with a positive slope
 - D) Most of the data points must fall exactly on a horizontal straight line
- ANSWER: B**

Chapter 2

41. The following stem-and-leaf output has been generated by Minitab.

```
Stem-and-leaf
N = 10
Leaf Unit = 0.10
2      - 1 53
4      - 0 97
(2)    - 0 65
4       0 3
3       0 6
2       1 3
1       1 8
```

Which of the following statements is true?

- A) This data set has a mean that is negative.
- B) This data set has a median less than -0.5
- C) This data set has six negative values
- D) All of the above are correct

ANSWER: D

42. Which of the following statements is true?

- A) Measures of variability are numbers that describe the scatter of the data or the extent to which the data values are spread out.
- B) The range is the most useful measure of variability.
- C) The weighted mean is the most useful measure of variability.
- D) All of the above.

ANSWER: A

43. Which of the following is not a measure of variability?

- A) Interquartile range
- B) Variance
- C) Weighted mean
- D) Range

ANSWER: C

44. The standard deviation of the sample data 13, 14, 17, and 20 equals

- A) 2.74
- B) 3.16
- C) 7.98
- D) 9.16

ANSWER: B

Describing Data: Numerical

45. The strength of the linear relationship between two numerical variables may be measured by the
- A) correlation coefficient
 - B) coefficient of variation
 - C) interquartile range
 - D) third quartile
- ANSWER: A**
46. Which of the following statements is not true?
- A) Quartiles divide the values into 4 parts of equal size, each comprising 25% of the observations.
 - B) Measures of variability describe typical values in the data.
 - C) The variance and standard deviation are the most useful statistical measures of variability.
 - D) The coefficient of variation is a measure of relative dispersion.
- ANSWER: B**
47. For any set of grouped or ungrouped data, which measures of central location always have only one value?
- A) Arithmetic mean and median
 - B) Median and mode
 - C) Mode and arithmetic mean
 - D) Geometric mean and mode
- ANSWER: A**
48. Which of the following statement is true?
- A) The range is found by taking the difference between the high and low values and dividing that value by 2.
 - B) The interquartile range is found by taking the difference between the 1st and 3rd quartiles and dividing that value by 2.
 - C) The standard deviation is expressed in terms of the original units of measurement but the variance is not.
 - D) The values of the standard deviation may be either positive or negative, while the value of the variance will always be positive.
- ANSWER: C**

Chapter 2

49. What is the lowest level of measurement that is needed for the median to be computed?

- A) Nominal
- B) Ordinal
- C) Interval
- D) Ratio

ANSWER: B

50. Which of the following statements is true?

- A) The mean is a measure of the deviation in a data set.
- B) The standard deviation is a measure of variability.
- C) The range is a measure of central location.
- D) The median is a measure of variability.

ANSWER: B

51. A sample of 15 observations has a standard deviation of 4. The sum of the squared deviations from the sample mean is:

- A) 19
- B) 56
- C) 60
- D) 224

ANSWER: D

52. Which of the following statements are *not* correct?

- A) The most useful measures of variability are based on deviations from the mean.
- B) The empirical rule applies to any distributions.
- C) The sum of $(x - \bar{x})$ will always be zero.
- D) For distributions that are bell-shaped and symmetric, approximately 68% of the observations will fall within one standard deviation of the mean.

ANSWER: B

53. According to the Empirical Rule, the percentage of observations in a data set that should fall within two standard deviations of their mean is approximately:

- A) 90%
- B) 95%
- C) 97.5%
- D) 100%

ANSWER: B

Describing Data: Numerical

54. The Empirical Rule states that the percentage of observations in a data set (providing that the data set has a bell-shaped and symmetric distribution) that fall within one standard deviation of their mean is approximately:

- A) 68%
- B) 75%
- C) 95%
- D) 99%

ANSWER: A

55. Which of the following is *not* a correct statement?

- A) The mean is a measure of central tendency.
- B) Chebychev's Theorem applies only to non-mounded distributions.
- C) The sum of $(x - \bar{x})$ will always be zero.
- D) The calculation of the range does not consider all values.

ANSWER: B

56. According to the Empirical Rule, if the distribution is mounded, then within one standard deviation of the mean, there will be approximately:

- A) 75% of the data.
- B) 85% of the data.
- C) 95% of the data.
- D) None of the data.

ANSWER: D

57. Which of the following statements about the median is not true?

- A) It is a measure of central tendency
- B) It is equal to the second quartile
- C) It is more affected by extreme values than the mean
- D) It is equal to the mean in bell-shaped distributions

ANSWER: C

58. For any distribution, the percent of observations that lie within four standard deviations of the mean is:

- A) 93.75% or more.
- B) 93.75% or less.
- C) 6.25% or more.
- D) 6.25% or less.

ANSWER: A

Chapter 2

59. For a sample of size 5, if $x_1 - \bar{x} = -5$, $x_2 - \bar{x} = 9$, $x_3 - \bar{x} = -7$, and $x_4 - \bar{x} = -2$, then the sample standard deviation is

- A) 5.639
- B) 6.782
- C) 6.066
- D) 6.305

ANSWER: B

60. For any bell-shaped and symmetric distribution,

- A) the mean equals the median
- B) the median equals the mode
- C) the mode equal the mean
- D) All of the above

ANSWER: D

61. The covariance of the following sample data of four (X, Y) pairs: $(1, 5)$, $(2, 10)$, $(4, 7)$, and $(5, 9)$ equals

- A) 1.25
- B) 2.50
- C) 3.75
- D) 3.69

ANSWER: A

62. A random sample from an unknown population had a sample standard deviation of zero. Which of the following is a reasonable conclusion?

- A) The sample range must be zero.
- B) An error was made in computing the sample standard deviation. It must always be greater than zero.
- C) The population standard deviation must be zero.
- D) None of the above

ANSWER: A

Describing Data: Numerical

THE NEXT FOUR QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

The following data represent a sample of 10 scores on a statistics quiz: 16, 16, 16, 16, 16, 18, 18, 20, 20, and 20.

63. The mean score is

- A) 17.4
- B) 15.8
- C) 12.2
- D) 10.4

ANSWER: C

64. The median score is

- A) 16
- B) 17
- C) 18
- D) 19

ANSWER: B

65. The modal score is

- A) 16
- B) 17
- C) 18
- D) 20

ANSWER: A

66. The standard deviation of the scores is

- A) 4.682
- B) 5.174
- C) 6.720
- D) 7.083

ANSWER: D

67. The range of the scores is

- A) 4
- B) 6
- C) 8
- D) 9

ANSWER: A

Chapter 2

68. Which of the following is used as a divisor in the sample variance s^2 , where n is the sample size?

- A) $n+1$
- B) n
- C) $n-1$
- D) $n-2$

ANSWER: C

69. Which of the following represents a disadvantage of using the sample range to measure spread or dispersion?

- A) It produces spreads that are too large.
- B) The sample range is not measured in the same units as the data.
- C) The largest or smallest observation (or both) may be an outlier.
- D) None of the above is correct.

ANSWER: C

70. The correlation coefficient of the following sample data of four (X, Y) pairs: (1, 5), (2, 10), (4, 7), and (5, 9) equals

- A) 0.263
- B) 0.412
- C) 0.528
- D) 0.364

ANSWER: B

71. The following ten scores were obtained on a 20-point quiz: 4, 5, 8, 9, 11, 13, 15, 18, 18, and 20. The teacher computed the usual descriptive measures of center (central tendency) and variability (dispersion) for these data, and then discovered an error was made. One of the 18's should have been a 16. Which of the following measures, calculated on the corrected data, would change from the original computation?

- A) Median
- B) Mean
- C) Standard deviation
- D) Both (B) and (C)

ANSWER: D

Describing Data: Numerical

72. A college placement office conducted a survey of 100 engineers who had graduated from Stanford University. For these engineers, the mean salary was computed to be \$72,000 with a standard deviation of \$8,000. The percentage of these engineers who earn more than \$96,000 or less than \$48,000 is

- A) Approximately 0%.
- B) At least 5.6% (1/18 of the engineers).
- C) At most 5.6% (1/18 of the engineers).
- D) At most 11.1% (1/9 of the engineers).

ANSWER: D

73. For which measures of central location will the sum of the deviations of each value from the data's average will always be zero?

- A) Arithmetic mean
- B) Geometric mean
- C) Median
- D) Mode

ANSWER: A

74. Which one of the values below represents a lower quartile for the data set 23, 24, 21, and 20?

- A) 22.0
- B) 22.5
- C) 20.25
- D) 23.5

ANSWER: C

75. Which of the following statements is true for the following data values: 17, 15, 16, 14, 17, 18, and 22?

- A) The mean, median and mode are all equal
- B) Only the mean and median are equal
- C) Only the mean and mode are equal
- D) Only the median and mode are equal

ANSWER: A

76. What is the smallest measure of central tendency in a positively skewed distribution?

- A) The arithmetic mean
- B) The median
- C) The mode
- D) Can't be determined

ANSWER: C

Chapter 2

77. Which of the following measures of dispersion are based on deviations from the mean?

- A) Standard deviation
- B) Variance
- C) Range
- D) Both A and B

ANSWER: D

78. At a highway checkpoint, the average speed of a passing car was measured at 50 miles per hour with a standard deviation of 5 miles per hour. According to Chebychev's Theorem, what percentage of cars would you expect to be traveling between 42.5 and 57.5 miles per hour?

- A) At least 50%
- B) At least 55.6%
- C) At least 75%
- D) At least 88.9%

ANSWER: B

79. Which of the following statements is true?

- A) The sum of the deviations from the mean is always zero
- B) The sum of the squared deviations from the mean is always zero
- C) The mean is always less than the median
- D) The standard deviation is always smaller than the variance

ANSWER: A

80. Which one of the values below represents the third quartile of the data set 10, 12, 16, 7, 9, 7, 41, and 14?

- A) 8.0
- B) 15.5
- C) 7.0
- D) 24.0

ANSWER: B

81. Expressed in percentiles, the interquartile range is the difference between the

- A) 30% and 80% values.
- B) 45% and 95% values.
- C) 25% and 75% values.
- D) 20% and 70% values.

ANSWER: C

82. What is the median of 36, 40, 37, 42, 45, 41, 34, and 39?

Describing Data: Numerical

- A) 39
- B) 39.5
- C) 40
- D) 41

ANSWER: B

83. If two data sets have the same range, then:

- A) the distances from the smallest to largest observations in both sets will be the same
- B) the smallest and largest observations are the same in both sets
- C) both sets will have the same variance
- D) both sets will have the same interquartile range

ANSWER: A

84. A sample of eight retired persons receiving social security payments revealed the following monthly benefits: \$985, \$798, \$1,210, \$1,356, \$1,087, \$869, \$987, and \$1,045. How many observations are below the median?

- A) 2
- B) 3
- C) 4
- D) 4.5

ANSWER: C

85. For a data set with 10 numerical values arranged in an ascending order, the median is the arithmetic mean of the

- A) third and fourth values
- B) fourth and fifth values
- C) fifth and sixth values
- D) first and tenth values

ANSWER: C

86. Since the population is always larger than the sample, the population mean:

- A) is always larger than the sample mean
- B) is always smaller than the sample mean
- C) is always larger than or equal to or smaller than or equal to the sample mean
- D) can be smaller than, or larger than, or equal to the sample mean

ANSWER: D

87. The average score for a class of 35 students was 70. The 20 male students in the class averaged 73. What was the average score for the 15 female students in the class?

Chapter 2

- A) 73
B) 70
C) 66
D) 60
ANSWER: C
88. Which of the following summary measures is affected most by outliers?
- A) The first quartile
B) The second quartile
C) The third quartile
D) None of the above
ANSWER: D
89. When extreme values are present in a set of data, which pair of the following descriptive summary measures of central tendency and dispersion is most appropriate?
- A) Mean and standard deviation
B) Median and interquartile range
C) Range and coefficient of variation
D) Mode and variance
ANSWER: B
90. Which measures of central tendency are not affected by extremely small or extremely large values?
- A) Arithmetic mean and median
B) Arithmetic mean and mode
C) Arithmetic mean and geometric mean
D) Median and mode
ANSWER: D
91. If a distribution is highly skewed, what measure of central tendency should be avoided?
- A) Arithmetic mean
B) Median
C) Mode
D) All of the above
ANSWER: A
92. A question in a survey asks for a respondent's favorite sport. Which measure of central tendency should be used to summarize this question?

Describing Data: Numerical

- A) Arithmetic mean
- B) Geometric mean
- C) Median
- D) Mode

ANSWER: D

93. According to Chebychev's Theorem, at least what percent of the observations lie within ± 1.25 standard deviations of the mean?

- A) 36%
- B) 56%
- C) 76%
- D) 96%

ANSWER: A

94. Which measure of central location is used to determine an average annual percent increase?

- A) Arithmetic mean
- B) Weighted mean
- C) Geometric mean
- D) Median

ANSWER: C

95. The five-number summary includes all of the following except the

- A) first quartile
- B) second quartile
- C) third quartile
- D) mode

ANSWER: D

96. In the calculation of the arithmetic mean for grouped data, which value is used to represent all the values in a particular class?

- A) The lower limit of the class
- B) The upper limit of the class
- C) The frequency of the class
- D) The midpoint of the class

ANSWER: D

97. A sample of college students revealed their last month income as follows: \$765, \$680, \$623, \$980, \$875, and \$985. How many observations are below the median?

- A) 1

Chapter 2

- B) 2
- C) 3
- D) 4

ANSWER: C

98. A question in a market survey asks for a respondent's favorite car model. Which measure of central location should be used to summarize this question?

- A) Arithmetic mean
- B) Geometric mean
- C) Median
- D) Mode

ANSWER: D

99. According to Chebychev's Theorem, what percent of the observations lie within ± 2.25 standard deviations of the mean?

- A) At least 80.25%
- B) At least 75.25%
- C) At least 55.56%
- D) At least 95.25%

ANSWER: A

True-False Questions

100. The sample covariance must take a value between -1 and $+1$ inclusive.
ANSWER: F
101. The sample covariance may never be negative.
ANSWER: F
102. The correlation coefficient measures the strength of a linear relationship between two variables.
ANSWER: T
103. A correlation coefficient of zero indicates a lack of relationship between the two variables of interest.
ANSWER: F
104. The value of the correlation coefficient may be used to confirm a non-linear relationship.
ANSWER: F
105. The mean is generally the preferred measure of central tendency to describe numerical data, but not categorical data.
ANSWER: T
106. For any set of numerical data values arranged in an ascending or descending order, the value of the observation in the center is called the weighted mean.
ANSWER: F
107. Categorical data are best described by the mode or the median.
ANSWER: T
108. The median should always be preferred to the mean when the population or sample is skewed to the right or left.
ANSWER: F
109. One possible source of skewness is the presence of outliers, and sometimes skewness is simply inherent in the distribution.
ANSWER: T
110. Although the range measures the total spread of the data, the interquartile range (IQR) measures only the spread of the middle 50% of the data.
ANSWER: T
111. In a negatively skewed distribution, the mean is always greater than the median.
ANSWER: F

Chapter 2

112. The coefficient of variation measures variability in a positively skewed data set relative to the size of the median.
ANSWER: F
113. When the data values are arranged in an ascending order, the third quartile (Q_3) is located in the $0.75(n+1)^{\text{th}}$ position, and first quartile (Q_1) is located in the $0.25(n+1)^{\text{th}}$ position.
ANSWER: T
114. The five – number summary refers to the five descriptive measures: minimum, mean, median, mode, and maximum; therefore it is sometimes known as the five–m summary.
ANSWER: F
115. If the interquartile range for a set of data is 10 minutes, this means that the data have a spread of only 10 minutes.
ANSWER: F
116. Since the interquartile range takes into account only two of the data values, it is susceptible to considerable distortion if there is an unusual number of extreme observations (outliers).
ANSWER: F
117. Although range and interquartile range measure the spread of data, both measures take into account only two of the data values, regardless of the size of the data.
ANSWER: T
118. For any symmetrical distribution, the standard deviation is equal to the variance.
ANSWER: F
119. If the population variance σ^2 is unknown, a sample variance s^2 is a better estimator of σ^2 if the denominator in s^2 formula is $(n-1)$ rather than n .
ANSWER: T
120. For any distribution, the number of values above the mean and below it is the same.
ANSWER: F
121. Suppose that the average score on an exam is 73 with a standard deviation of 2. According to Chebychev’s theorem, at least 60% of the scores are in the interval between 70 and 76.
ANSWER: F
122. The advantage of Chebychev’s theorem is that its applicability extends to any population regardless of its shape. However, it is within this guarantee that its major drawback lies.
ANSWER: T

Describing Data: Numerical

123. For any distribution, the empirical rule estimates that approximately 95% of the observations will fall within two standard deviations of the mean.
ANSWER: F
124. The coefficient of variation (CV) is a measure of relative dispersion that expresses the standard deviation as a percentage of the mean (provided the mean is positive).
ANSWER: T
125. If the mean is greater than the median, then the distribution is skewed to the right.
ANSWER: T
126. Any set of ordinal, interval or ratio level data may only have one mode.
ANSWER: F
127. The median and the 50th percentile are always equal.
ANSWER: T
128. The interquartile range measures the spread of the lower 50% of data values.
ANSWER: F
129. The mode is the most useful measure of central tendency if the data is ordinal.
ANSWER: T
130. Suppose you have a set of data which has a mound-shaped histogram. Compare the inter-quartile range with the range from one standard deviation below the mean to one standard deviation above the mean. The inter-quartile range is larger.
ANSWER: T
131. In a positively skewed distribution, the mode is greater than the median.
ANSWER: F
132. A variable measured at the interval or ratio level can have more than one arithmetic mean.
ANSWER: F
133. Consider two possible investments with the same expected rate of return. Over the past several months, investment A has had an average closing price of \$14.00 and a standard deviation of \$4.00. Investment B has had an average closing price of \$58.00 and a standard deviation of \$15.00. The market value of investment A fluctuates relatively more than investment B.
ANSWER: T
134. For salaries of \$108,000, \$102,000, \$52,000, \$105,000, 107,000 and \$101,000, the arithmetic mean would be an appropriate measure of central tendency.
ANSWER: F

Chapter 2

135. The mean is a better measure of central tendency than the median when there are outliers.
ANSWER: F
136. The variance of a set of data can never be negative.
ANSWER: T
137. Quartiles and percentiles are two of the most popular measures of dispersion.
ANSWER: F
138. The median, the second quartile, and the 50th percentile are all the same.
ANSWER: T
139. The first quartile, Q_1 , is a number such that at most 25 of the data values are smaller in value than Q_1 and at most 75 of the data values are larger.
ANSWER: F
140. The interquartile range is very unique in the sense that it is a measure of central tendency as well as a measure of dispersion.
ANSWER: F
141. Percentiles are values of the variable that divide a set of ranked data into 100 equal subsets.
ANSWER: T
142. Each set of data has 100 percentiles.
ANSWER: F
143. The 30th percentile, P_{30} , is a value such that at most 30% of the data are smaller in value than P_{30} and at most 70% of the data are larger.
ANSWER: T
144. The first quartile and the 25th percentile are the same.
ANSWER: T
145. For any data set, the variance is the average of the sum of the squared deviations between each observation and the median.
ANSWER: F
146. The mean, median, the second quartile, and the 50th percentile are all the same.
ANSWER: F
147. The interquartile range is the average of the first and third quartiles.
ANSWER: F

Describing Data: Numerical

148. The 5-number summary divides a set of data into four subsets, with one-quartile of the data in each subset.
ANSWER: T
149. The interquartile range measures the spread of the middle 50% of the data.
ANSWER: T
150. *Chebychev's* Theorem says that within two standard deviations of the mean, you will always find at least 88.9% of the data.
ANSWER: F
151. The standard deviation, as a measure of variation (dispersion), can be understood by examining two statements that tell us how the standard deviation relates to the data: the Empirical Rule and Chebychev's Theorem.
ANSWER: T
152. If the mean of a quantitative data set exceeds the median, the data are considered to be symmetrical.
ANSWER: F
153. The Empirical rule is frequently applied as an interpretive guide to any mounded distribution.
ANSWER: T
154. The Empirical rule applies to any distribution, regardless of its shape, as an interpretive guide to the distribution.
ANSWER: F
155. A student scores 89, 75, 94, and 88 on four exams during the semester and 97 on the final exam. If the final is weighted double and the four others weighted equally, the student's final average would be 90.
ANSWER: T
156. In a mound-shaped distribution, there is no difference in the values of the mean, and median.
ANSWER: T
157. A variable measured at the interval or ratio level can only have one arithmetic mean.
ANSWER: T
158. The covariance measures the direction and strength of any relationship between two variables.
ANSWER: F

Chapter 2

159. In an accounting exam, the standard deviation of the scores of female students is six and the standard deviation of the scores of male students is ten. These statistics indicate that there is more spread in the scores of the female students.
ANSWER: F
160. Three persons earn \$9 an hour, five earn \$10 an hour, and one earns \$13 an hour. The weighted mean hourly wage is \$10.
ANSWER: T
161. A distribution that has the same shape on either side of the center is said to be symmetrical.
ANSWER: T
162. For a set of numerical data values arranged in ascending order, the value of the observation in the center is called the geometric mean.
ANSWER: F
163. The geometric mean of a set of 10 positive numbers is the 10th root of the product of the 10 values.
ANSWER: T
164. A negatively skewed distribution is not symmetrical. The long tail is to the right.
ANSWER: F
165. If the mean of a symmetrical data set is less the median, the data are considered to be negatively skewed.
ANSWER: T
166. Measures of central tendency provide numerical information about a “typical” observation in the data.
ANSWER: T
167. For a set of numerical data, the geometric mean is the n th root of the sum of the n observations.
ANSWER: F
168. The median can be determined for any set of ordinal, interval or ratio-level data.
ANSWER: T
169. Extremely small or large values in a data set affect the value of the median as well as the mode.
ANSWER: F
170. The sum of the deviations from the mean and the median for the set of numbers 1, 3, and 5 will equal zero.
ANSWER: T

Describing Data: Numerical

171. For any distribution, there are an equal number of observations above and below the mean.

ANSWER: F

Chapter 2

Short Answer and Applied Questions

172. Why do we have so many different measures of central tendency? Are all really necessary or do they essentially provide the same information?

ANSWER:

For certain types of well-behaved data, all the measures will provide very similar information. However, because much data is not well-behaved, it is often beneficial to look at a number of measures of central tendency and report the one that best describes the location and average of the data. Wherever possible, the mean is the preferred measure of location, because it uses all the data values. However, for certain types of data that contain severe outliers (like income data), the median is the preferable measure of central tendency because the outliers do not distort it.

173. Why is it necessary for a measure of variation to accompany a measure of central tendency?

ANSWER:

A measure of central tendency alone does not give a complete picture of the data set. The object of summary measures is to visualize the data set based on these measures. Hence, the measure of central tendency locates the data set, but a measure of variation completes the picture by describing the dispersion in the data about the location measure.

174. In spite of its advantage in discounting extreme observations, the median is used less frequently than the mean. Why?

ANSWER:

In spite of its advantage in discounting extreme observations, the median is used less frequently than the mean. The reason is that the theoretical development of inferential procedures based on the mean, and measures related to it, is considerably more straightforward than the development of procedures based on the median.

175. Give an example to illustrate that the median should not always be preferred to the mean when the population or sample is skewed.

ANSWER:

There are times when the mean would still be the preferred measure even if the distribution were skewed. Consider an insurance company that most likely faces a right-skewed distribution of claims. If the company wants to know the most typical claim size, the median is preferred. However, suppose the company wants to know how much money needs to be budgeted to cover claims. Then the mean is preferred.

Describing Data: Numerical

176. The management at a small manufacturing plant has noticed that the price of steel has increased significantly over the past several years. Looking over their records, they find that over the four-year period, prices have increased by 40%. They expect this same trend in prices for next year. In budgeting for next year, by how much should they expect prices to increase?

ANSWER:

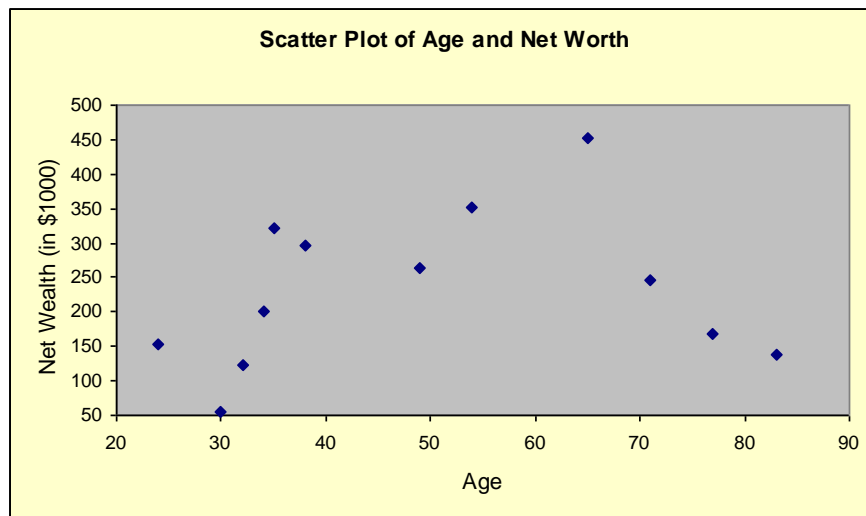
$$(1 + K)^4 = 1.4 \Rightarrow 1 + K = (1.4)^{0.25} \Rightarrow K = 0.0878 \text{ or } 8.78\%$$

THE NEXT FOUR QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

A researcher is interested in examining how the net wealth of individuals changes over the course of their lifetimes. She has collected the following data regarding the age X , in years, and net worth Y , measured in thousands of dollars, of 12 individuals in the form of (X, Y) pairs: (24, 153), (34, 201), (38, 297), (83, 139), (77, 167), (32, 123), (71, 247), (49, 263), (54, 352), (35, 321), (65, 453), and (30, 54).

177. Prepare a scatter plot of this data.

ANSWER:



178. What conclusions can you draw about the relationship between age and net wealth based on the scatter plot in the previous question?

ANSWER:

It appears that as you get older, at first your wealth increases until the age of 65, then starts to decrease.

179. Calculate the correlation coefficient between the age and net worth of individuals.

ANSWER:

$$r = 0.207$$

Chapter 2

180. Use your answer to one of the previous questions to determine if there is a relationship between the age and net worth of individuals.

ANSWER:

A useful rule of thumb is that a relationship exists if $|r| \geq 2/\sqrt{n}$. Since $2/\sqrt{n} = 2/\sqrt{12} = 0.577$ and $|r| = 0.207$, we may conclude that no relationship exists between the age and net worth of individuals.

THE NEXT TWO QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

The police lieutenant in charge of the traffic division has reviewed the number of traffic citations issued per day by each of the 10 police officers in his division. The data were: 13, 21, 12, 34, 31, 13, 22, 26, 25, and 23.

181. What is the standard deviation for the number of citations issued per day?

ANSWER:

$$s^2 = \sum (x - \bar{x})^2 / (n - 1) = 514 / 9 = 51.11 \Rightarrow s = 7.56$$

182. What is the inter-quartile range for the number of citations issued per day?

ANSWER:

Location of $Q_3 = 0.75(n+1) = 0.75(11) = 8.25$; Value of $Q_3 = 26 + 0.25(31 - 26) = 27.25$

Location of $Q_1 = 0.25(n+1) = 0.25(11) = 2.75$; Value of $Q_1 = 13 + 0.75(3 - 3) = 13.0$

$IQR = Q_3 - Q_1 = 27.25 - 13.0 = 14.25$

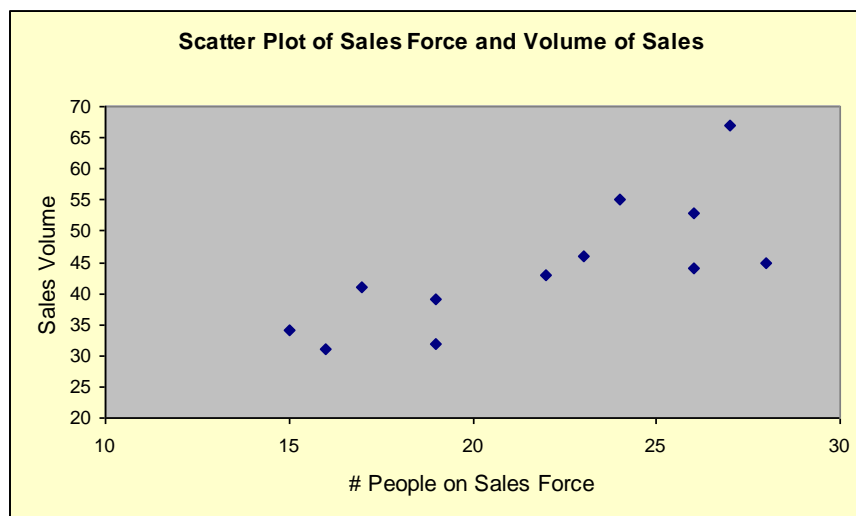
Describing Data: Numerical

THE NEXT FOUR QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

You are interested in looking at the relationship between the number of people on the sales force and the dollar volume of sales. The following data show gross sales, Y , measured in millions of dollars, and the number of people on the sales force, X , in the form of (X, Y) pairs for 12 people: (15, 34), (24, 55), (27, 67), (16, 31), (19, 32), (26, 44), (19, 39), (23, 46), (26, 53), (22, 43), (28, 45), and (17, 41).

183. Prepare a scatter plot of this data.

ANSWER:



184. What conclusions can you draw about the relationship between the size of the sales force and gross sales based on the scatter plot in the previous question?

ANSWER:

In general, there is a positive relationship between the two variables; that is, as the number of people on the sales force increases, so do sales.

185. Calculate the correlation coefficient between the size of the sales force and gross sales.

ANSWER:

$$r = 0.765$$

Chapter 2

186. Use your answer to one of the previous questions to determine if there is a relationship between the size of the sales force and gross sales.

ANSWER:

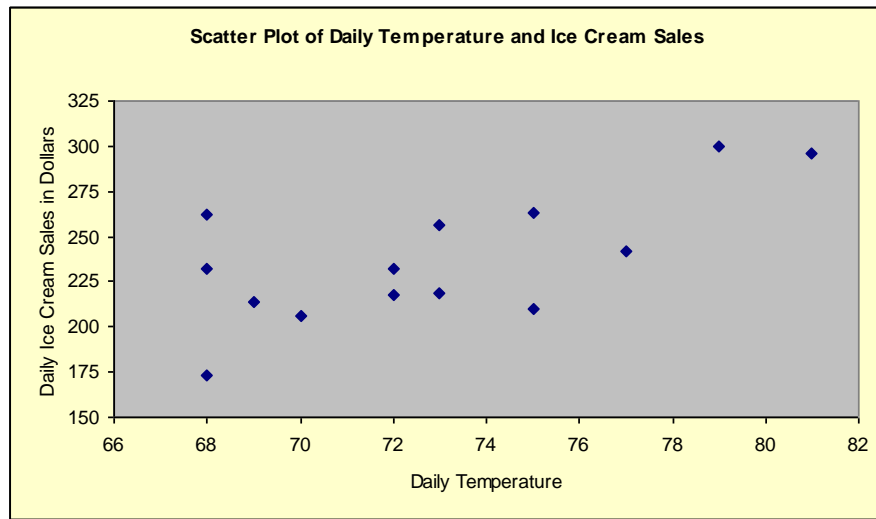
A useful rule of thumb is that a relationship exists if $|r| \geq 2/\sqrt{n}$. Since $2/\sqrt{n} = 2/\sqrt{12} = 0.577$, and $|r| = 0.765$, we may conclude that a positive relationship exists between the size of the sales force and gross sales.

THE NEXT FOUR QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

Ahmed is trying to figure out the relationship between daily sales from his ice cream truck and the daily temperature in Dearborn, Michigan during the month of August. He has collected data on sales Y , measured in dollars, and temperature X , measured in Fahrenheit degrees, over the past 14 days in the form of (X, Y) pairs: (72, 232), (77, 242), (73, 219), (69, 214), (68, 262), (72, 218), (75, 263), (70, 206), (79, 300), (73, 256), (68, 173), (75, 210), (81, 296), and (68, 232).

187. Prepare a scatter plot of this data.

ANSWER:



188. What conclusions can you draw about the relationship between ice cream sales and daily temperature based on the scatter plot in the previous question?

ANSWER:

There appears to be a slight positive relationship between daily temperature and ice cream sales.

189. Calculate the correlation coefficient between ice cream sales and temperature.

ANSWER:

$r = 0.679$

Describing Data: Numerical

190. Use your answer to one of the previous questions to determine if there is a relationship between ice cream sales and temperature.

ANSWER:

A useful rule of thumb is that a relationship exists if $|r| \geq 2/\sqrt{n}$. Since $2/\sqrt{n} = 2/\sqrt{14} = 0.535$, and $|r| = 0.679$, we may conclude that a positive relationship exists between ice cream sales and temperature.

THE NEXT SIX QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

A set of data is mound, with a mean of 500 and a variance of 576.

191. Approximately what proportion of the observations is greater than 476?

ANSWER:

The value 476 is one standard deviation below the mean; hence the empirical rule implies that the area between 476 and the mean is approximately $0.68 / 2 = 0.34$. Therefore the proportion of the observations greater than 476 is approximately 0.84.

192. Approximately what proportion of the observations is less than 548?

ANSWER:

The value 548 is two standard deviations above the mean; hence the Empirical Rule implies that the area between the mean and 548 is approximately $0.95 / 2 = 0.475$. Therefore the proportion of the observations less than 548 is approximately 0.975.

193. Approximately what proportion of the observations is greater than 572?

ANSWER:

The value 572 is three standard deviations above the mean; hence the Empirical Rule implies that the area between the mean and 572 is approximately $1.00 / 2 = 0.50$. Therefore the proportion of the observations greater than 572 is approximately 0.0.

194. Approximately what proportion of the observations is between 452 and 548?

ANSWER:

The values 452 and 548 are within two standard deviations of the mean; hence the Empirical Rule implies that the proportion of the observations between these two values is approximately 0.95

Chapter 2

195. Approximately what proportion of the observations is between 428 and 572?

ANSWER:

The values 428 and 572 are within three standard deviations of the mean; hence the Empirical Rule implies that the proportion of the observations between these two values is approximately 1.0. That is, almost all the observations are contained between 428 and 572.

196. Approximately what proportion of the observations is between 476 and 524?

ANSWER:

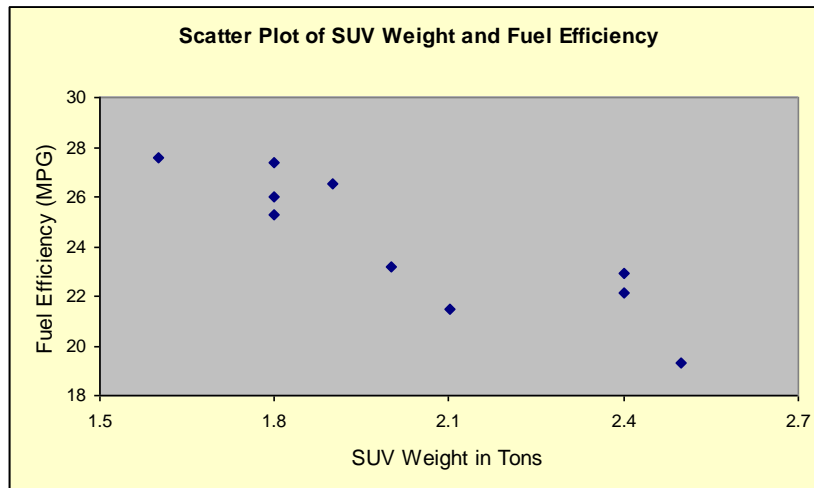
The values 476 and 524 are within one standard deviation of the mean; hence the Empirical Rule implies that the proportion of the observations between these two values is approximately 0.68

THE NEXT FOUR QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

Some people would argue that the trend of consumers toward purchasing large sport utility vehicles (SUVs) has been detrimental to the environment. The following data show the vehicle weight X , measured in tons, and the corresponding fuel efficiency Y , measured by the average miles per gallon (mpg), in the form of (X,Y) pairs for 10 vehicles: (1.9, 26.5), (2, 23.2), (2.4, 22.1), (1.8, 26), (2.4, 22.9), (2.1, 21.5), (1.8, 25.3), (2.5, 19.3), (1.8, 27.4), and (1.6, 27.6).

197. Prepare a scatter plot of this data.

ANSWER:



198. What conclusions can you draw about the relationship between weight and fuel efficiency based on the scatter plot in the previous question?

ANSWER:

In general, there is a negative relationship between the two variables; that is, as the weight of the SUV increases, the gas mileage per gallon (SUV efficiency) decreases.

Describing Data: Numerical

199. Calculate the correlation coefficient between vehicle weight and fuel efficiency.

ANSWER:

$$r = -0.887$$

200. Use your answer to one of the previous questions to determine if there is a relationship between vehicle weight and fuel efficiency.

ANSWER:

A useful rule of thumb is that a relationship exists if $|r| \geq 2/\sqrt{n}$. Since $2/\sqrt{n} = 2/\sqrt{10} = 0.632$ and $|r| = 0.887$, we may conclude that a negative relationship exists between vehicle weight and fuel efficiency.

THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

Chebychev's theorem is used to approximate the proportion of observations for any data set, regardless of the shape of the distribution. Assume that a distribution has a mean of 255 and standard deviation of 20.

201. Approximately what proportion of the observations is between 195 and 315?

ANSWER:

$K = 3$; hence at least 88.9% of the observations are in the interval between 195 and 315.

202. Approximately what proportion of the observations is between 210 and 290?

ANSWER:

$K = 2$; hence at least 75% of the observations are in the interval between 215 and 295.

203. Approximately what proportion of the observations is between 225 and 285?

ANSWER:

$K = 1.5$; hence at least 55.6% of the observations are in the interval between 225 and 285.

THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

The annual percentage returns on two stocks over a 7-year period were as follows:

Stock A:	4.01%	14.31%	19.01%	-14.69%	-26.49%	8.01%	5.81%
Stock B:	6.51%	4.41%	3.81%	6.91%	8.01%	5.81%	5.11%

204. Compare the means of these two population distribution.

ANSWER:

$$\mu_A = 1.89\% \text{ and } \mu_B = 5.80\%$$

205. Compare the standard deviations of these two population distributions.

Chapter 2

ANSWER:

$$\sigma_A = 0.151\% \text{ and } \sigma_B = 0.0147$$

206. Compute an appropriate measure of dispersion for both stocks to measure the risk of these investment opportunities. Which stock is more volatile?

ANSWER:

The coefficients of variation are computed for both stocks to measure and compare the risk of these two investment opportunities. Since $CV_A = 7.989\%$ and $CV_B = 0.253\%$, we conclude that stock A is more volatile than stock B.

207. Calculate the coefficient of variation for the following sample data: 13.2, 14.7, 17.2, 12.1, 21.8, 8.4, 14.3, 11.0, 9.3, and 8.7

ANSWER:

$$\bar{x} = \sum x/n = 13.07$$

$$s^2 = \sum (x - \bar{x})^2 / (n - 1) = 17.5334 \Rightarrow s = 4.1873$$

$$CV = (s/\bar{x}) \cdot 100\% = (4.1879/13.07) \cdot 100\% = 32.04\%$$

THE NEXT TWO QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

The following numbers represent the distance, in miles, that randomly selected ten employees of a firm must travel each way to work from home: 6.5, 14.8, 18.6, 6.5, 17.4, 12.3, 1.9, 12.9, 11.1, and 8.0.

208. The mean number of miles driven by the ten employees,

ANSWER:

11.0.

209. The standard deviation of the number of miles driven by the ten employees is

ANSWER:

5.28.

Describing Data: Numerical

THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

In a recent survey, 12 students at a local university were asked approximately how many hours per week they spend on the Internet. Their responses were: 13, 0, 5, 8, 22, 7, 3, 0, 15, 12, 13, and 17.

210. What are the mean and standard deviation for this data?

ANSWER:

$$\bar{x} = \sum x / n = 9.58$$

$$s^2 = (\sum (x - \bar{x})^2) / (n - 1) = 47.7197 \Rightarrow s = 6.91$$

211. What is the coefficient of variation for this data?

ANSWER:

$$CV = (s / \bar{x}) \cdot 100\% = (6.91 / 9.58) \cdot 100\% = 72.13\%$$

212. From the data presented above, calculate the inter-quartile range.

ANSWER:

Location of $Q_3 = 0.75(n+1) = 0.75(13) = 9.75$; Value of $Q_3 = 13 + 0.75(15 - 13) = 14.5$

Location of $Q_1 = 0.25(n+1) = 0.25(13) = 3.25$; Value of $Q_1 = 3 + 0.25(5 - 3) = 3.5$
 $IQR = Q_3 - Q_1 = 14.5 - 3.5 = 11.0$

THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

In a recent survey, 200 top executives were asked how many hours they spend each year in community service. The data are presented below.

# of Hours	0 but < 20	20 but < 40	40 but < 60	60 but < 80	80 but < 100	100 but < 120	120 but < 140
# of Executives	11	27	33	53	47	22	7

213. Calculate the quantities $\sum f_i$, $\sum f_i m_i$, and $\sum f_i (m_i - \bar{x})^2$.

ANSWER:

$$\sum f_i = n = 200, \sum f_i m_i = 13,840, \sum f_i (m_i - \bar{x})^2 = 175,072.$$

214. What is the estimated mean amount of time spent by these executives in community service?

ANSWER:

$$\bar{x} = \sum f_i m_i / n = 13,840 / 200 = 69.2 \text{ hours.}$$

Chapter 2

215. What is the estimated standard deviation for this data?

ANSWER:

$$s^2 = \sum f_i (m_i - \bar{x})^2 / (n-1) = 175,072 / 199 = 879.7588$$

Then $s = 29.66$ hours.

THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

The number of students eating breakfast at the school dining commons was recorded over 110 days last semester. These data are presented below.

# of Students	160 but < 190	190 but < 220	220 but < 250	250 but < 280	280 but < 310
# of Days	11	27	42	23	7

216. Calculate the quantities $\sum f_i$, $\sum f_i m_i$, and $\sum f_i (m_i - \bar{x})^2$.

ANSWER:

$$\sum f_i = n = 110, \sum f_i m_i = 25,490, \sum f_i (m_i - \bar{x})^2 = 108,621.819.$$

217. What is the estimated mean number of students showing up for breakfast?

ANSWER:

$$\bar{x} = \sum f_i m_i / n = 25,490 / 110 = 231.73.$$

218. What is the estimated standard deviation for this data?

ANSWER:

$$s^2 = \sum f_i (m_i - \bar{x})^2 / (n-1) = 108,621.819 / 109 = 996.53 \Rightarrow s = 31.57.$$

THE NEXT FIVE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

A small accounting office is trying to determine its staffing needs for the coming tax season. The manager has collected the following data: 46, 27, 79, 57, 99, 75, 48, 89, and 85. These values represent the number of returns the office completed each year over the past nine years it has been doing tax returns:

219. For this data, what is the mean number of tax returns completed each year?

ANSWER:

$$\mu = \sum x / N = 605 / 9 = 67.22$$

Describing Data: Numerical

220. For this data, what is the median number of tax returns completed each year?

ANSWER:

Ranked data: 27, 46, 48, 57, 75, 79, 85, 89, 99

Location of median = $0.50(n+1) = 0.50(10) = 5^{\text{th}}$ position. Hence, median = 75

221. For this data, what is the variance of the number of tax returns completed each year?

ANSWER:

$$\sigma^2 = \sum (x - \mu)^2 / N = 4541.56/9 = 504.62$$

222. For this data, what is the interquartile for the number of tax returns completed each year?

ANSWER:

Location of $Q_1 = 0.25(n+1) = 0.25(10) = 2.5$; Value of $Q_1 = (45+48)/2 = 47$

Location of $Q_3 = 0.75(n+1) = 0.75(10) = 7.5$; Value of $Q_3 = (85+89)/2 = 87$

$$IQR = Q_3 - Q_1 = 87 - 47 = 40$$

223. For this data, what is the coefficient of variation for the number of tax returns completed each year?

ANSWER:

$$\sigma = \sqrt{504.62} = 22.46$$

$$CV = (\sigma / \mu) \cdot 100\% = (22.46/67.22) \cdot 100\% = 33.41\%$$

224. Compute the mean, standard deviation and inter-quartile range for the following sample data: 12.2, 15.9, 8.1, 19.2, 13.7, 7.2, 12.2, 10.9, 5.4, and 16.8.

ANSWER:

Mean = 12.16, standard deviation = 4.42, and $IQR = Q_3 - Q_1 = 16.125 - 7.875 = 8.25$.

225. The manager of 45 sales people examined their monthly expenditures on entertaining clients. He found that the mean amount was \$237.50 with a standard deviation of \$27.40. Based on this data, would a claim for the amount of \$300 be considered unlikely? Why or why not?

ANSWER:

We would expect that virtually all of the sales people had expenses within 3 standard deviations of the mean. Although not very likely, it is possible that a sales person will have expenses in the amount of \$300. Less than 5% of the sales people would have sales expenses of this amount or more.

Chapter 2

226. A researcher interested in determining the average monthly expenditures of college students on DVDs finds that for a sample of 25 students, the mean expenditure was \$24.40, and the median expenditure was \$21.76. Specify the shape of the histogram for this data. Does this shape make sense? Why?

ANSWER:

The distribution is skewed to the right, implying that there are a few students (outliers) who spend a lot of money on DVDs, raising the average well above the typical or median student.

227. The following values represent the annual snowfall, measured in inches, at a local ski resort over the past 10 years: 123, 87, 143, 133, 182, 176, 96, 104, 201, and 152. What is the mean snowfall over the past 10 years? What is the standard deviation?

ANSWER:

Mean = 139.7, standard deviation = 38.46.

228. The shift supervisor at a local manufacturing plant collected data on the number of defects coming off an assembly line over the past eight weeks (40 work days). She summarized the data in the following frequency table.

# of Errors	0	1	2	3	4	5
# of Weeks	1	4	7	16	10	2

What is the weighted average number of errors per night?

ANSWER:

Weighted mean = 2.9

THE NEXT FOUR QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

An investment counselor recently reviewed the account activity of a sample of 10 of his clients and calculated the average number of stock trades per month over the past year for each client. He obtained the following data values: 10.2, 2.5, 11.4, 3.2, 1.1, 3.4, 8.4, 9.7, 11.2, and 2.4.

229. Calculate the average number of trades per month for these 10 clients.

ANSWER:

Mean = 6.35

230. Calculate the standard deviation.

ANSWER:

Standard deviation = 4.16.

Describing Data: Numerical

231. Calculate the median number of trades per month for these 10 clients.

ANSWER:

$$\text{Median} = (3.4 + 8.4) / 2 = 5.9$$

232. Describe the shape of the distribution of the number of trades per month for these 10 clients. Justify your answer.

ANSWER:

The distribution is skewed to the right (or positively skewed) since mean = 6.35 > median = 5.9.

234. The manager of the help-line at a local software manufacturing firm collected data on the number of calls the help desk received per shift for the last 100 days. The data is summarized in the table below.

# of Calls	80 but < 120	120 but < 160	160 but < 200	200 but < 240	240 but < 280	280 but < 320
# of Days	1	21	29	32	15	2

What is the estimated mean number of calls per night? What is the estimated standard deviation?

ANSWER:

Mean = 198, standard deviation = 43.06.

235. For a particular aptitude test, the mean score was 83.2. Suppose you were told that your score of 87 placed you in the 85th percentile. Provide an estimate of the standard deviation of the test.

ANSWER:

At the 85th percentile, you outperformed more than 68% of the class, so we know that you scored more than 1 standard deviation above the mean. Therefore the standard deviation must be less than 87 - 83.2 or 3.8 points.

236. The supervisor of a tourist information desk at a local airport is interested in how long it takes an employee to serve a customer. Using a stopwatch, he measures the amount of time it takes for each of 10 customers. These times, measured in minutes, are reported as follows: 2.3, 1.5, 3.9, 0.6, 2.7, 3.1, 2.8, 0.9, 1.4, and 2.6. Calculate the standard deviation and the inter-quartile range.

ANSWER:

Standard deviation = 1.046, IQR = $Q_3 - Q_1 = 2.875 - 1.275 = 1.6$.

Chapter 2

237. For the sample data shown below, compute the five-number summary, and sketch out the Box-and-Whisker plot.

34	76	29	59	71	43	24	92	82	79
49	62	34	92	87	23	64	37	54	27

ANSWER:

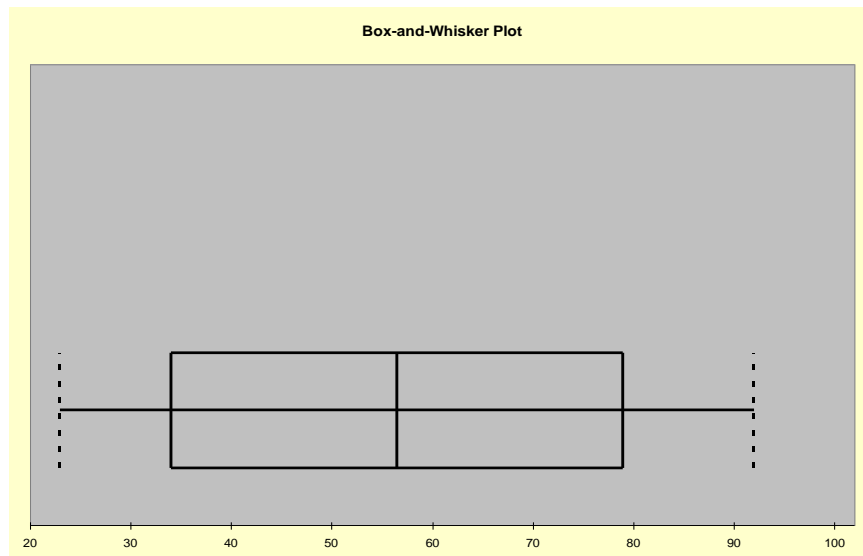
The five-number summary values are:

Minimum = 23, $Q_1 = 30.25$, $Q_2 = 56.5$, $Q_3 = 78.25$, Maximum = 92

NOTE: You obtain different results for some or all values of the quartiles if you use Excel and the syntax method or Minitab.

THE NEXT TWO QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

Consider the following (X, Y) sample data: (40, 13), (15, 33), (16, 42), (39, 9), (20, 38), (35, 15), (25, 21), (34, 17), (20, 24), and (34, 21).



238. Calculate the variances s_x^2 and s_y^2 , and the covariance s_{xy} of the sample data.

ANSWER:

$$s_x^2 = 92.844, s_y^2 = 121.122, \text{ and } s_{xy} = -95.267.$$

239. Calculate the correlation coefficient sample data.

ANSWER:

$$r = \text{Cov}(x, y) / (s_x \cdot s_y) = -0.898$$

THE NEXT SIX QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

Starting with a sample of two values 70 and 100, add three data values to your sample to obtain a new sample with certain statistics.

240. What are the three data values such that the new sample has a mean of 100? Justify your answer.

ANSWER:

Many different answers are possible. The sum of the five numbers needs to be 500; therefore we need any three numbers that total 330, such as 100, 110, 120. Thus, the new sample mean $\bar{x} = 500 / 5 = 100$.

241. What are the three data values such that the new sample has a median of 70? Justify your answer.

ANSWER:

Many different answers are possible. Need two numbers smaller than 70 and one number larger than 70. For example, we may choose 50, 60, and 80. Thus the five numbers are 50, 60, 70, 80, 100, and the median is 70.

242. What are the three data values such that the new sample has a mode of 87? Justify your answer.

ANSWER:

Many different answers are possible. Need multiple 87's. For example, we may choose 87, 87, and 95. Thus, the five numbers are 70, 87, 87, 95, 100, and the mode = 87.

243. What are the three data values such that the new sample has a mean of 100 and a median of 70? Justify your answer.

ANSWER:

Many different answers are possible. Need two numbers smaller than 70 and one number larger than 70 so that their total is 330. For example, we may choose the numbers 65, 65, and 200. Thus the five numbers are 65, 65, 70, 100, and 200. Hence, $\bar{x} = 500/5 = 100$, and the median is 70.

244. What are the three data values such that the new sample has a mean of 100 and a mode of 87? Justify your answer.

ANSWER:

Many different answers are possible. Need two numbers of 87 and a number large enough so that the total of all five numbers is 500. Therefore the three numbers are 87, 87, 156. The five numbers are 70, 87, 87, 100, and 156. Thus the mode = 87, and $\bar{x} = 500 / 5 = 100$.

Chapter 2

245. What are the three data values such that the new sample has a mean of 100, a median of 70, and a mode of 87? Justify your answer.

ANSWER:

Many different answers are possible. There must be two 87's in order to have a mode of 87, and there can only be two data values larger than 70 in order for 70 to be the median, which is impossible since 100 is one of the numbers, and that makes three of the five numbers larger than 70.

THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

Consider the following (x, y) sample data: (53, 37), (34, 26), (10, 29), (63, 55), (28, 36), (58, 48), (28, 41), (50, 42), (39, 21), and (35, 46).

246. Calculate the variances s_x^2 and s_y^2 , and the covariance s_{xy} of the sample data.

ANSWER:

$$s_x^2 = 63.511, s_y^2 = 110.767, \text{ and } s_{xy} = 100.578.$$

247. Calculate the correlation coefficient sample data.

ANSWER:

$$r = \text{Cov}(x, y) / (s_x \cdot s_y) = 0.589$$

248. In general, which of the covariance and the sample correlation coefficient is a more useful measure of the relationship between the two variables?

ANSWER:

The sample correlation coefficient is generally a more useful measure, as it provides both the direction and the strength of the relationship.

249. Consider the following sample data: 153, 178, 203, 410, 310, 231, 190, and 225. Compute the mean and median. Is the distribution of these numbers skewed to the right, skewed to the left or symmetric? Why?

ANSWER:

Mean = 237.5, median = $(203+225)/2 = 214$. Since mean > median, the distribution of these numbers is skewed to the right.

250. Since 1985, the daily temperature in Daytona Beach, Florida during spring break week has a mean of 74.6 degrees and a standard deviation of 2.2 degrees. Assuming that the distribution of temperatures is mound-shaped, how likely is it that the average daily temperature this year will be above 71 degrees? Explain.

ANSWER:

The probability is higher than 95%. This means that 95% of the time, the temperature should be between 70.2 and 79.0.

Describing Data: Numerical

THE NEXT SEVEN QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

Production records for an automobile manufacturer show the following figures for production per shift (maximum production is 720 cars per shift):

693 716 630 706 693 672 699 635 552 708 693 702 708
661 682 705 707 693 696 669 693 684 713 704 672 708

251. Would the mode be a useful summary statistic for these data? Why?

ANSWER:

The mode is defined to be the most common value, and is most often used to describe qualitative data. Here, the data are quantitative. The mode is not very useful for such data.

252. Find the median.

ANSWER:

There are $n = 26$ data values; the median is the $(26 + 1) / 2 = 27 / 2 = 13.5^{\text{th}}$ value, the average of the 13^{th} and 14^{th} ordered values. When we arrange the data in order from low to high, the 13^{th} and 14^{th} values are both 693, so the median is 693.

253. Find the mean.

ANSWER:

$$\bar{x} = \sum x_i / n = 17794 / 26 = 684.38$$

254. What does the relation between the mean and median indicate about the shape of the data?

ANSWER:

Since the mean is less than the median, this indicates that the distribution of the data is skewed to the left (negatively skewed).

255. Find the standard deviation.

ANSWER:

The deviation is $s = 34.83$.

Chapter 2

256. How well does the Empirical Rule work for the fraction of data falling within one standard deviation of the mean?

ANSWER:

Recall that the Empirical Rule says that about 68% of the data should fall within one standard deviation of the mean. This rule works best for well – behaved data, without skewness or outliers. $\bar{x} \pm 1s = 384.38 \pm 34.83 = 649.55$ to 719.21

This range includes all but three of the 26 measurements, missing only 630, 635, and 552 values. So the interval includes $23 / 26 = .885$ or 88.5% of the data, much more than the 68% indicated by the Empirical Rule. (Presumably, the reason why the rule works badly here is that the 552 outlier makes the standard deviation very big).

257. Find the IQR.

ANSWER:

To find the IQR, we must obtain the 1st and 3rd quartiles. The 1st quartile Q_1 is located in the $0.25(25+1)$ th ordered position = 6.5 th ordered position. Thus, $Q_1 = (667 + 667) / 2 = 667$. The 3rd quartile is located in the $0.75(25+1)$ th ordered position = 19.5 th ordered position. Thus, $Q_3 = (700 + 701) / 2 = 700.5$. Therefore $IQR = 700.5 - 667 = 33.5$.

258. Suppose that you are working with a data set and want to check for any outliers. What should you do? Suppose you detect an outlier. What are some of your options, and how would you make your decision?

ANSWER:

Inspect the data using either graphical tools or descriptive statistics. If the mean is quite a bit different from the median, there may be an outlier or outliers. We would want to examine the outlier to make sure that it was a legitimate value. If so, we should keep it in the data set. Otherwise we may want to remove it from the data set.

259. For data that has a mound-shaped distribution, will the inter-quartile range span a larger set of values than the range from one standard deviation below the mean to one standard deviation above the mean? Explain why or why not.

ANSWER:

Inter-quartile range contains the middle 50%. The mean ± 1 standard deviation captures the middle 68% of the observations. Therefore the inter-quartile range has to be smaller.

260. Summarize the Empirical Rule.

ANSWER:

For distributions that are bell-shaped and symmetrical:

- a) Approximately 68% of observations will fall within 1 standard deviation of the mean.
 - b) Approximately 95% of observations will fall within 2 standard deviations of the mean.
 - c) Practically all observations will fall within 3 standard deviations of the mean.
261. For a particular sample of 50 scores on a statistics exam, the following results were obtained:

Mean = 78	Median = 80	Mode = 84	Range = 52
First quartile = 68	Third quartile = 94	Standard deviation = 11	

What score was earned by more students than any other score? Why?

ANSWER:

84; since it is the mode

262. For a particular sample, the mean is 3.7 and the standard deviation is 1.2. A new sample is formed by adding 6.3 to every item of data in the original sample. Find the mean and standard deviation of the new sample.

ANSWER:

$\bar{x}_{new} = 10.0$ and $s_{new} = 1.2$

263. For the following three samples, for which sample is the data most closely grouped about the sample mean? Give a written explanation that supports your conclusion.

Sample 1: 15, 16, 19, 21, 28;
 Sample 2: 44, 49, 50, 51, 57; and
 Sample 3: 122.8, 123.7, 124.6, 130.5, 135.8.

ANSWER:

Since the sample standard deviation s measures dispersion about the mean, we compute s of each sample. Since $s_1 = 5.17$, $s_2 = 4.66$, and $s_3 = 5.54$, sample 2 has the smallest standard deviation, and the data most closely grouped about its mean.

Chapter 2

264. Consider the following two sets of data:

Set 1: 45 55 50 48 52
 Set 2: 35 50 65 47 53

Compare the following measures for both sets: $\sum x$, $\sum (x - \bar{x})$, $\sum (x - \bar{x})^2$, and the range. Comment on the meaning of these comparisons.

ANSWER:

Set 1:

x	$x - \bar{x}$	$(x - \bar{x})^2$
45	-5	25
5	+5	25
50	0	0
48	-2	4
52	+2	4
250	0	58

Set 2:

x	$x - \bar{x}$	$(x - \bar{x})^2$
35	-15	225
50	0	0
65	+15	225
47	-3	9
53	+3	9
250	0	468

Comparisons:

	$\sum x$	$\sum (x - \bar{x})$	$\sum (x - \bar{x})^2$	Range
Set1	250	0	58	10
Set2	250	0	468	30

The values of $\sum (x - \bar{x})^2$ and the range reflect the fact that there is more variability in data set 2 than in data set 1.

$\sum x$ is the same for both sets reflect the fact that both sets have the same mean $\bar{x} = 50$

$\sum (x - \bar{x}) = 0$ for both sets of data (in fact this is always true for any data).

THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

Consider the following (x, y) sample data: (24, 24), (19, 33), (21, 31), (10, 36), (22, 30), (13, 36), (21, 32), (23, 26), (20, 26), and (21, 31).

265. Calculate the variances s_x^2 and s_y^2 and the covariance s_{xy} .

ANSWER:

$s_x^2 = 19.822$, $s_y^2 = 16.944$, and $s_{xy} = -14.889$.

Describing Data: Numerical

266. Compute and interpret the sample correlation coefficient.

ANSWER:

The sample correlation coefficient $r = Cov(x, y)/(s_x \cdot s_y) = -0.812$. This indicates that there is a strong negative linear relationship between the two variables.

267. Compute and interpret b_1 ; the slope of the least squares regression line.

ANSWER:

$$b_1 = Cov(x, y)/s_x^2 = -14.889 / 19.822 = -0.7511$$

This means that for every unit increase in x , y is expected to decrease on average by about 0.75 units.

268. The following subscripted x 's represent a sample of size $n = 67$ which has been ranked from smallest (x_1) to largest (x_{67}): $x_1, x_2, x_3, \dots, x_{65}, x_{66}, x_{67}$. Prepare a 5-number summary for this sample in terms of the subscripted x 's.

ANSWER:

$$\text{Minimum} = x_1, \quad Q_1 = x_{17}, \quad \text{Median} = x_{34}, \quad Q_3 = x_{51}, \quad \text{Maximum} = x_{67}$$

269. A sample has a mean of 100.0 and a standard deviation of 15.0. According to Chebychev's Theorem, at least 8/9 of all of the data will lie between what two values?

ANSWER:

55.0 and 145.0

270. A sample of size 50 has a mean of 60.0 and a standard deviation of 10.0. According to Chebychev's Theorem, at least what percent of the data is between 10 and 110?

ANSWER:

96%

271. A sample of size 100 from a normal population has a mean of 110 and a standard deviation of 10.0. Using the Empirical Rule, about how many items of the sample will be above 130?

ANSWER:

Approximately 2 to 3 items

Chapter 2

THE NEXT FOUR QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

A sample of 26 offshore oil workers took part in a simulated escape exercise, resulting in the accompanying data on time (sec) to complete the escape:

373	370	364	366	364	325	339	393	356	359	363	375
424	325	394	402	392	369	374	359	356	403	334	397

272. Construct a stem-and-leaf display of the data. How does it suggest that the sample mean and median will compare?

ANSWER:

Stem: ones

Leaf: Tenths

32	55
33	49
34	
35	6699
36	34469
37	03345
38	9
39	2347
40	23
41	
42	4

The stem-and-leaf display is reasonably symmetric, so the mean and median will be close.

273. Calculate the values of the sample mean and median.

ANSWER:

The sample mean $\bar{x} = 9638/26 = 370.7$, and sample median = $(369+370)/2=369.50$.

274. By how much could the largest time, currently 424, be increased without affecting the value of the sample median? By how much could this value be decreased without affecting the value of the sample mean?

ANSWER:

The largest value (currently 424) could be increased by any amount. Doing so will not change the fact that the middle two observations are 369 and 170, and hence, the median will not change. However, the value $x = 424$ can not be changed to a number less than 370 (a change of $424-370 = 54$) since that *will* lower the values(s) of the two middle observations.

Describing Data: Numerical

275. What are the values of \bar{x} and \tilde{x} when the observations are re-expressed in minutes?

ANSWER:

Expressed in minutes, the mean is $(370.7 \text{ sec})/(60 \text{ sec}) = 6.18 \text{ min}$; and the median is $(369.50 \text{ sec}) / (60 \text{ sec}) = 6.16 \text{ min}$.

THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

Consider the following observations on shear strength of a joint bonded in a particular manner:

30.0 4.4 33.1 66.7 81.5 22.2 40.4 16.4 73.7 36.6 109.9

276. Determine the value of the sample mean.

ANSWER:

The sum of the $n = 11$ data points is 514.90, so $\bar{x} = 514.90/11 = 46.81$.

277. Determine the value of the sample median.

ANSWER:

The sample size $n = 11$ is odd, so there will be a middle value. Sorting the data values from smallest to largest produce the following; 4.4, 16.4, 22.2, 30.0, 33.1, 36.6, 40.4, 66.7, 73.7, 81.5, and 109.9. The sixth value, 36.6 is the middle, or median, value.

278. Why is the median so different from the mean?

ANSWER:

The mean differs from the median because the largest sample observations are much further from the median than are the smallest values.

279. The first four deviations from the mean in a sample of $n = 5$ reaction times were .6, .9, 1.0, and 1.5. What is the fifth deviation from the mean? Give a sample for which these are the five deviations from the mean.

ANSWER:

Let d denote the fifth deviation. Then $.6+.9+1.0+1.5+d = 0$ or $4.0 + d = 0$, so $d = -4.0$. One sample for which these are the deviations is $x_1 = 4.6$, $x_2 = 4.9$, $x_3 = 5.0$, $x_4 = 5.5$, $x_5 = 0$. (Obtained by adding 4.0 to each deviation; adding any other number will produce a different sample with the desired property).

Chapter 2

THE NEXT TWO QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

Calculate the following sample observations on fracture strength: 128, 131, 142, 168, 87, 93, 105, 114, 96, and 98.

280. Calculate and interpret the value of the sample mean

ANSWER:

The sample mean, $\bar{x} = \sum x_i / n = 1,162 / 10 = 116.2$.

On average, we would expect fracture strength of 116.2.

281. Calculate and interpret the value of the sample standard deviation,

ANSWER:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{5967.6}{9}} = 25.75$$

In general, the size of a typical deviation from the sample mean (116.2) is about 25.75. Some observations may deviate from 116.2 by more than this and some by less.

THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

A sample of eight doctors was asked how many flu shots they had given to patients this fall. The numbers of flu shots were 6, 3, 5, 24, 2, 6, 0, and 8.

282. Find the sample mean.

ANSWER:

$$\bar{x} = 6.75$$

283. Find the median time to learn this task.

ANSWER:

Median = 5.5 flu shots

284. Based on the values of the mean and median in the previous two questions, are the measurements symmetric or skewed? Why?

ANSWER:

Since the mean is larger than the median, we conclude that the measurements are positively skewed (skewed to the right.)

Describing Data: Numerical

THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

The following data represent scores on a 15 point aptitude test: 8, 10, 15, 12, 14, and 13.

285. Subtract 5 from every observation and compute the sample mean for the original data and the new data.

ANSWER:

$$\bar{x}_{org.} = 12, \text{ and } \bar{x}_{new} = 7.$$

286. Subtract 5 from every observation and complete the sample variance for the original data and the new data.

ANSWER:

$$s_{org.}^2 = 6.80, \text{ and } s_{new}^2 = 6.80$$

287. What effect, if any, does subtracting 5 from every observation have on the sample mean and sample variance?

ANSWER:

The sample mean \bar{x} is shifted to the left (decreased) by 5, but the sample variance remains unchanged.

288. A large sample is selected from a normal distribution. The middle 99.7% of the sample data falls between 24.2 and 69.2. Estimate the sample mean and the sample standard deviation.

ANSWER:

$$\bar{x} - 3s = 24.2, \text{ and } \bar{x} + 3s = 69.2 \Rightarrow \bar{x} = 46.7, \text{ and } s = 7.5$$

THE NEXT FOUR QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

A sample of 33 students were asked to rate themselves on whether they were outgoing or not using this five point scale: 1 = extremely extroverted, 2 = extroverted, 3 = neither extroverted nor introverted, 4 = introverted, or 5 = extremely introverted. The results are shown in the table below:

Rating x_i	1	2	3	4	5
Frequency f_i	1	7	20	5	0

289. Calculate the sample mean.

ANSWER:

$$\bar{x} = 2.88$$

Chapter 2

290. Calculate the median.

ANSWER:

Median = 3

291. Calculate the sample standard deviation.

ANSWER:

$s = 0.69$

292. Find the percentage of measurements in the intervals $\bar{x} \pm s$ and $\bar{x} \pm 2s$. Compare these results with the Empirical Rule percentages, and comment on the shape of the distribution.

ANSWER:

Sixty-one percent of the observations are in the interval $\bar{x} \pm s = (2.19, 3.57)$; the Empirical Rule says if the data set is mound-shaped, we should expect to see approximately 68% of the data within one standard deviation of the mean.

Ninety-seven percent of the observations are in the interval $\bar{x} \pm 2s = (1.50, 4.26)$; the Empirical Rule says that if the data set is mound-shaped, we should expect to see approximately 95% of the observations within two standard deviations of the mean. Since the percentages of measurements in the intervals $\bar{x} \pm s$ and $\bar{x} \pm 2s$ are close to those predicted by the Empirical Rule, the data must be approximately mound-shaped.

THE NEXT FOUR QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

Consider the following scores on a 20 point aptitude test for two samples of eight students each: Sample 1: 18, 19, 17, 15, 14, 20, 14, and 16.

Sample 2: 14, 15, 13, 11, 10, 16, 10, and 12.

293. Calculate the mean score in sample 1

ANSWER:

$\bar{x}_1 = 16.625$

294. Calculate the mean score in sample 2

ANSWER:

$\bar{x}_2 = 12.625$

295. Calculate the variance for the scores in sample 1

ANSWER:

$s_1^2 = 5.1248$

296. Calculate the variance for the scores in sample 2

ANSWER:

$$s_2^2 = 5.1248$$

297. You may have noticed that each score in sample 2 is obtained by subtracting 4 from the corresponding score in sample 1. Review your answers to the previous four questions and write your conclusion.

ANSWER:

The mean in the second sample is shifted to the left (decreased) by 4 from the mean in the first sample, but the variance remained unchanged.

THE NEXT TWO QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

In a time study, conducted at a manufacturing plant, the length of time to complete a specified operation is measured for each on $n = 40$ workers. The mean and standard deviation are found to be 15.2 and 1.40, respectively.

298. Describe the sample data using the Empirical Rule.

ANSWER:

To describe the data, calculate these intervals:

$$(\bar{x} \pm s) = 15.2 \pm 1.40, \text{ or } 13.8 \text{ to } 16.6$$

$$(\bar{x} \pm 2s) = 15.2 \pm 2.80, \text{ or } 12.4 \text{ to } 18.0$$

$$(\bar{x} \pm 3s) = 15.2 \pm 4.20, \text{ or } 11.0 \text{ to } 19.4$$

If the distribution of measurements is mound-shaped, you can apply the Empirical Rule and expect approximately 68% of the measurements to fall into the interval from 13.8 to 16.6, approximately 95% to fall into the interval from 12.4 to 18.0, and all or almost all to fall into the interval from 11.0 to 19.4.

299. Describe the sample data using Chebychev's Theorem.

ANSWER:

If you doubt that the distribution of measurements is mound-shaped, or if you wish for some other reason to be conservative, you can apply Chebychev's Theorem and be absolutely certain of your statements. Chebychev's Theorem tells you that at least $3/4$ of the measurements fall into the interval from 12.4 to 18.0 and at least $8/9$ into the interval from 11.0 to 19.4

Chapter 2

THE NEXT FOUR QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

The following data represents the number of minutes an athlete spends training per day.

73	74	76	77	79	79	83	84	88
84	84	85	86	86	87	87	88	91
92	92	93	97	98	98	81	82	

The mean and standard deviation were computed to be 85.54 and 6.97, respectively.

300. Create a stem and leaf plot for the distribution of training times. Is the distribution relatively mound-shaped?

ANSWER:

<u>Stems</u>	<u>Leaves</u>
7	34
7	6799
8	123444
8	5667788
9	1223
9	788

The distribution of training times appears to be relatively mound-shaped.

301. What percentage of measurements would you expect to be between 71.60 and 99.48?

ANSWER:

Since the distribution appears to be relatively mound-shaped, the Empirical Rule applies. The interval (71.60, 99.48) represents 2 standard deviations from the mean, so one would expect approximately 95% of the measurements to lie in this interval.

302. What percentage of the measurements lies in the interval (71.60, 99.48)?

ANSWER:

26 of the 26 measurements or 100% of the measurements lie in the given interval.

303. Do your answers to the previous two questions agree? If not, what conclusion can be drawn?

ANSWER:

The two percentages do not agree exactly indicating that the distribution of training times is not perfectly mound-shaped. However, it is very close!