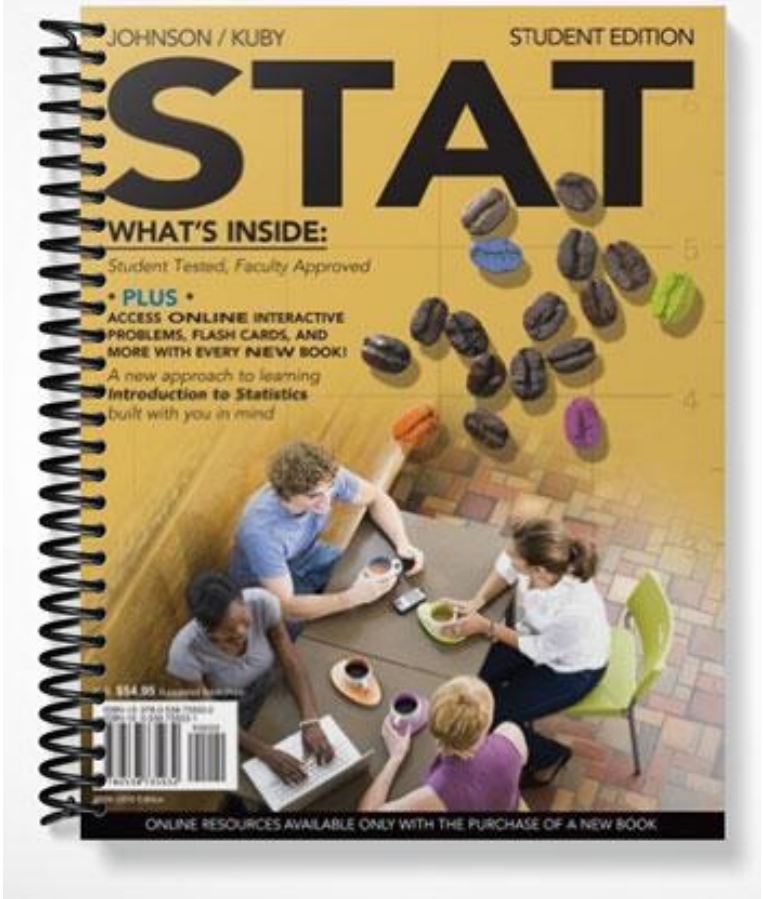
TEST BANK


JOHNSON / KUBY
STUDENT EDITION
STAT
WHAT'S INSIDE:
Student Tested, Faculty Approved
• PLUS •
ACCESS ONLINE INTERACTIVE PROBLEMS, FLASH CARDS, AND MORE WITH EVERY NEW BOOK!
A new approach to learning
Introduction to Statistics
built with you in mind
ONLINE RESOURCES AVAILABLE ONLY WITH THE PURCHASE OF A NEW BOOK

Chapter 2

Descriptive Analysis and Presentation of Single-Variable Data

Section 2.1

True-False Questions

1. Circle graphs and bar graphs are graphs that are used to summarize qualitative, or attribute, or categorical data.
   ANSWER: T

2. All graphic representations of sets of data need to be completely self-explanatory. That includes a descriptive meaningful title, and identification of the vertical and horizontal scales.
   ANSWER: T

3. The *stem-and-leaf display* for summarizing numerical data is a combination of a graphic technique and a sorting technique.
   ANSWER: T

4. There is no single correct answer when constructing a graphic display. The analyst's judgment and the circumstances surrounding the problem play a major role in the development of the graphic.
   ANSWER: T

5. Circle graphs and bar graphs are graphs that are used to summarize quantitative data.
   ANSWER: F

6. Circle graphs (pie diagrams) show the amount of data that belong to each category as a proportional part of a circle.
   ANSWER: T

7. Circle graphs show the amount of data that belong to each category as a frequency.
   ANSWER: F

8. Bar graphs show the amount of data that belong to each category as a proportionally sized rectangular area.
   ANSWER: T

9. Bar graphs of attribute data should be drawn with connected bars of equal width.
ANSWER: F

10. One major reason for constructing a graph of quantitative data is to display its distribution.
ANSWER: T

Multiple-Choice Questions

11. Which of the following statements is false?

A) Pareto diagram is a bar graph with the bars arranged from the most numerous categories to the least numerous categories.
B) Pareto diagram includes a line graph displaying the cumulative percentages and counts for the bars.
C) A Pareto diagram of types of defects will show the ones that have the greatest effect on the defective rate in order of effect. It is then easy to see which defects should be targeted in order to most effectively lower the defective rate.
D) None of the above.
ANSWER: D

12. Which of the following statements is false?

A) Dotplot displays the data of a sample by representing each data with a dot positioned along a scale. This scale can be either horizontal or vertical. The frequency of the values is represented along the other scale.
B) Pareto diagram includes a line graph displaying the frequency (counts) for the bars.
C) Dotplot display is a convenient technique to use as you first begin to analyze the data. It results in a picture of the data as well as sorts the data into numerical order.
D) The stem-and-leaf display is a combination of a graphic technique and a sorting technique. This display is simple to create and use, and it is well suited to computer applications.
ANSWER: B

Short-Answer Questions

13. Complete the following statement: A stem-and-leaf display is a combination of a sorting technique and a _____ technique.

ANSWER:
graphing

14. Complete the following statement: Circle graphs and bar graphs are often used to summarize _____ data.

ANSWER:
attribute

15. Data for the distribution of land in a particular county is given in percentages. Name two types of graphs that would be most appropriate to display these results.

ANSWER:
Bar graph or circle graph

Applied and Computational Questions

16. Construct a stem-and-leaf display for the data below.

219	225	222	243	234	241	231	235	234
231	240	231	246	232	229	233	233	226
225	227	230	229	227	218	216	234	240

ANSWER:

21	**6 8 9**	
22	**2 5 5 6 7 7 9 9**	
23	**0 1 1 1 2 3 3 4 4 4 5**	
24	**0 0 1 3 6**	

17. The number of vehicles passing a tollgate between 7 a.m. and 8 a.m. were recorded for twenty different days. Construct a stem-and-leaf display for these data.

10 26 32 15 16 22 31 46 27 33 27 15 16 19 20 16 12 22 30 41

ANSWER:

1	**0 2 5 5 6 6 6 9**
2	**0 2 2 6 7 7**
3	**0 1 2 3**
4	**1 6**

18. A group of hypertensive patients (with diastolic blood pressure between 110 and 130) were given a medication for reducing elevated blood pressure. The decreases in blood pressure produced by the medication were categorized into four categories as follows:

Category	*Decrease in Pressure*
A--Marked decrease in blood pressure	15 or more units
B--Moderate decrease in blood pressure	10 to less than 15 units
C--Slight decrease in blood pressure	5 to less than 10 units
D--Stationary blood pressure	0 to less than 5 units

Thirty patients who used the medication experienced the following blood pressure reductions. Give the height of each at the four bars of a bar graph for these results.

12	15	6	4	20	17	25	4	5	18
10	12	18	13	14	20	30	12	14	17
30	18	10	8	16	32	27	13	8	4

ANSWER:

Category	*Height of bar*
A	14
B	9
C	4
D	3

19. A random sample of test scores was taken from two different sections of an introductory statistics course. Construct a back-to-back stem-and-leaf display for this set of data.

Section A:	46	97	99	64	78	76	45	73	81	51	68	81	81	79	100
Section B:	80	69	92	75	88	47	98	92	90	81	42	50	59	66	67	66

ANSWER:

Sec. A		***Sec. B***
56	4	27
1	5	09
48	6	6679
3689	7	5
111	8	018
79	9	0228
0	10	

20. The total amount spent for textbooks (to the nearest dollar) was recorded for several students. Some of the information was collected for the summer session (denoted by S), and some was collected for the fall semester (denoted by F). Construct a back-to-back stem-and-leaf display for this set of data.

Semester:	S	F	F	S	F	F	F	F	S	F	S
Amount:	25	90	115	40	80	75	95	60	29	120	46
Semester:	S	F	F	S	F	F	F	S	F	F	S
Amount:	35	75	80	50	122	95	79	20	95	65	42
Semester:	F	F	F	F	F	S	F	F			
Amount:	80	69	112	105	108	37	98	92			

ANSWER:

Summer		*Fall*
059	02	
57	03	
026	04	
0	05	
	06	059
	07	559
	08	000
	09	025558
	10	58
	11	25
	12	02

21. A department of mathematical sciences has majors in four areas.

Major	*Number of Majors*
Mathematics	50
Computer Science	22
Actuarial Science	15
Statistics	10

If a circle graph is constructed for these data, what would be the percentage of the graph for each major?

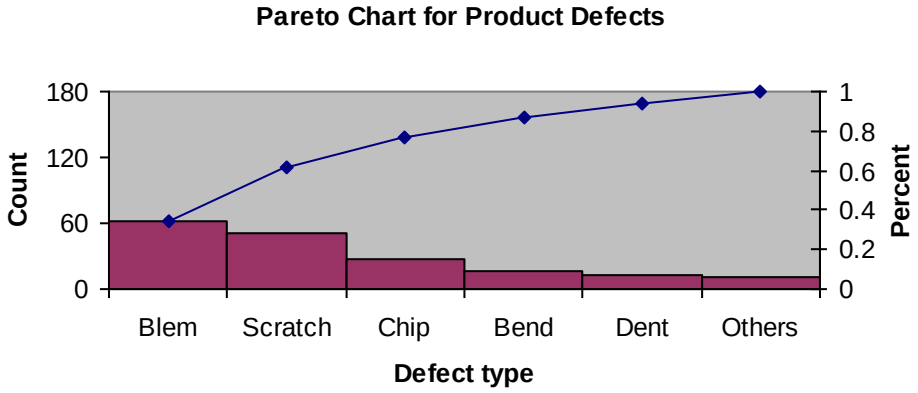
ANSWER:

Major	*% of Majors*
Mathematics	51.5
Computer Science	22.7
Actuarial Science	15.5
Statistics	10.3

QUESTIONS 22 THROUGH 25 ARE BASED ON THE FOLLOWING INFORMATION:

The final-inspection defect report for an assembly line is reported on the table and Pareto diagram as shown below:

Defect	Blemish	Scratch	Chip	Bend	Dent	Others
Count	61	50	28	17	13	11



22. What is the total defect count in the report?

ANSWER:
180 defects

23. Find the percentage for "chip" defect items.

ANSWER:
Percent of chip = $(50/180) \cdot 100\% = 15.56\%$

24. Find the "cum % for bend", and explain what that value means.

ANSWER:
$[(61+50+28+17)/180] \cdot 100\% = (156/180) \cdot 100\% = 86.67\%$. The value 86.67% is the sum of the percentages for all defects that occurred more often than Bend, including Bend.

25. Management has given the production line the goal of reducing their defects by 50%. What two defects would you suggest they give special attention to in working toward this goal? Explain.

ANSWER:
The two defects, Blemish and Scratch, total 61.67%. If they can control these two defects, the goal should be within reach.

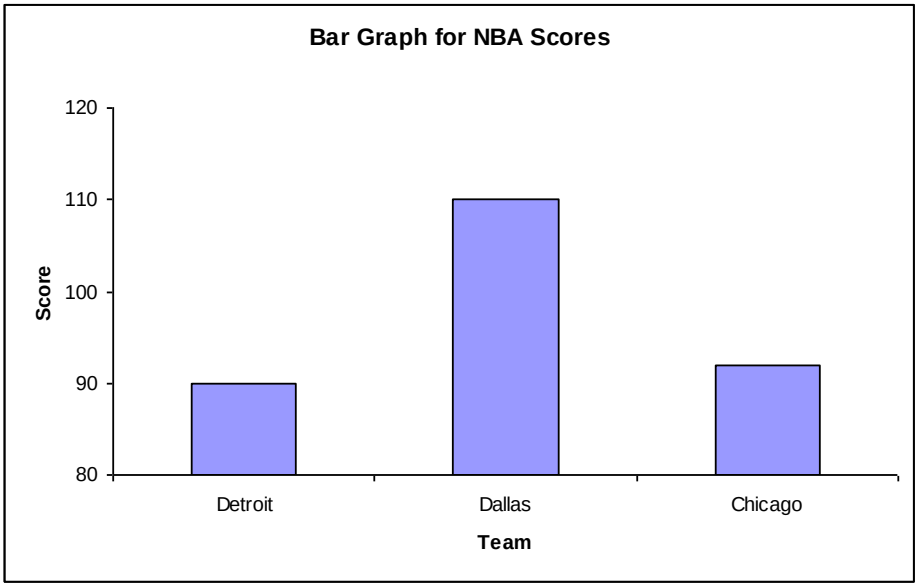
QUESTIONS 26 THROUGH 29 ARE BASED ON THE FOLLOWING INFORMATION:

The points scored by the winning teams on opening night of a recent NBA season are shown in the table below:

Team	Detroit	Dallas	Chicago
Score	90	110	92

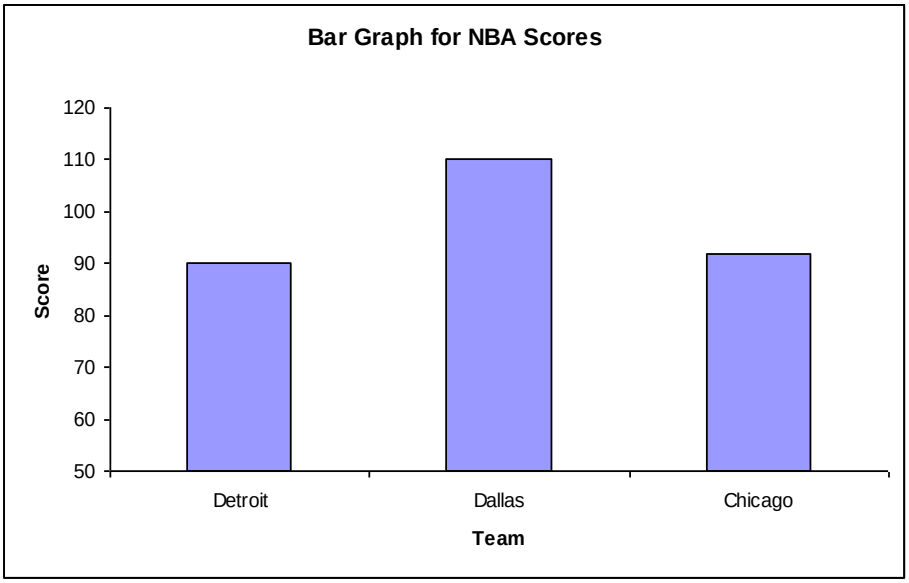
26. Draw a bar graph of these scores using a vertical scale ranging from 80 to 120.

ANSWER:



27. Draw a bar graph of the scores using a vertical scale ranging from 50 to 120.

ANSWER:

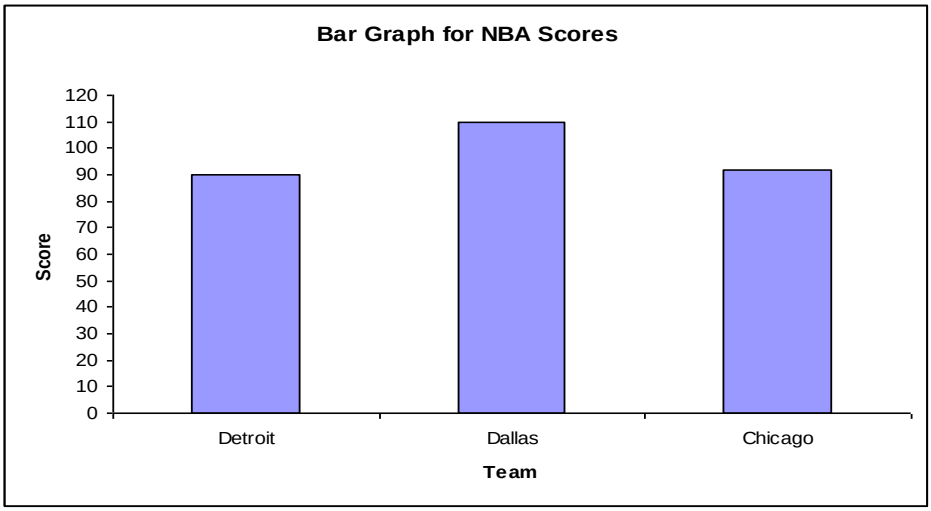


28. In which bar graph does it appear that the NBA scores vary more? Why?

ANSWER:
Bar graph in question 27 emphasizes the variation in the scores as it focuses only on the variation and not the relative size of the scores.

29. How could you create an accurate representation of the relative size and variation between these scores? Draw this new bar graph.

ANSWER:
An accurate representation of both the size and variation of the values would be best served by starting the vertical scale at zero.

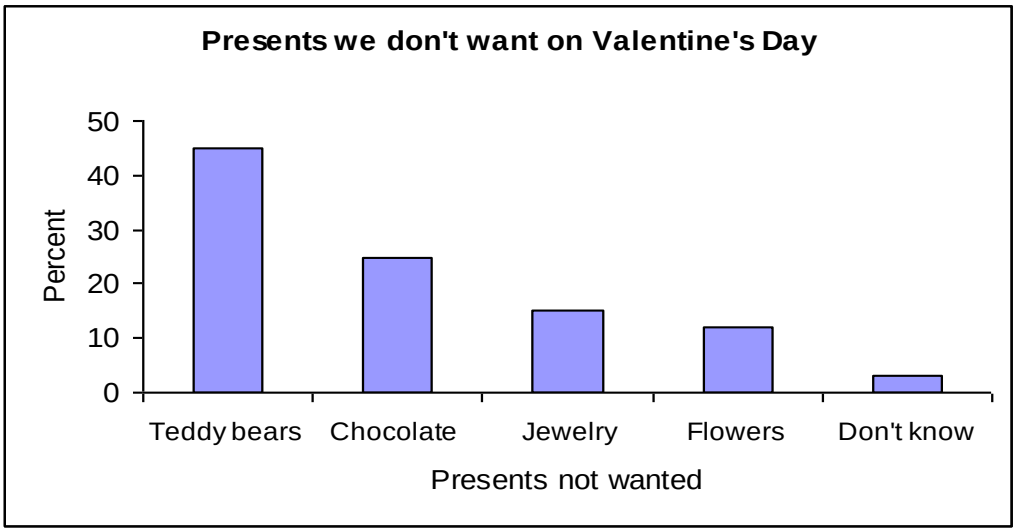


QUESTIONS 30 THROUGH 33 ARE BASED ON THE FOLLOWING INFORMATION:
What not to get them on Valentines Day! A recent study among adults in USA shows that adults prefer not to receive certain items as gifts on Valentine's Day as shown below:

Teddy bears: 45%; Chocolate: 25%; Jewelry: 15%; Flowers: 12%; Don't Know: 3%.

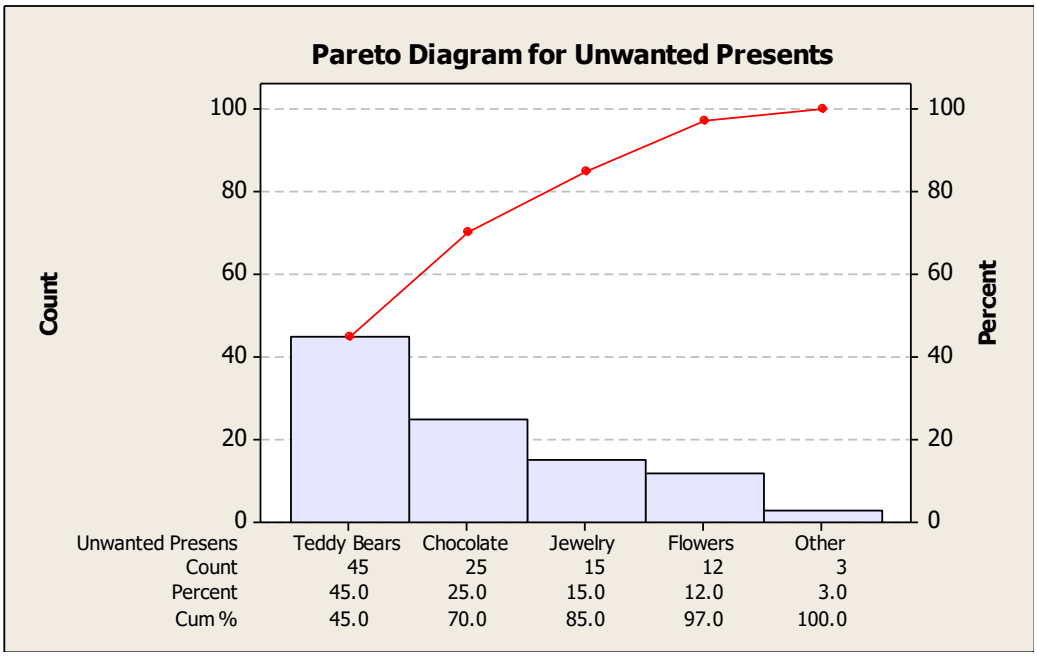
30. Draw a bar graph picturing the percentages of "Presents not wanted".

ANSWER:



31. Draw a Pareto diagram picturing the “Presents not wanted”.

ANSWER:



32. If you want to be 80% sure you did not get your valentine something unwanted, what should you avoid buying? How does the Pareto diagram show this?

ANSWER:
Teddy bears, chocolates, jewelry; these are listed first in the Pareto diagram.

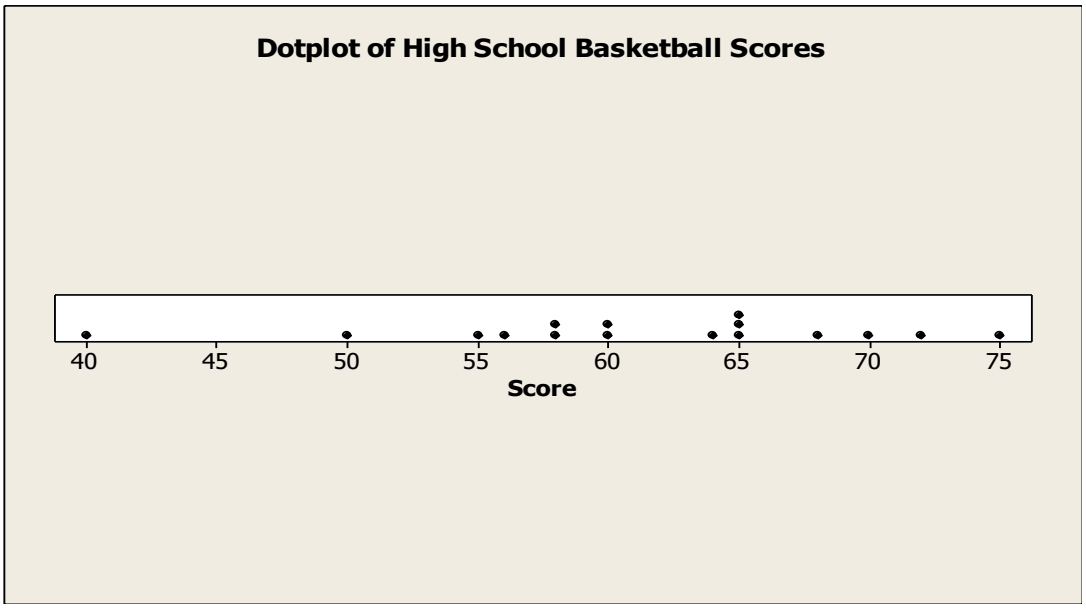
33. 400 adults are to be surveyed, what frequencies would you expect to occur for each unwanted item listed on the snapshot?

ANSWER:
The frequencies are 180, 100, 60, 48, and 12 for teddy bears, chocolates, jewelry, flowers, and don’t know, respectively.

QUESTIONS 34 THROUGH 36 ARE BASED ON THE FOLLOWING INFORMATION:
The points scored during each game by the Big Rapids High School basketball team last season were: 60, 58, 65, 75, 50, 65, 60, 72, 64, 70, 58, 65, 56, 40, 68, and 55.

34. Construct a dotplot of these data.

ANSWER:



35. Use the dotplot in question 34 to uncover the lowest and highest scores.

ANSWER:
The lowest score was 40 and the highest was 75.

36. Use the dotplot in question 34 to determine the most common score? How many teams share that score?

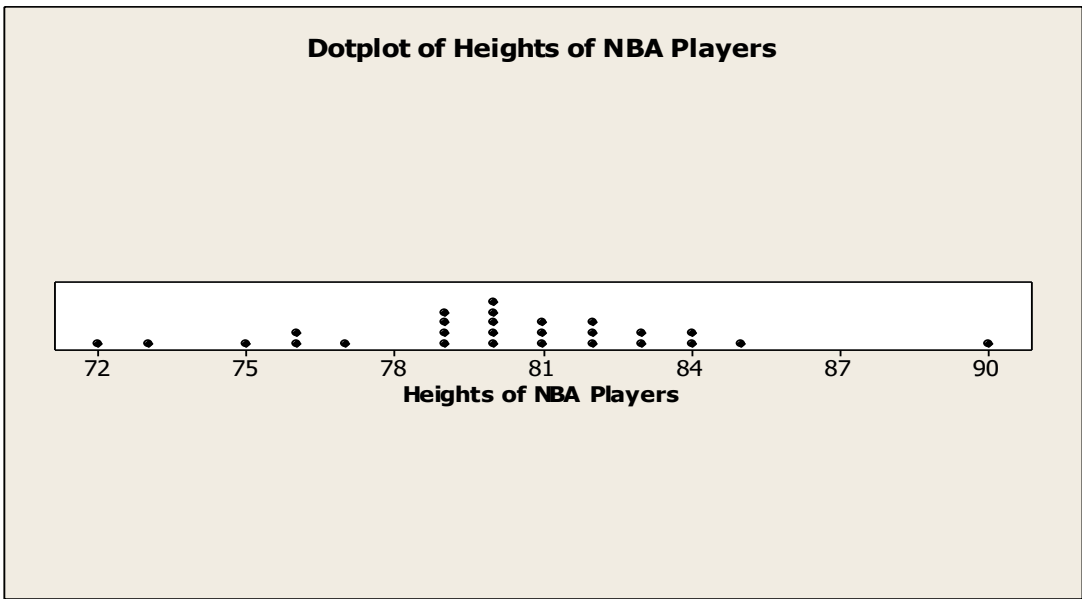
ANSWER:
65; three teams share that score

QUESTIONS 37 THROUGH 40 ARE BASED ON THE FOLLOWING INFORMATION:
The data shown below are the heights (in inches) of the basketball players who were the first round picks by the professional NBA teams in a recent year.

83	83	75	80	76	80	81	84	79	80
84	86	72	82	82	79	81	79	80	73
90	82	81	75	77	80	79	76	85	

37. Construct a dotplot of the heights of these players.

ANSWER:



38. Use the dotplot in question 37 to uncover the shortest and the tallest players.

ANSWER:
The shortest player is 72 inches and the tallest player is 90 inches.

39. Use the dotplot in question 37 to determine the most common height and how many players share that height?

ANSWER:
The most common height is 80 inches, shared by 5 players.

40. What feature of the dotplot in question 37 illustrates the most common height?

ANSWER:
The height of column of dots illustrates the most common height.

Sections 2.2 through 2.5

True-False Questions

41. A histogram is used to summarize attributive data.
ANSWER: F

42. One major reason for constructing a graph of quantitative data is to display its distribution.
ANSWER: T

43. In a *J*-shaped histogram, there is one tail on the side of the class with the highest frequency.
ANSWER: F

44. A line graph of a cumulative frequency or cumulative relative frequency distribution is referred to as an ogive.
ANSWER: T

45. The frequency of a class is the number of pieces of data whose values fall within the boundaries of that class.
ANSWER: T

46. Frequency distributions are used in statistics to present large quantities of repeating values in a concise form.
ANSWER: T

47. If grouping data are used to form a frequency distribution, the class width is the difference between the upper and lower class boundaries.
ANSWER: T

48. If grouping data are used to form a frequency distribution, the class midpoint (sometimes called the class mark) is the numerical value that is exactly in the middle of each class. It is found by adding the class boundaries and dividing by 2.
ANSWER: T

49. A histogram is a bar graph that represents a frequency distribution of categorical data.
ANSWER: F

50. A bimodal distribution has two high-frequency classes separated by classes with lower frequencies. It is not necessary for the two high frequencies to be the same.
ANSWER: T

51. Relative frequency can be expressed as a common fraction, in decimal form, but not as a percentage.
ANSWER: F

52. The histogram of a sample should have a distribution shape very similar to that of the population from which the sample was drawn.
ANSWER: T

53. An ogive is a line graph of a cumulative frequency or cumulative relative frequency distribution.
ANSWER: T

54. Every ogive starts on the left with a cumulative relative frequency of zero at the lower class boundary of the first class and ends on the right with a cumulative relative frequency of 100% at the upper class boundary of the last class.
ANSWER: T

55. Measures of central tendency measure the spread of a set of data about its center.
ANSWER: F

56. For every set of data, the value of the median will always be one of the original items of data.
ANSWER: F

57. In a sample of size *n*, the median of the sample is $(n+1)/2$.
ANSWER: F

58. The midrange for a set of data is found by subtracting the lowest valued data *L* from the highest valued data *H*.
ANSWER: F

59. The mean, median and mode are the most common measures of dispersion (spread).
ANSWER: F

60. Measures of central tendency are numerical values that locate, in some sense, the center of a set of data.
ANSWER: T

61. The mean, median and mode for the set of data {3, 5, 3, 8, 6} are all the same value.
ANSWER: F

62. The mean of a sample always divides the data into two equal halves (half larger and half smaller in value than itself).
ANSWER: F

63. A measure of central tendency is a quantitative value that describes how widely the data are dispersed about a central value.
ANSWER: F

64. For any distribution, the sum of the deviations from the mean equals zero.
ANSWER: T

65. Measures of central tendency are attribute data that locate, in some sense, the center of a set of data.
ANSWER: F

66. The term average is often associated with all measures of central tendency.
ANSWER: T

67. The population mean, μ (lowercase mu in the Greek alphabet), is the mean of all *x* values in the entire population.
ANSWER: T

68. The median is the value of the data that occupies the middle position when the data are ranked in order according to size.
ANSWER: T

69. The sample median is represented by $\bar{x}$.
ANSWER: F

70. The midrange is the number exactly midway between a lowest value data *L* and a highest value data *H*. It is found by averaging the low and high values.
ANSWER: T

71. The sample mean is represented by $\tilde{x}$ (read "x-tilde").
ANSWER: F

72. The population median is represented by *M* (the uppercase mu in the Greek alphabet).
ANSWER: T

73. When *n* is odd, the depth of the median, $d(\bar{x})$ will always be an integer.
ANSWER: T

74. When *n* is even, the depth of the median, $d(\bar{x})$ will always be an integer or a half-number.
ANSWER: F

75. According to your book, if two or more values in a sample are tied for the highest frequency (number of occurrences), we say there is no mode.
ANSWER: T

76. The midrange is the range of the middle two values.
ANSWER: F

77. There are several kinds of measures ordinarily known as averages and each gives a different picture of the figures it is called on to represent.
ANSWER: T

78. The standard deviation is the positive square root of the variance.
ANSWER: T

79. The sum of the squares of the deviations from the mean $\sum (x - \bar{x})^2$, will sometimes be negative.
ANSWER: F

80. The standard deviation for the set of values 5, 5, 5, 5, and 5 is 5.
ANSWER: F

81. The sample variance, s^2, is the mean of the squared deviations of *x* values from the sample mean $\bar{x}$, calculated using *n* – 1 as the divisor.
ANSWER: T

82. The measures of dispersion include the *range, variance*, and *standard deviation*.
ANSWER: T

83. The unit of measure for the variance is the same as the unit of measure for the data.
ANSWER: F

84. There is no limit to how widely spread out the data can be; therefore, measures of dispersion can be very large.
ANSWER: T

85. Although the mean deviation is always zero, it is a useful statistic in some occasions.
ANSWER: F

86. The range is the difference in value between the highest-valued (*H*) and the lowest-valued (*L*) data.
ANSWER: T

87. The sample variance, s^2 is the mean of the deviations of *x* values from the sample mean $\bar{x}$.
ANSWER: F

88. The standard deviation of a sample is the square of the sample variance.
ANSWER: F

89. If a rounded value of $\bar{x}$ is used, then $\sum(x-\bar{x})$ will not always be exactly zero. It will, however, be reasonably close to zero.
ANSWER: **T**

90. In a box-and-whisker display, the length of the "box" is the same as the interquartile range.
ANSWER: **T**

91. Each set of data has four quartiles; they divide the ranked data into four equal quarters.
ANSWER: **F**

92. The numerical value midway between the first quartile and the third quartile is referred to as the midquartile.
ANSWER: **T**

93. Each set of data has 100 percentiles; they divide the ranked data into 100 equal subsets.
ANSWER: **F**

94. The median, the midrange, and the midquartile are always the same value, since each is a middle value.
ANSWER: **F**

95. The interquartile range is the difference between the first and third quartiles; it is the range of the middle 50% of the data.
ANSWER: **T**

96. The standard score (or *z*-score) identifies the position a particular value of *x* has relative to the mean, measured in standard deviations; that is, $z=(x-\bar{x})/s$.
ANSWER: **T**

97. On a test Jim scored at the 50th percentile and Jean scored at the 25th percentile; therefore, Jim's test score was twice Jean's test score.
ANSWER: **F**

98. The unit of measure for the standard score is always in standard deviations.
ANSWER: **T**

99. Data must be ranked before calculating many of the measures of position.
ANSWER: **T**

100. Each set of data has four quartiles.
ANSWER: **F**

101. Measures of position are used to describe the position a specific data value possesses in relation to the mean of the data.
ANSWER: **F**

102. Measures of position are used to describe the position a specific data value possesses in relation to the rest of the data.
ANSWER: **T**

103. Quartiles and percentiles are two of the most popular measures of dispersion.
ANSWER: **F**

104. The median, the second quartile, and the 50th percentile are all the same.

ANSWER: **T**

105. 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**

106. The median, the midrange, and the midquartile are not necessarily the same value. Each is the middle value, but by different definitions of "middle."
ANSWER: **T**

107. Percentiles are values of the variable that divide a set of ranked data into 100 equal subsets.
ANSWER: **T**

108. Each set of data has 100 percentiles.
ANSWER: **F**

109. 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**

110. The first quartile and the 25th percentile are the same.
ANSWER: **T**

111. The mean, median, the second quartile, and the 50th percentile are all the same.
ANSWER: **F**

112. The midquartile is a measure of central tendency.
ANSWER: **T**

113. The 5-number summary divides a set of data into four subsets, with one-quartile of the data in each subset.
ANSWER: **T**

114. The median, the midrange, and the midquartile are the same, since each is the middle value.
ANSWER: **F**

115. The midquartile is the numerical value midway between the first quartile and the third quartile.
ANSWER: **T**

116. The interquartile range is the average of the first and third quartiles.
ANSWER: **F**

117. The interquartile range is the range of the middle 50% of the data.
ANSWER: **T**

118. 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**

119. Since the *z*-score is a measure of relative position with respect to the mean, it can be used to help us compare two raw scores that come from separate populations.
ANSWER: **T**

120. The midquartile, defined as the average of the first and third quartiles, is a measure of position, simply because quartiles are one of the most popular measures of position.
ANSWER: F

Multiple-Choice Questions

121. At a large company, the majority of the employees earn from $20,000 to $30,000 per year. Middle management employees earn between $30,000 and $50,000 per year while top management earn between $50,000 and $100,000 per year. A histogram of all salaries would have which of the following shapes?

A) Symmetrical
B) Uniform
C) Skewed to right
D) Skewed to left
ANSWER: C

122. Which of the following statements is false regarding an ogive?

A) The horizontal scale identifies the upper class boundaries.
B) The vertical scale identifies either the cumulative frequencies or the cumulative relative frequencies.
C) Every ogive starts on the left with a cumulative relative frequency of one at the upper class boundary of the first class.
D) None of the above
ANSWER: C

123. Which of the following statements is false regarding an ogive?

A) It is a line graph of a cumulative frequency or cumulative relative frequency distribution.
B) Its horizontal scale is always based on the lower class boundaries.
C) Its vertical scale identifies either the cumulative frequencies or the cumulative relative frequencies.
D) None of the above
ANSWER: B

124. Which of the following are graphic representations of sets of data?

A) A descriptive and meaningful title
B) A proper identification of the vertical scale
C) A proper identification of the horizontal scale
D) All of the above
ANSWER: D

125. Which of the following statements is false?

A) Relative frequencies are often useful in a presentation because nearly everybody understands fractional parts when expressed as percents.
B) Relative frequencies are particularly useful when comparing the frequency distributions of two different size sets of data.
C) The histogram of a sample should have a distribution shape that is bimodal.
D) A stem-and-leaf display contains all the information needed to create a histogram.
ANSWER: C

126. The following set of data represents letter grades on term papers in a rhetoric class:

A, A, A, B, B, B, B, C, C, C, C, C, C, C, C, C, C, D, D, D, F.

Select the most appropriate measure of central tendency for the data described.

A) Mean
B) Median
C) Mode
D) Midrange
ANSWER: C

127. The following set of data represents the ages of students in a small seminar: 20, 21, 22, 25, 26, 27, and 68. Select the most appropriate measure of central tendency for the data described.

A) Mean
B) Median
C) Mode
D) Midrange
ANSWER: B

128. The following set of data represents the temperature high for seven consecutive days in February in Chicago: 22, 14, 26, 27, 35, 38, and 41. Select the most appropriate measure of central tendency for the data described.

A) Mean
B) Median
C) Mode
D) Midrange
ANSWER: A

129. Which of the following is not affected by extreme values?

A) Median
B) Tenth percentile
C) Third quartile
D) All of the above
ANSWER: D

130. The measure most affected by extreme values is the:

A) mean
B) second quartile
C) first quartile
D) midquartile
ANSWER: A

131. Which of the following is *not* a measure of central tendency?

A) Mean
B) Median
C) Midrange
D) None of the above
ANSWER: D

132. Which of the following statements is not true?

A) When *n* is odd, the depth of the median, $d(\tilde{x})$, will always be an integer.
B) When *n* is even, the depth of the median, $d(\tilde{x})$, will always be a half-number.
C) The midrange is a measure of position.
D) None of the above.
ANSWER: C

133. The following data set represents shirt sizes for girls' field hockey team:

S, S, S, M, M, M, M, M, M, M, M, M, M, L, L, L, L, L, XL, XL

Select the most appropriate measure of central tendency for the data described.

A) Mean
B) Median
C) Mode
D) Midrange
ANSWER: C

134. Adding 5 to each value in a data set would *not* change which of the following measures?

A) Mode
B) Mean
C) Mid-range
D) Standard deviation
ANSWER: D

135. Which of the following is a correct statement?

A) The interquartile range is found by taking the difference between the first and third quartiles and dividing that value by 2.
B) The standard deviation is expressed in terms of the original units of measurement but the variance is not.
C) The values of the standard deviation may be either positive or negative, while the value of the variance will always be positive.
D) A large measure of dispersion is the result of an error of calculation because there is a limit to how widely spread out data can be.
ANSWER: B

136. Which of the following is a correct statement?

A) The mean is a measure of the deviation in a data set.
B) The standard deviation is a measure of dispersion.
C) The range is a measure of central tendency.
D) The median is a measure of dispersion.
ANSWER: B

137. The difference between the largest and smallest values in an ordered array is called the:

A) standard deviation
B) variance
C) interquartile range
D) range
ANSWER: D

138. Which of the following is the weakest measure of dispersion?

A) Range
B) Variance
C) Standard deviation
D) None of them
ANSWER: A

139. Which of the following statements is false?

A) The measures of dispersion include the *range, variance*, and *standard deviation*.
B) The numerical values of measures of dispersion describe the amount of spread, or variability that is found among the data values.
C) Closely grouped data have relatively small measures of dispersion values, and more widely spread-out data have larger values.
D) None of the above
ANSWER: D

140. Which of the following statements is false?

A) $\sum(x-\bar{x})$ is always zero even if a rounded value of $\bar{x}$ is used.
B) The standard deviation of a sample, *s*, is the positive square root of the variance.
C) The unit of measure for the standard deviation is the same as the unit of measure for the data.
D) The unit of measure for the variance is *units squared* of the unit of measure for the data.
ANSWER: A

141. Which of the following types of graphs would *not* be good for qualitative data?

A) Box-and-whiskers display
B) Circle graph
C) Bar graph
D) Pareto diagram
ANSWER: A

142. For a normal distribution, a value that is two standard deviations below the mean would be closer to which of the following?

A) Third percentile
B) First quartile
C) Fortieth percentile
D) Median
ANSWER: A

143. Which of the following statements is correct?

A) Measures of position are used to describe the position a specific data value possesses in relation to the rest of the data.
B) Quartiles and percentiles are two of the most popular measures of position.
C) Quartiles are values of the variable that divide the ranked data into 4 equal subsets called quarters.
D) All of the above.
ANSWER: D

144. Which is the depth of Q_1 for a ranked set of 40 exam scores?

A) 9.5
B) 10.0
C) 10.5
D) 11.0
ANSWER: C

145. Which of the following statements is false?

A) The first quartile, Q_1, is a number such that at most 25% of the data are smaller in value than Q_1 and at most 75% are larger.
B) The second quartile is the mean.
C) The third quartile, Q_3, is a number such that at most 75% of the data are smaller in value than Q_3, and at most 25% are larger.
D) None of the above
ANSWER: B

146. Which is the depth of the 65th percentile for a ranked set of 50 student ages?

A) 32.5
B) 33.0
C) 33.5
D) 34.0
ANSWER: B

147. If the 70th percentile for a set of exam scores is 82, what does this mean?

A) At most 70% of the exam scores are smaller in value than 82
B) At most 82% of the exam scores are smaller in value than 70
C) At least 70% of the exam scores are larger in value than 82
D) At least 82% of the exam scores are larger in value than 70
ANSWER: A

148. The 5-number summary divides a set of data into how many subsets?

A) 6
B) 5
C) 4
D) 3
ANSWER: C

149. Which of the following statements is false?

A) The 5-number summary is more informative when it is displayed on a diagram drawn to scale. A computer-generated graphic display that accomplishes this is known as the box-and-whiskers display.
B) The position of a specific value in a set of data can be measured in terms of the mean and variance using the standard score, commonly called the *z*-score.
C) The *z*-scores are typically range in value from approximately -3.00 to +3.00.
D) None of the above
ANSWER: B

150. Which is the depth of the 5th percentile for a ranked set of 35 student weights?

A) 1.50
B) 2.00
C) 2.50
D) 3.00
ANSWER: B

Short-Answer Questions

151. Explain the difference between a *J*-shaped histogram and a skewed histogram.

ANSWER:
J-shaped histogram has only one tail with the highest frequency as an end class. A skewed histogram has tails on both sides of the class with the highest frequency, with one tail being considerably longer.

152. If a histogram is constructed for the following frequency distribution, what shape would it have?

Class Boundaries	*Frequency*
$20 \le x < 30$	5
$30 \le x < 40$	15
$40 \le x < 50$	20
$50 \le x < 60$	18
$60 \le x < 70$	13
$70 \le x < 80$	10
$80 \le x < 90$	5
$90 \le x \le 100$	1

ANSWER:
Skewed to right or positively skewed

153. What is the largest possible value needed on the vertical axis of a relative frequency histogram?

ANSWER:
One

154. A relative frequency distribution was constructed for a sample of size $n = 120$. The relative frequency for the third class was 0.15. How many items of data fell into the third class?

ANSWER:
18

155. A relative frequency distribution was constructed for a sample of size $n = 150$. The relative frequency for the second class was 0.067. How many items of data fell into this class?

ANSWER:
10

156. In an ogive, what does the vertical scale identify?

ANSWER:
The vertical scale identifies either the cumulative frequencies or the cumulative relative frequencies.

157. In an ogive, what does the horizontal scale identify?

ANSWER:
The horizontal scale identifies the upper class boundaries. Until the upper boundary of a class has been reached, you cannot be sure you have accumulated all the data in that class. Therefore, the horizontal scale for an ogive is always based on the upper class boundaries.

158. Explain what is wrong with the statement, "The mean is always the best measure of central tendency."

ANSWER:
It depends on the type of data, and what would be an appropriate measure of central tendency.

159. A company found that the mean number of sales for the 20 salesmen during the past month was 8.5. What was the total number of sales for the salesmen?

ANSWER:
170

160. For a particular sample $\bar{x} = 14.7$ and s = 3.5. A new sample is formed by subtracting 2 from each value in the original sample. Find $\bar{x}$ for this new sample.

ANSWER:
$\bar{x} = 12.7$

161. Explain why it is possible to find the mean for the data of a quantitative variable, but not for a qualitative variable.

ANSWER:
Quantitative variable results in numbers for which arithmetic is meaningful; qualitative variable does not

162. Find the median height of cheerleaders of a college basketball team: 66, 69, 65, 63 and 67 inches.

ANSWER:
$\tilde{x}$ = 66 inches

163. Explain why the standard deviation is not always less than the variance and give an example.

ANSWER:

If $s < 1$, then s will be larger than s^2; example: s = 0.25, then s^2 = 0.0625.

164. Which of the three measures of variability, range, standard deviation, and variance, does *not* preserve the same unit of measurement as the observations themselves?

ANSWER:

Variance

165. If a sample has a standard deviation of 4.5, what is its variance?

ANSWER:
20.25

166. For a particular sample $\bar{x} = 14.7$ and s = 3.5. A new sample is formed by subtracting 2 from each value in the original sample. Find s for this new sample.

ANSWER:
s = 3.5

167. For a particular sample of size n = 10, the sample variance is 4.8 and $\bar{x} = 0.5$. For this sample, find $\sum x^2$.

ANSWER:
45.7

168. Why the sum of the deviations, $\sum(x - \bar{x})$, is always zero?

ANSWER:
$\sum(x - \bar{x})$, is always zero because the deviations of x values smaller than the mean (which are negative values) cancel out x values larger than the mean (which are positive).

169. Explain the meaning of the following statement "The data value x = 30 has a deviation value of 6."

ANSWER:
The value x = 30 is 6 larger than the mean.

170. Explain the meaning of the following statement "The data value x = 80 has a deviation value of -15."

ANSWER:
The value x = 80 is 15 smaller than the mean.

171. A particular standardized test has a mean score of 455 with a standard deviation of 112. A student scored 575 on this test. Determine the student's z-score.

ANSWER:
1.07

172. On a standardized test, a student's z-score was near zero. What does this tell us about the student's actual score on the test?

ANSWER:
The actual score was near the mean.

173. For a particular sample, the mean is 4.74, and the standard deviation is 3.10. What score in the sample has a z-score equal to -0.40?

ANSWER:

3.5

174. What statistical measure gives the range of the middle 50% of the data?

ANSWER:
Interquartile range

175. An aptitude test is known to have a mean score of 37.75 with a standard deviation equal to 3.5. A company requires a standard score of at least 1.5 for employment as one of its requirements. What must your test score be in order to be considered for employment?

ANSWER:
43 or larger

176. A normal distribution has a mean equal to 55.0 and a standard deviation equal to 7.5. Find the value of the midquartile.

ANSWER:
55.0

177. For a particular sample $\bar{x}$ = 4.2, one item in the sample is x = 4.8. This item has a z-score at 2.50. Find the sample standard deviation.

ANSWER:
s = 0.24

178. For a particular sample $\bar{x}$ = 4.4, an item in the sample is x = 3.4, and the z-score of this item is equal to –1.25. Find the sample variance.

ANSWER:
s^2 = 0.64

179. Determine your raw score on a test that has a sample mean of 65 and a sample variance of 121 if your instructor told you that your standard score is 1.50.

ANSWER:

$$z = \frac{x - \bar{x}}{s} \Rightarrow 1.50 = \frac{x - 65}{11} \Rightarrow x = 81.5$$

180. In general, the median, the midrange, and the midquartile are not necessarily the same value. Each is the middle value, but by different definitions of “middle”. What property does the distribution need for these three measures to all be the same value?

ANSWER:
The distribution of the data needs to be symmetric for these three measures to all be the same value.

181. What does it mean to say that x = 163 has a standard score of +1.60?

ANSWER:
It means that x = 163 is 1.60 standard deviations above the mean.

182. Determine your raw score on a test that has a sample mean of 74 and a sample standard deviation of 12 if your instructor told you that your standard score is -0.50.

ANSWER:

$$z = \frac{x - \bar{x}}{s} \Rightarrow -0.50 = \frac{x - 74}{12} \Rightarrow x = 68$$

183. What does it mean to say that a particular value of *x* has a *z* score of -1.94?

ANSWER:
It means that value of *x* is 1.94 standard deviations below the mean.

184. In general, the standard score is a measure of what?

ANSWER:
The standard score is a measure of the number of standard deviations from the mean.

Applied and Computational Questions

QUESTIONS 185 THROUGH 190 ARE BASED ON THE FOLLOWING INFORMATION:
The frequency distribution below gives the weight loss in pounds for 90 patients.

Class Number	*Class Boundaries*	*f*
1	$0.0 \le x < 5.0$	5
2	$5.0 \le x < 10.0$	12
3	$10.0 \le x < 15.0$	16
4	$15.0 \le x < 20.0$	27
5	$20.0 \le x < 25.0$	19
6	$25.0 \le x < 30.0$	9
7	$30.0 \le x \le 35.0$	2

185. What is the upper class boundary of the fifth class?

ANSWER:
25.0

186. What is the class width?

ANSWER:
5.0

187. What is the class mark of the third class?

ANSWER:
12.5

188. What is the value of $\sum f$?

ANSWER:
90

189. What is the sample size?

ANSWER:
90

190. Convert the above table to a relative frequency distribution.

ANSWER:

Class Number	*Class Boundaries*	*Relative frequency*
1	$0.0 \le x < 5.0$	0.056
2	$5.0 \le x < 10.0$	0.133
3	$10.0 \le x < 15.0$	0.178
4	$15.0 \le x < 20.0$	0.300
5	$20.0 \le x < 25.0$	0.211
6	$25.0 \le x < 30.0$	0.100
7	$30.0 \le x \le 35.0$	0.022

QUESTIONS 191 THROUGH 195 ARE BASED ON THE FOLLOWING INFORMATION:
A sample of families living in a large, suburban subdivision resulted in the following frequency distribution, where: x = number of children in the family.

x	f
0	8
1	11
2	23
3	21
4	13
5	7
6	2

191. What does the "3" represent?

ANSWER:
3 children per family for 21 families

192. What does the "7" represent?

ANSWER:
7 families with 5 children each

193. How many families were used to form this sample?

ANSWER:
85

194. How many children are included in this sample?

ANSWER:
219

195. Determine the mean number of children per family in the sample.

ANSWER:
$\overline{x}$ =219 / 85=2.58

196. A sample of twenty-five snow blowers of a given brand were filled with gasoline (one gallon) and allowed to run until the tank was empty. The times (in minutes) that the snow blowers operated were recorded as follows:

65	70	60	65	67	68	63	62	63	70	72	66	63
66	66	62	70	58	60	60	60	62	67	71	65	

Form a frequency distribution consisting of 5 classes.

ANSWER:

Class Boundaries	*Frequency*
$58 \le x < 61$	5
$61 \le x < 64$	6
$64 \le x < 67$	6
$67 \le x < 70$	3
$70 \le x \le 73$	5

QUESTIONS 197 THROUGH 201 ARE BASED ON THE FOLLOWING INFORMATION:

The frequency distribution below gives the daily high temperature for 40 consecutive winter days in northern Wisconsin.

Class Boundaries	*f*
$0 \le x < 3$	2
$3 \le x < 6$	4
$6 \le x < 9$	7
$9 \le x < 12$	10
$12 \le x < 15$	8
$15 \le x < 18$	6
$18 \le x \le 21$	3

197. In how many days was the daily high temperature between 9 and 12 degrees?

ANSWER:
10 days

198. Convert the above frequency distribution to a relative frequency distribution.

ANSWER:

Class Boundaries	*Relative frequency*
$0 \le x < 3$	0.050
$3 \le x < 6$	0.100
$6 \le x < 9$	0.175
$9 \le x < 12$	0.250
$12 \le x < 15$	0.200
$15 \le x < 18$	0.150
$18 \le x \le 21$	0.075

199. What is the proportion of days in which the daily high temperature was between 15 and 18?

ANSWER:
0.15

200. Construct the cumulative frequency distribution.

ANSWER:

Class Boundaries	*Cumulative frequency*
$0 \le x < 3$	2
$3 \le x < 6$	6
$6 \le x < 9$	13
$9 \le x < 12$	23
$12 \le x < 15$	31
$15 \le x < 18$	37
$18 \le x \le 21$	40

201. Construct the cumulative relative frequency.

ANSWER:

Class Boundaries	*Cumulative Relative frequency*
$0 \le x < 3$	0.050
$3 \le x < 6$	0.150
$6 \le x < 9$	0.325
$9 \le x < 12$	0.575
$12 \le x < 15$	0.775
$15 \le x < 18$	0.925
$18 \le x \le 21$	1.000

202. The following frequency distribution gives the pay ranges (in thousands of dollars) for all middle management personnel in large company.

Class Boundaries	*f*
$20 < x < 30$	4
$30 \le x < 40$	27
$40 \le x < 50$	29
$50 \le x < 60$	25
$60 \le x < 70$	17

Describe what shape a histogram of this data would have.

ANSWER:
Skewed to left (or negatively skewed)

QUESTIONS 203 THROUGH 206 ARE BASED ON THE FOLLOWING INFORMATION:

The ages of 50 students who are attending a community college in Iowa are shown below:

20	20	19	21	21	22	19	19	21	19
18	21	19	18	22	21	24	20	24	17
21	19	22	19	18	20	23	19	19	20
19	20	21	22	21	20	22	20	21	20
21	19	21	21	19	19	20	19	19	19

203. Prepare an ungrouped frequency distribution of these ages.

ANSWER:

Age	17	18	19	20	21	23	23	24
Frequency	1	3	16	10	12	5	1	2

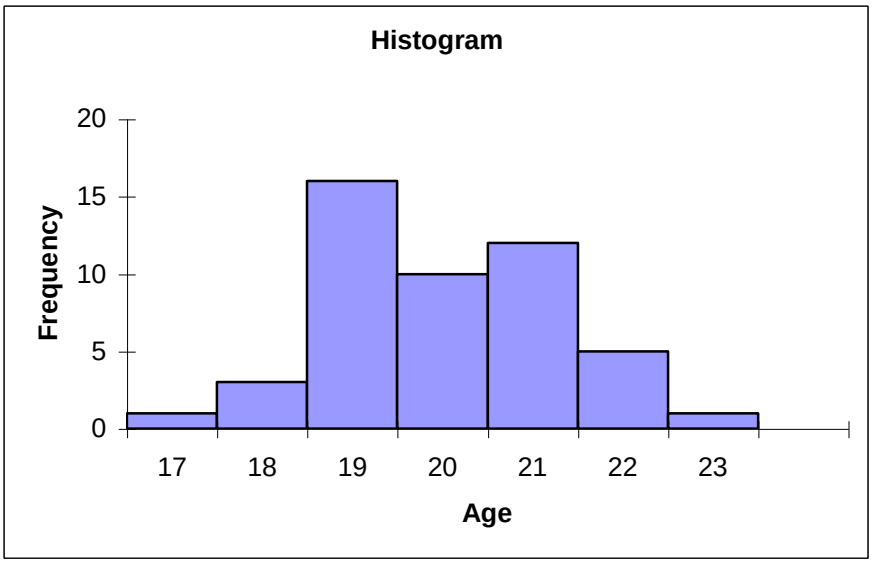
204. Prepare an ungrouped relative frequency distribution of the same data.

ANSWER:

Age	17	18	19	20	21	23	23	24
Rel. Freq.	0.02	0.06	0.32	0.20	0.24	0.10	0.02	0.04

205. Prepare a frequency histogram of these data.

ANSWER:



206. Prepare a cumulative relative frequency distribution of the same data.

ANSWER:

Age	17	18	19	20	21	23	23	24
Cum. rel. freq.	0.02	0.08	0.40	0.60	0.84	0.94	0.96	1.00

207. Briefly discuss the basic guidelines to follow in constructing a grouped frequency distribution.

ANSWER:
(a) Each class should be the same width.
(b) Classes (sometimes called bins) should be set up so that they do not overlap and so that each data belongs to exactly one class.
(c) Five to twelve classes are most desirable. (The square root of *n* is a reasonable guideline for the number of classes with samples of fewer than 125 data.)
(d) Use a system that takes advantage of a number pattern to guarantee accuracy.
(e) When it is convenient, an even class width is often advantageous.

208. The terms “symmetrical, uniform, skewed, J-shaped, bimodal, and normal” are usually used to describe histograms. Discuss each term briefly.

ANSWER:
Symmetrical: Both sides of this distribution are identical (halves are mirror images).
Uniform (rectangular): Every value appears with equal frequency.
Skewed: One tail is stretched out longer than the other. The direction of skewness is on the side of the longer tail.
J-Shaped: There is no tail on the side of the class with the highest frequency.

Bimodal: The two most populous classes are separated by one or more classes. This situation often implies that two populations are being sampled.
Normal: A symmetrical distribution is mounded up about the mean and becomes sparse at the extremes.

QUESTIONS 209 AND 210 ARE BASED ON THE FOLLOWING INFORMATION:
The following frequency distribution provides the number of managers and their annual salaries (in $1000):

Annual Salary ($1000)	15-25	25-35	35-45	45-55	55-65
Number of Managers	24	74	52	38	12

209. Prepare a cumulative frequency distribution for this frequency distribution.

ANSWER:

Class Boundaries	Cumulative Frequency
$15 \le x \le 25$	24
$15 \le x \le 25$	98
$15 \le x \le 25$	150
$15 \le x \le 25$	188
$15 \le x \le 25$	200

210. Prepare a cumulative relative frequency distribution for this frequency distribution.

ANSWER:

Class Boundaries	Cumulative Frequency
$15 \le x \le 25$	0.12
$15 \le x \le 25$	0.49
$15 \le x \le 25$	0.75
$15 \le x \le 25$	0.94
$15 \le x \le 25$	1.00

QUESTIONS 211 THROUGH 214 ARE BASED ON THE FOLLOWING INFORMATION:
The players on a professional soccer team scored 40 goals during last season.

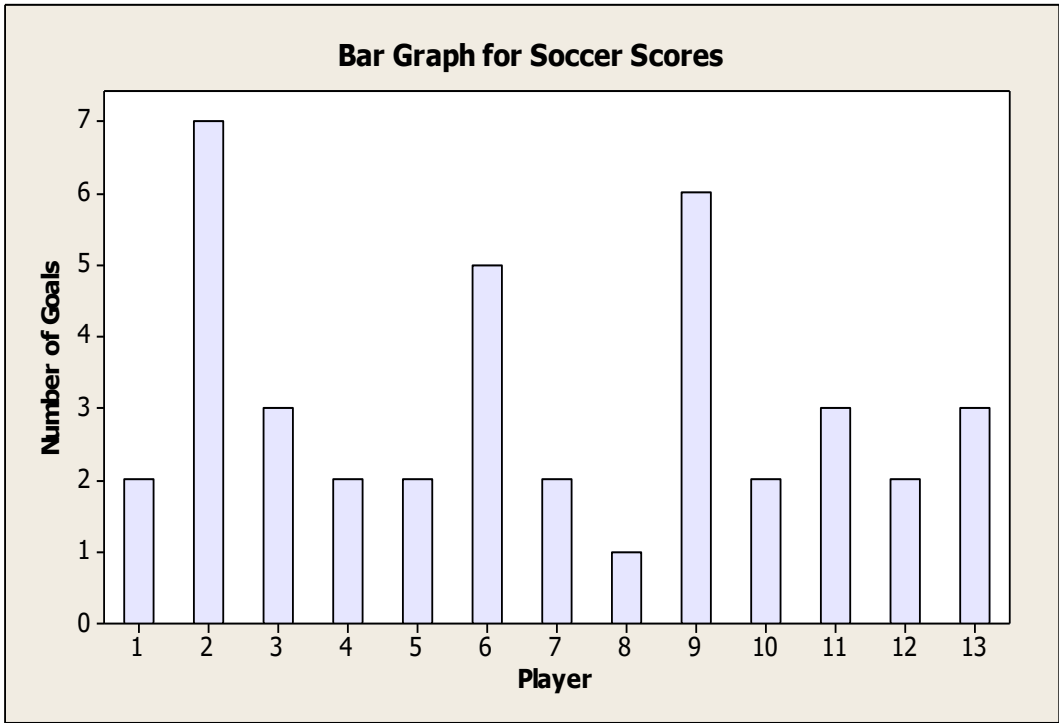
Player	1	2	3	4	5	6	7	8	9	10	11	12	13
Goals	2	7	3	2	2	5	2	1	6	2	3	2	3

211. If you want to show the number of goals scored by each player, would it be more appropriate to display this information on a bar graph or a histogram? Explain.

ANSWER:
In order to show the number of goals scored by each player, it would be more appropriate to display this information on a bar graph

212. Construct the appropriate graph for question 211.

ANSWER:

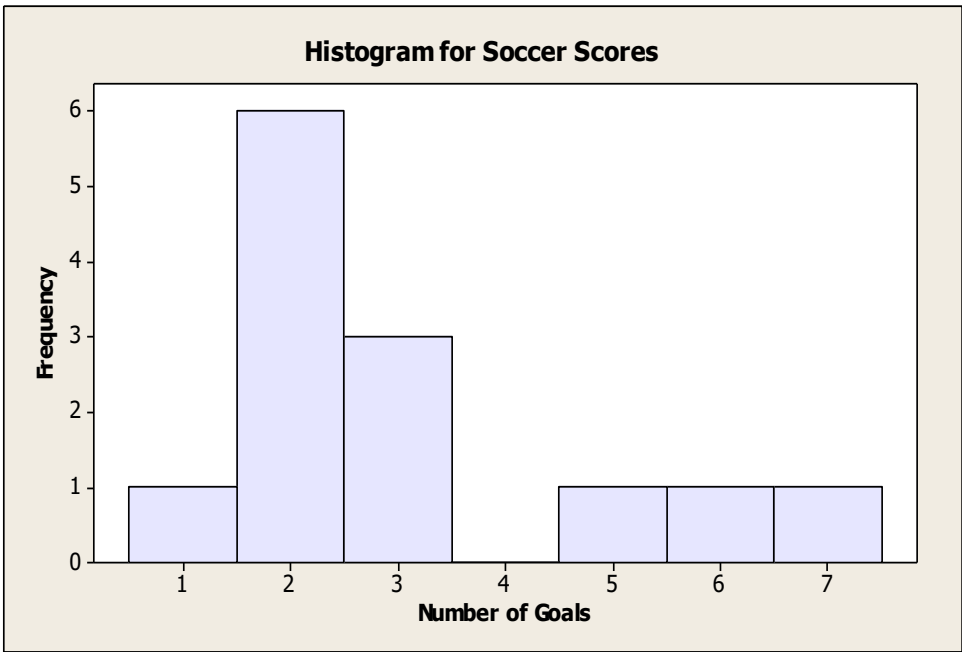


213. If you wanted to show (emphasize) the distribution of scoring by the team, would it be more appropriate to display this information on a bar graph or a histogram? Explain.

ANSWER:
If we want to emphasize the distribution of scoring by the team, it would be more appropriate to display this information on a histogram.

214. Construct the appropriate graph for question 213.

ANSWER:



QUESTIONS 215 AND 216 ARE BASED ON THE FOLLOWING INFORMATION:
A sample of twenty-five snow blowers of a given brand were filled with gasoline (one gallon) and allowed to run until the tank was empty. The times (in minutes) that the snow blowers operated were recorded as follows:

65 70 60 65 67 68 63 62 63 70 72 66 63
66 66 62 70 58 60 60 60 62 67 71 65

215. Construct a stem-and-leaf display.

ANSWER:

Stems	*Leaves*
5	8
6	0 0 0 0 2 2 2 3 3 3 5 5 5 6 6 6 7 7 8
7	0 0 0 1 2

216. Find the mean, median, mode, and midrange.

ANSWER:
Mean = 64.84, Median = 65.0, Mode = 60.0, Midrange = 65.0

217. Nine households had the following number of children per household: 2, 0, 2, 2, 1, 2, 4, 3, 2. Find the mean, median, mode, and midrange for these data.

ANSWER:
Mean = 2, Median = 2, Mode = 2, Midrange = 2

QUESTIONS 218 THROUGH 219 ARE BASED ON THE FOLLOWING INFORMATION:
The commuting distance was determined for each of 10 employees at Acme manufacturing. One of the employees lives in another town and has a large commuting distance. The 10 distances were as follows: 5, 10, 7, 15, 10, 12, 8, 120, 20, 18.

218. Find the mean distance.

ANSWER:
22.5

219. Find the median distance.

ANSWER:
11

220. Which measurement, A or B, is most representative for the data? Why?

ANSWER:
Median; since median is not affected by outliers.

221. Consider the following sample: 26, 49, 9, 42, 60, 11, 43, 26, 30, and 14. Find the mean.

ANSWER:
31.0

222. Consider the following sample: 26, 49, 9, 42, 60, 11, 43, 26, 30, and 14. Find the median.

ANSWER:
28.0

223. Consider the following sample: 26, 49, 9, 42, 60, 11, 43, 26, 30, and 14. Find the midrange.

ANSWER:
34.5

224. For a particular sample of 50 scores on a statistics exam, the following results were obtained:

Mean = 78	Midrange = 72	Third quartile = 94	Mode = 84
Median = 80	Standard deviation = 11	Range = 52	First quartile = 68

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

ANSWER:
84; since it is the mode

225. If a sample with a mean of 10.5 and a standard deviation of 2.30 has every item multiplied by 10, find the mean of the new sample.

ANSWER:
105

226. 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 of the new sample.

ANSWER:
10.0

227. Find the mean, median, mode, and midrange for the following data:

x	*f*
10	2
11	4
12	7
15	3
20	1

ANSWER:
Mean = 12.5, Median = 12, Mode = 12, Midrange = 15

228. A student computed the mean of a particular sample to be 40.0. After computing the mean, he discovered that he forgot to include the number 36 in the sample. When this number was included, the sample mean changed to 39.5. What is the sample size when the number 36 is correctly included in the sample?

ANSWER:
$n = 8$

QUESTIONS 229 THROUGH 235 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.

229. 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 $\overline{x} = 500 / 5 = 100$.

230. 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.

231. 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.

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

ANSWER:
Many different answers are possible. Need any two numbers that total 140 for the extreme values *L* and *H*, where one is 100 or larger. For example, we may choose the numbers 40, 50, and 60. Thus the five numbers are 40, 50, 60, 70, 100, and midrange = (*L*+*H*)/2 = (40+100)/2 = 70.

233. 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, 200. Hence, $\overline{x}$ = 500/5 = 100, and the median is 70.

234. 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, 156. Thus the mode = 87, and $\overline{x}$ = 500 / 5 = 100.

235. 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.

236. The Next Door Store kept track of the number of paying customers it had during the noon hour each day for 100 days. The following are the resulting statistics rounded to the nearest integer: Mean = 95, Median = 97, Mode = 98, First quartile = 85, Third quartile = 107, Midrange = 93, Range = 56, and Standard deviation = 12. The Next Door Store served what number of paying customers during the noon hour more often than any other number? Explain how you determined your answer.

ANSWER:
98 customers; this is the mode.

237. A statistics test was given with the following results:

80, 69, 92, 75, 88, 37, 98, 92, 90, 81, 32, 50, 59, 66, 67, 66

Find the range, standard deviation, and variance for the scores.

ANSWER:
Range = 66, s = 19.64, s^2 = 385.85

QUESTIONS 238 THROUGH 245 ARE BASED ON THE FOLLOWING INFORMATION:
Starting with a sample of two values 75 and 105, add three data values to your sample to obtain a new sample with certain statistics.

238. What are the three data values such that the new sample has a mean of 110 (Hint: Many different answers are possible). Justify your answer.

ANSWER:
$\sum x$ needs to be 550; therefore, need any three numbers that total 370, such as 110, 120, and 140. Hence, the mean $\bar{x} = \sum x / n$ = 550 / 5 = 110

239. What are the three data values such that the new sample has a median of 75 (Hint: Many different answers are possible). Justify your answer.

ANSWER:
Need two numbers smaller than 75 and one number larger. For example, choose the numbers 60, 70, and 80. Hence, the five data values are 60, 70, 75, 80, 105, and d($\tilde{x}$) = (n+1)/2 = (5+1)/2 = 3^{rd} value; therefore the median $\tilde{x}$ = 75.

240. What are the three data values such that the new sample has a mode of 85 (Hint: Many different answers are possible). Justify your answer.

ANSWER:
Choose three numbers, each is 85. Hence the five data values are 75, 85, 85, 85, 105, and the mode = 85.

241. What are the three data values such that the new sample has a midrange of 80 (Hint: Many different answers are possible). Justify your answer.

ANSWER:
Need any two numbers that total 160 for the extreme values where one is 105 or larger. For example, choose the values 40, 50, and 120. Hence the five data values are 40, 50, 75, 105, and 120. Therefore, midrange = (L+H)/2 = (40+120)/2 = 80.

242. What are the three data values such that the new sample has a mean of 110 and a median of 75 (Hint: Many different answers are possible). Justify your answer.

ANSWER:
Need two numbers smaller than 75 and one larger than 75 so that their total is 365. For example, choose the values 65, 70, and 230. Hence the five data values are 65, 70, 75, 105, and 230. Hence the mean $\bar{x} = \sum x / n$ = 550 / 5 = 110, and d($\tilde{x}$) = (n+1)/2 = (5+1)/2 = 3^{rd} value; therefore the median $\tilde{x}$ = 75.

243. What are the three data values such that the new sample has a mean of 110 and a mode of 80 (Hint: Many different answers are possible). Justify your answer.

ANSWER:
Need two numbers of 80 and a third number large enough so that the total of all five values is 550. Then the third number must be 210. Hence, the five values are 75, 80, 80, 105, and 210. Hence, the mean $\bar{x} = \sum x / n$ = 550 / 5 = 110, and the mode = 80.

244. What are the three data values such that the new sample has a mean of 110 and a midrange of 80 (Hint: Many different answers are possible). Justify your answer.

ANSWER:
We started with the data values 75 and 105. A mean of 110 requires the five data values to total 550 and a midrange of 80 requires the total of the lowest value *L* and the highest value *H* to be 160. The sum of 75 and 105 is 180; hence, the total of the other three remaining numbers is 370. Since *L* + *H* must be 160, then the fifth number must be 210, which would then become *H* and change the value of the midrange. So, this situation is impossible!

245. What are the three data values such that the new sample has a mean of 110, a median of 75, and a mode of 85 (Hint: Many different answers are possible). Justify your answer.

ANSWER:
There must be two 85's in order to have a mode of 85, and there can only be two data values larger than 75 in order for 75 to be the median, but since 105 is one of the starting numbers, then we have three data values larger than 75; namely 85, 85, and 105. As a result, 75 can't be the median. So, this situation is impossible!.

246. A sample of twenty-five snow blowers of a given brand were filled with gasoline (one gallon) and allowed to run until the tank was empty. The times (in minutes) that the snow blowers operated were recorded as follows:

65	70	60	65	67	68	63	62	63	70	72	66	63
66	66	62	70	58	60	60	60	62	67	71	65	

Find the standard deviation and the range.

ANSWER:
Standard deviation *s* = 3.91, Range=14

247. A group of children had the following heights in inches: 45, 46, 42, 56, 37, 50, 51, 50, 47, 47. Find the range, standard deviation, and variance for the scores.

ANSWER:

Range = 19, s = 5.22, and s^2 = 27.211

QUESTIONS 248 AND 249 ARE BASED ON THE FOLLOWING INFORMATION:
Consider the following sample: 26, 49, 9, 42, 60, 11, 43, 26, 30, and 14.

248. Find the variance.

ANSWER:
294.89

249. Find the standard deviation

ANSWER:
17.17

QUESTIONS 250 AND 251 ARE BASED ON THE FOLLOWING INFORMATION:
For a particular sample of 50 scores on a statistics exam, the following results were obtained:

Mean = 78	Midrange = 72	Third quartile = 94	Mode = 84
Median = 80	Standard deviation = 11	Range = 52	First quartile = 68

250. What was the highest score earned on the exam?

ANSWER:
98 [Recall that: Midrange = (L+H)/2, and Range = H-L]

251. What was the lowest score earned on the exam?

ANSWER:
46

252. If a sample with a mean of 10.5 and a standard deviation of 2.30 has every item multiplied by 10, find the variance of the new sample.

ANSWER:
529

QUESTIONS 253 AND 254 ARE BASED ON THE FOLLOWING INFORMATION:
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.

253. Find the standard deviation of the new sample.

ANSWER:
1.20

254. Find the variance of the new sample.

ANSWER:
1.44

255. 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. Sample 1, *s* = 5.17; Sample 2, *s* = 4.66; Sample 3, *s* = 5.54. Since sample 2 has the smallest standard deviation, data most closely grouped about its mean.

256. The mean for 50 pressure readings equals 5.5, and the sum of the squares of the readings equals 1622.75. Find the standard deviation of these pressure readings.

ANSWER:
$s = \sqrt{[\sum x^2 - n\overline{x}^2]/(n-1)} = \sqrt{[1622.75 - 50(5.5)^2]/49} = \sqrt{2.25}$ = 1.5

257. A set of 25 measurements has a mean of 24.5 and a standard deviation equal to 4.0. Find $\sum x^2$.

ANSWER:
$s = \sqrt{[\sum x^2 - n\overline{x}^2]/(n-1)} \Rightarrow$ 4.0 = $\sqrt{[\sum x^2 - 25(24.5)^2]/24} \Rightarrow \sum x^2$ = 15,390.25

QUESTIONS 258 AND 259 ARE BASED ON THE FOLLOWING INFORMATION:
Consider the following data: 21, 41, 41, 36, 39, 23, 30, 30, 34, 31, 26, 25, 29, 28, 36.

258. Find the mean.

ANSWER:
$\overline{x}$ = 31.3

259. Find the standard deviation.

ANSWER:
s = 6.3

260. Consider the following two sets of data:

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

Both sets have the same mean $\overline{x} = 50$. Compare the following measures for both sets: $\sum (x - \overline{x})$, *SS*(*x*), and range. Comment on the meaning of these comparisons.

ANSWER:
Set 1:

x	$x - \overline{x}$	$(x - \overline{x})^2$
45	-5	25
55	+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 $SS(x)$ [recall $SS(x)=\sum(x-\bar{x})^2$] and range reflect the fact that there is more variability in the data forming set 2 than in the data of set 1. $\sum(x-\bar{x})=0$ for both sets of data (in fact this is always true for any data).

QUESTIONS 261 AND 262 ARE BASED ON THE FOLLOWING INFORMATION:

A sample of twenty-five snow blowers of a given brand were filled with gasoline (one gallon) and allowed to run until the tank was empty. The times (in minutes) that the snow blowers operated were recorded as follows:

65 70 60 65 67 68 63 62 63 70 72 66 63
66 66 62 70 58 60 60 60 62 67 71 65

261. Find the first quartile

ANSWER:
$Q_1 = 62$

262. Find the ninetieth percentile.

ANSWER:
$P_{90} = 70$

263. For a particular sample of 50 scores on a statistics exam, the following results were obtained: Mean = 78, Midrange = 72, third quartile = 94, Mode = 84, Median = 80, Standard deviation = 11, Range = 52, and first quartile = 68. How many students scored between 68 and 94 on the exam?

ANSWER:
25

QUESTIONS 264 THROUGH 266 ARE BASED ON THE FOLLOWING INFORMATION:

Consider the following sample of size *n* = 65, ordered from smallest to largest:

124 127 128 129 133 134 137 139 141 143
147 148 156 159 163 166 169 170 173 179
199 201 207 210 213 217 219 222 225 228
234 238 244 259 261 262 263 264 266 268
279 280 286 298 299 305 306 307 311 313

320 328 333 345 350 351 361 362 363 364
378 388 390 400 417

264. Prepare a five-number summary for this set of data.

ANSWER:
L = 124, $Q_1 = 169$, $\tilde{x} = 244$, $Q_3 = 311$, $H = 417$

265. Find the 80th percentile.

ANSWER:
P_{80} =330.5

266. Find the 29th percentile.

ANSWER:
P_{29} = 173115.

QUESTIONS 267 AND 268 ARE BASED ON THE FOLLOWING INFORMATION:
Consider the following sample of size, n = 60 ordered from smallest to largest:

24	27	28	29	33	34	37	39	41	43	47	48
56	59	63	66	69	70	73	79	99	21	27	10
13	17	19	22	25	28	34	38	44	59	61	62
63	64	66	68	79	80	86	98	99	35	36	37
11	13	20	28	33	45	50	51	61	62	63	64

267. Prepare a five-number summary for this set of data.

ANSWER:
L =10, $Q_1 = 28$, $\tilde{x} = 44.5$, $Q_3 = 63.25$, H = 99

268. Find the 20th percentile.

ANSWER:
P_{20} =27

269. Consider the sample 9, 11, 17, 23, 26, 38, 47. Find the z-score for the data point of "11."

ANSWER:
$\bar{x} = 24.4286$ and $s = 13.9745$. Then, $z = (x - \bar{x}) / s$ = (11.0 – 24.4286) / 13.9745 = -0.96

270. In which of these situations (A, B, or C) is the x-value lowest in relation to the sample from which it comes? These samples come from three different populations.

Situation A: $x = 6$, $\bar{x} = 20.0$, $s = 9.0$
Situation B: $x = 350$, $\bar{x} = 400.0$, $s = 20.0$
Situation C: $x = 1.6$, $\bar{x} = 2.00$, $s = 0.30$

ANSWER:
In situations A, B, and C; z = -1.56, -2.50, and –1.33, respectively. In situation B we see the lowest z-score of –2.50. Therefore, the x-value in B is lowest in relation to the sample from which it comes.

271. Find the first quartile and the third quartile for the following data:

2.1, 2.1, 2.2, 2.4, 2.5, 2.5, 2.5, 2.5, 2.6, 2.6, 2.6, 2.7, 2.7,
2.7, 2.8, 2.9, 3.0, 3.0, 3.2, 3.2, 3.3, 3.3, 3.5, 3.6, 4.0

ANSWER:
$Q_1 = 2.5,\ Q_3 = 3.2$

272. Consider the following measurements of ozone concentration (in ppm):

11.1, 11.5, 11.9, 12.0, 11.6, 12.2, 11.9, 12.5, 12.8, 19.0,
10.9, 11.6, 12.7, 5.0, 11.5, 12.6, 19.5, 12.7, 4.0, 19.1

The mean equals 12.31 and the variance equals 14.2884. Find the standard score for the smallest and largest data values.

ANSWER:
The smallest data value is 4.0, and its *z*-score is -2.198, while the largest data value is 19.5 and its *z*-score is 1.902.

273. Use the following stem-and-leaf display to find the tenth percentile for the distribution of lengths:

Stems	*Leaves*
2.1	0 2 1
2.3	3 6 5 2 1 1
2.4	1 1
2.5	0 1 2
2.7	7 7 8
3.1	2 2 4
3.5	1 1 2 1 1

ANSWER:
$P_{10} = 2.12$

274. The interquartile range (IQR) of a set of measurements is defined to be the difference between the upper and lower quartiles. Find the IQR for the following HLT scores that measure the degree of hostility: 80, 70, 63, 92, 81, 76, 78, 88, 70, 83, 74, 77, and 85.

ANSWER:
IQR = $Q_3 - Q_1$ = 83 – 74 = 9

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

ANSWER:
$L = x_1,\quad Q_1 = x_{17},\quad \tilde{x} = x_{34},\quad Q_3 = x_{51},\quad H = x_{67}$

276. What does it mean to say that *x* = 152 have a standard score of +1.5?

ANSWER:
152 is one and one-half standard deviations above the mean.

277. What does it mean to say that a particular value of *x* has a *z*-score of –2.1?

ANSWER:
The score is 2.1 standard deviations below the mean.

278. In general, the standard score is a measure of what?

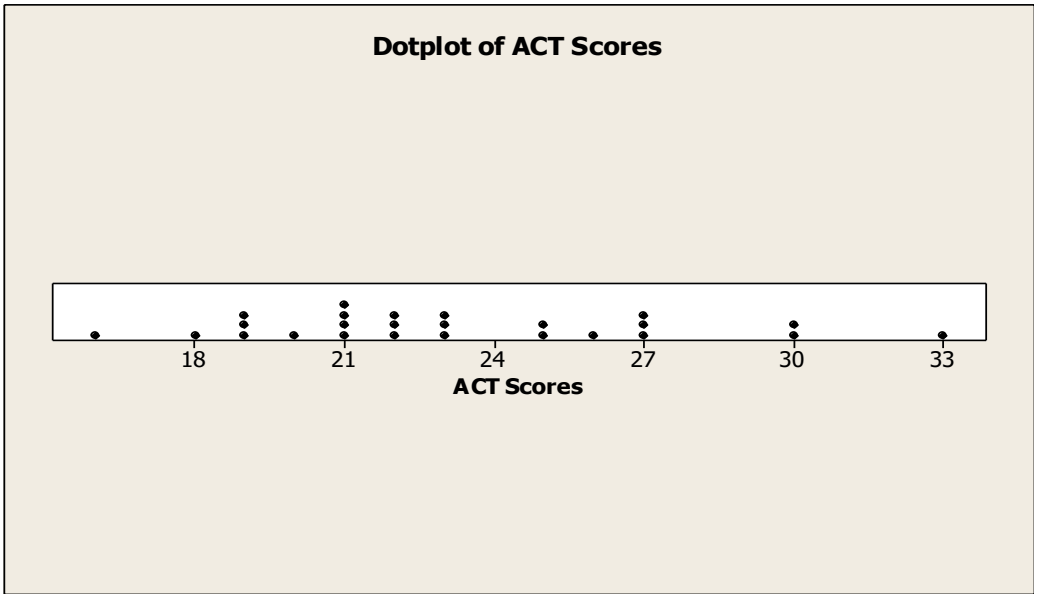
ANSWER:
The standard score is a measure of the number of standard deviations from the mean.

QUESTIONS 282 THROUGH 293 ARE BASED ON THE FOLLOWING INFORMATION:
Below are the ACT scores attained by the 25 members of a local high school graduating class.

23	26	25	19	33	21	21	22	21	27
19	25	18	23	22	30	27	27	23	16
21	19	20	30	22					

279. Draw a dotplot of the ACT scores.

ANSWER:



280. Using the concept of depth, describe the position of 26 in the set of 25 ACT scores in two different ways.

ANSWER:
The data values in ascending are:

16	18	19	19	19	20	21	21	21	21
22	22	22	23	23	23	25	25	26	27
27	27	30	30	33					

Therefore, the value 26 is in the 19th position from *L* = 16, and in the 7th position from *H* = 33.

281. Find P_5 for the ACT scores.

ANSWER:
nk / 100 =(25)(5) / 100 = 1.25. Hence, *d*(P_5) = 2, and P_5 = 18

282. Find P_{10} for the ACT scores.

ANSWER:
nk / 100 =(25)(10) / 100 = 2.5. Hence *d*(P_{10}) = 3, and P_{10} = 19

283. Find P_{20} for the ACT scores.

ANSWER:
nk / 100 =(25)(20) / 100 = 5. Hence *d*(P_{20}) = 5.5, and P_{20} =(19+20)/2 = 19.5

284. Find P_{99} for the ACT scores.

ANSWER:
Since *k* = 99 > 50, subtract 99 from 100 and use 100 – *k* in place of *k* to determine the depth, which is then counted from the largest-valued data *H*. Therefore, *n*(100 – *k*) / 100 = 25 (1) / 100 = 0.25; then *d*(P_{99}) = 1, and P_{99} = 33

285. Find P_{90} for the ACT scores.

ANSWER:
Since *k* = 90 > 50, subtract 90 from 100 and use 100 – *k* in place of *k* to determine the depth, which is then counted from the largest-valued data *H*. Therefore, P_{90} : *n*(100 – *k*) / 100 = 25 (10) / 100 = 2.5; then *d*(P_{90}) = 3, and P_{90} = 30

286. Find P_{80} for the ACT scores.

ANSWER:
Since *k* = 80 > 50, subtract 80 from 100 and use 100 – *k* in place of *k* to determine the depth, which is then counted from the largest-valued data *H*. Therefore, *n*(100 – *k*) / 100 = 25 (20) / 100 = 5; then *d*(P_{80}) = 5.5, and P_{80} = (27+27) / 2 = 27

287. Find the first quartile, Q_1, for the ACT scores.

ANSWER:
nk / 100 =(25)(25) / 100 = 6.25. Hence *d*(Q_1) = 7, and Q_1 = 21

288. Find the second quartile, Q_2 for the ACT scores.

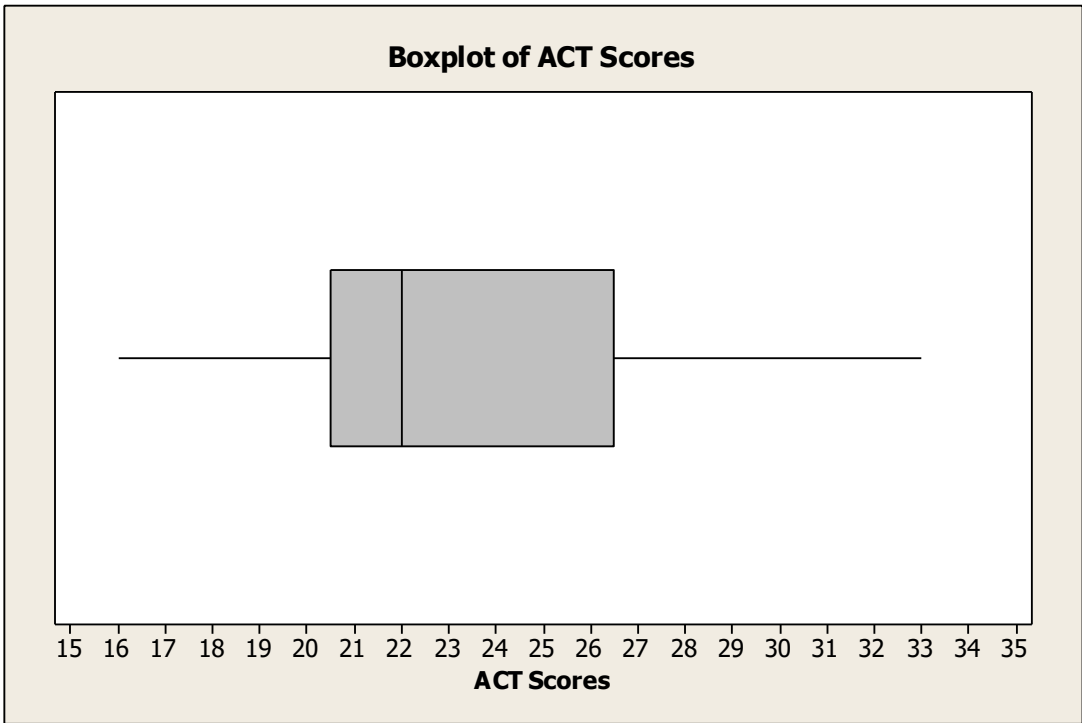
ANSWER:
nk / 100 =(25)(50) / 100 = 12.5. Hence *d*(Q_2) = 13, and Q_2 = 22

289. Find the third quartile, Q_3, for the ACT scores.

ANSWER:
Since *k* = 75 > 50, subtract 75 from 100 and use 100 – *k* in place of *k* to determine the depth, which is then counted from the largest-valued data *H*. Therefore, *n*(100 – *k*) / 100 = 25 (25) / 100 = 6.25; then *d*(Q_3) = 7, and Q_3 = 26.

290. Use Minitab to find the 5-number summary and draw a box-and-whiskers display.

ANSWER:
The five number summary reported by Minitab are: L = 16, Q_1 = 20.5, Q_2 = 22, Q_3 = 26.5, and H = 33.
Note that the values of Q_1 and Q_3 reported by Minitab are slightly different compared to our earlier calculations that showed Q_1 = 21 and Q_3 = 26.



291. The Next Door Store kept track of the number of paying customers it had during the noon hour each day for 100 days. The following are the resulting statistics rounded to the nearest integer:

Mean = 95	Median = 97	Mode = 98
First quartile = 85	Third quartile = 107	Midrange = 93
Range = 56	Standard deviation = 12	

On how many days were there between 85 and 107 paying customers during the noon hour? Explain how you determined your answer.

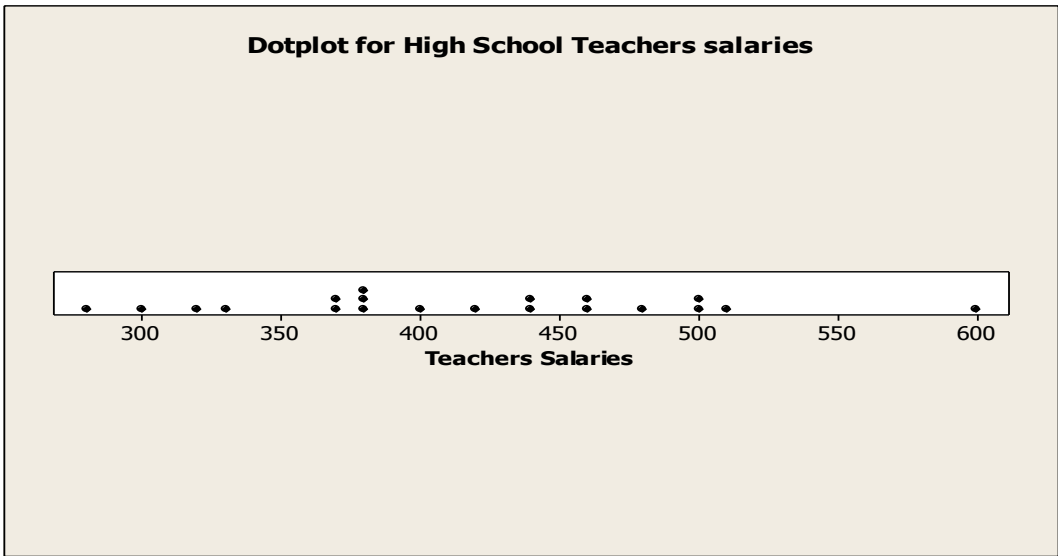
ANSWER:
50 days; since 50% of the 100 days fall between the first and third quartiles.

QUESTIONS 292 THROUGH 297 ARE BASED ON THE FOLLOWING INFORMATION:
The annual salaries (in $100) of high school teachers employed at one of the high schools in Kent County, Michigan are listed below:

600	440	461	419	397	477	464	275	507	497
332	373	440	373	501	382	377	301	323	383

292. Draw a dotplot of the salaries.

ANSWER:



293. Using the concept of depth, describe the position of 332 in the set of 20 salaries in two different ways.

ANSWER:
The data values in ascending are:

275	301	323	332	373	373	377	382	383	397
419	440	440	461	464	477	497	501	507	600

Therefore, the value 332 is in the 4th position from *L* = 270, and in the 17th position from *H* = 33.

294. Find the first quartile for these salaries, and interpret the result.

ANSWER:
nk / 100 =(20)(25) / 100 =5.0. Hence *d*(Q_1) =5.5, and Q_1 = (373+373)/2 = 373 or \$37,300. This means that at most 25% of high school teachers' salaries are lower than \$37,300 and at most 75% are higher.

295. Find the third quartile for these salaries, and interpret the result.

ANSWER:
Since *k* = 75 > 50, subtract 75 from 100 and use 100 – *k* in place of *k* to determine the depth, which is then counted from the largest-valued data *H*. Therefore,
n(100 – *k*) / 100 = 20 (25) / 100 = 5.0; then *d* (Q_3) = 5.5, and Q_3 = (464+477)/2 = 470.5 or \$47,050.
This means that at most 75% of high school teachers' salaries are lower than \$47,50 and at most 25% are higher.

296. Find the midquartile for these salaries, and interpret the result.

ANSWER:
Midquartile = ($Q_1 + Q_3$) / 2 = (373 + 470.5) / 2 = 421.75 or \$42,175.

This means that the salary midway between the first and third quartile is $42,175.

297. Find the interquartile range for these salaries, and interpret the result.

ANSWER:
Interquartile range = Q_3 - Q_1 **=** 470.5 – 373 = 97.5 or $9,750.
This means that the range of the middle 50% of the salaries is $9,750.

Sections 2.6 and 2.7

True-False Questions

298. *Chebyshev's* Theorem says that within two standard deviations of the mean, you will always find at least 89% of the data.
ANSWER: F

299. The Empirical Rule can be used to determine whether or not a set of data is approximately normally distributed.
ANSWER: T

300. For a bell-shaped distribution, the range will be approximately equal to six standard deviations.
ANSWER: T

301. The standard deviation is a kind of yardstick by which we can compare the variability of one set of data with another.
ANSWER: T

302. 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 Chebyshev's Theorem.
ANSWER: T

303. The Empirical Rule applies specifically to a normal (bell-shaped) distribution, but it is frequently applied as an interpretive guide to any mounded distribution.
ANSWER: T

304. The Empirical Rule applies to any distribution, regardless of its shape, as an interpretive guide to the distribution.
ANSWER: F

305. The Empirical Rule can be used to determine whether or not a set of data is approximately normally distributed.
ANSWER: T

306. The normal probability plot is an ogive drawn on probability paper.
ANSWER: T

307. The normal probability plot is a Dotplot drawn on probability paper.
ANSWER: F

308. In the event that the data do not display an approximately normal distribution, Chebyshev's Theorem gives us information about how much of the data will fall within intervals centered at the mean for all distributions.
ANSWER: T

309. Graphs in which the frequency scale starts at zero tend to emphasize the size of the numbers involved.
ANSWER: T

310. Graphs that are chopped off may tend to emphasize the variation in the numbers without regard to the actual size of the numbers.
ANSWER: T

311. Truncating scales on graphs often leads to misleading visual impressions.
ANSWER: T

Multiple-Choice Questions

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

A) Range is a measure of dispersion.
B) Chebyshev's Theorem applies only to non-normal distributions.
C) The sum of ($x - \overline{x}$) will always be zero.
D) The calculation of the range does not consider all values.
ANSWER: B

313. According to the Empirical Rule, if the variable is normally distributed, then within one standard deviation of the mean, there well be approximately:

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

314. The proportion of any distribution that lies 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

Short-Answer Questions

315. According to Chebyshev's Theorem, what percent of a set of data will be more than three standard deviations from the mean?

ANSWER:
About 11%

316. According to the Empirical Rule, at least what percent of a set of data will lie within two standard deviations from the mean?

ANSWER:
Approximately 95%

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

ANSWER:
55.0 and 145.0

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

ANSWER:
96%

319. 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

320. Complete the following statement: According to the Empirical Rule, _____ of the data for any distribution will occur within one standard deviations of the mean of the distribution.

ANSWER:
68%

321. The lifetimes of electronic components have a mean equal to 2.5 years and a standard deviation equal to 0.2 years. Within what time interval will at least 75% of the lifetimes fall?

ANSWER:
2.1 years to 2.9 years

322. A set of measurements has a mean equal to 35.5 and a standard deviation equal to 3.0. At least what percent of the data falls between 31.0 and 40.0?

ANSWER:
55.6%

323. For a normal distribution, a value that is one standard deviation above the mean would be approximately the same as what percentile?

ANSWER:
Eighty-fourth percentile

324. According to Chebyshev's Theorem, how many standard deviations on both sides of the mean do you need to go so that at least 96% of the distribution is covered?

ANSWER:
Five

325. According to Chebyshev's Theorem at least 75% of all the data in a particular sample lies between 74.5 and 82.9. Find the sample mean for this sample.

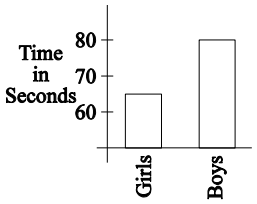
ANSWER:
$\bar{x} = 78.7$

326. According to Chebyshev's Theorem at least 75% of all the data in a particular sample lies between 74.5 and 82.9. Find the sample standard deviation for this sample.

ANSWER:
$s = 2.1$

Applied and Computational Questions

327. The bar graph below compares the mean time in seconds for seven-year-old girls to complete a certain task to the mean time in seconds for seven-year-old boys to complete the same task. There is statistical deception here. Explain what is deceptive about the bar graph.



ANSWER:
From the graph it appears that the time for the boys is twice the time for the girls. However, the time for the boys is 80 seconds while the time for the girls is 65 seconds. The deception is caused by the vertical scale not starting at zero.

328. 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:
$\bar{x} - 3s = 24.2, \text{ and } \bar{x} + 3s = 69.2 \Rightarrow \bar{x} = 46.7, \text{ and } s = 7.5$

QUESTIONS 329 AND 330 ARE BASED ON THE FOLLOWING INFORMATION:
The average clean-up time for a crew of a medium-size firm is 80.0 hours and the standard deviation is 6.5 hours. Assuming that the Empirical Rule is appropriate.

329. What proportion of the time will it take the clean-up crew 93.0 or more hours to clean the plant?

ANSWER:
z = (93 -80) / 6.5 = 2. Therefore, 93.0 is 2 standard deviations above the mean. Hence, 2.5% of the time more than 93.0 hours will be required.

330. The total clean-up time will fall within what interval 95% of the time?

ANSWER:
95% of the time, the total clean-up time will fall within 2 standard deviations of the mean; that is 80.0 $\pm$ 2 (6.5) or from 67 to 93 hours.

QUESTIONS 331 AND 332 ARE BASED ON THE FOLLOWING INFORMATION:

Chebyshev's Theorem can be stated in an equivalent form to that given in your book. For example, to say "at least 75% of the data fall within two standard deviations of the mean" is equivalent to stating that "at most, 25% will be more than two standard deviations away from the mean".

331. At most, what percentage of a distribution will be three or more standard deviations from the mean?

ANSWER:
At most 11%

332. At most, what percentage of a distribution will be four or more standard deviations from the mean?

ANSWER:
At most 6.25%

333. The Next Door Store kept track of the number of paying customers it had during the noon hour each day for 100 days. The following are the resulting statistics rounded to the nearest integer:

Mean = 95	Median = 97	Mode = 98
First quartile = 85	Third quartile = 107	Midrange = 93
Range = 56	Standard deviation = 12	

For how many of the 100 days was the number of paying customers within three standard deviations of the mean ($\overline{x} \pm 3s$) ? Explain how you determined your answer.

ANSWER:
According to Chebyshev's Theorem, the proportion of any distribution that lies within 3 standard deviations of the mean is at least 89%. Therefore, we should expect in at least 89 of the 100 days that the number of paying customers was within three standard deviations of the mean.

QUESTIONS 334 AND 335 ARE BASED ON THE FOLLOWING INFORMATION:

The mean lifetime of a certain tire is 50,000 miles and the standard deviation is 2,500 miles.

334. If we assume the mileages are normally distributed, approximately what percentage of all such tires will last between 42,500 and 57,500 miles?

ANSWER:
According to the Empirical Rule, approximately 99.7% of all such tires will last between 42,500 and 57,500 miles (i.e., within three standard deviations of the mean).

335. If we assume nothing about the shape of distribution, approximately what percentage of all such tires will last between 42,500 and 57,500 miles?

ANSWER:
According to Chebyshev's Theorem, at least 89% of all such tires will last between 42,500 and 57,500 miles (i.e., within three standard deviations of the mean).