The Wall Street Journal [illegible]
June 1995): B11.

■ 3-113 What are the mean and the median cost per passenger? Which would be the better figure to use for a new airline in developing its business plan? Merrill Lynch, the leading retail stock brokerage firm, commissioned a study of the wealth of American families. Using U.S. Census data from more than 38,000 families, Dr. Joseph Anderson concluded that the median level of financial assets (investable funds, excluding home equity) was $1,000, and the mean was approximately $30,000. Since $1,000 is less than the minimum amount needed to effectively invest in stocks, should the firm's managers think about getting out of selling stocks to the general public?

Source: B. Wysocki, "Many Baby Boomers Save Little, May Run Into Trouble Later On," *The Wall Street Journal* (5 June 1995): 1; details come from Merrill Lynch press release, "New Data Shows Wealth of American Families at 'Woeful Low,' " (21 December 1994).

● Chapter Concepts Test

Circle the correct answer or fill in the blank. *Answers are in the back of the book.*

T F 1. The value of every observation in the data set is taken into account when we calculate its median.

T F 2. When the population is either negatively or positively skewed, it is often preferable to use the median as the best measure of location because it always lies between the mean and the mode.

T F 3. Measures of central tendency in a data set refer to the extent to which the observations are scattered.

T F 4. A measure of the peakedness of a distribution curve is its skewness.

T F 5. With ungrouped data, the mode is most frequently used as the measure of central tendency.

T F 6. If we arrange the observations in a data set from highest to lowest, the data point lying in the middle is the median of the data set.

T F 7. When working with grouped data, we may compute an approximate mean by assuming that each value in a given class is equal to its midpoint.

T F 8. The value most often repeated in a data set is called the arithmetic mean.

T F 9. If the curve of a certain distribution tails off toward the left end of the measuring scale on the horizontal axis, the distribution is said to be negatively skewed.

T F 10. After grouping a set of data into a number of classes, we may identify the median class as being the one that has the largest number of observations.

T F 11. A mean calculated from grouped data always gives a good estimate of the true value, although it is seldom exact.

T F 12. We can compute a mean for any data set once we are given its frequency distribution.

T F 13. The mode is always found at the highest point of a graph of a data distribution.

T F 14. The number of elements in a population is denoted by n.

T F 15. For a data array with 50 observations, the median will be the value of the 25th observation in the array.

T F 16. Extreme values in a data set have a strong effect on the median.

T F 17. The difference between the largest and smallest observations in a data set is called the geometric mean.

T F 18. The dispersion of a data set gives insight into the reliability of the measure of central tendency

T F 19. The standard deviation is equal to the square root of the variance.

T F 20. The difference between the highest and lowest observations in a data set is called the quartile range.

T F 21. The interquartile range is based on only two values taken from the data set.

T F 22. The standard deviation is measured in the same units as the observations in the data set.

T F 23. A fractile is a location in a frequency distribution that a given proportion (or fraction) of the data lies at or above.

T F 24. The variance, like the standard deviation, takes into account every observation in the data set.

T F 25. The coefficient of variation is an absolute measure of dispersion.

T F 26. The measure of dispersion most often used by statisticians is the standard deviation.

T F 27. One of the advantages of dispersion measures is that any statistic that measures absolute variation also measures relative variation.

T F 28. One disadvantage of using the range to measure dispersion is that it ignores the nature of the variations among most of the observations.

T F 29. The variance indicates the average distance of any observation in the data set from the mean.

T F 30. Every population has a variance, which is signified by s^2.

T F 31. According to Chebyshev's theorem, no more than 11 percent of the observations in a population can have population standard scores greater than 3 or less than -3.

T F 32. The interquartile range is a specific example of an interfractile range.

T F 33. It is possible to measure the range of an open-ended distribution.

T F 34. The interquartile range measures the average range of the lower fourth of a distribution.

A B C D 35. When calculating the average rate of debt expansion for a company, the correct mean to use is the

(a) Arithmetic mean.
(b) Weighted mean.
(c) Geometric mean.
(d) Either (a) or (c).

A B C D 36. The mode has all of the following disadvantages except

(a) A data set may have no modal value.
(b) Every value in a data set may be a mode.
(c) A multimodal data set is difficult to analyze.
(d) The mode is unduly affected by extreme values.

A B C D 37. What is the major assumption we make when computing a mean from grouped data?

(a) All values are discrete.
(b) Every value in a class is equal to the midpoint.
(c) No value occurs more than once.
(d) Each class contains exactly the same number of values.

A B C D 38. Which of the following statements is NOT correct?

(a) Some data sets do not have means.
(b) Calculation of a mean is affected by extreme data values.
(c) A weighted mean should be used when it is necessary to take the importance of each value into account.
(d) All these statements are correct.

A B C D 39. Which of the following is the first step in calculating the median of a data set?

(a) Average the middle two values of the data set.
(b) Array the data.
(c) Determine the relative weights of the data values in terms of importance.
(d) None of these.

A B C D E 40. Which of the following is NOT an advantage of using a median?

(a) Extreme values affect the median less strongly than they do the mean.
(b) A median can be calculated for qualitative descriptions.
(c) The median can be calculated for every set of data, even for all sets containing open-ended classes.
(d) The median is easy to understand.
(e) All these are advantages of using a median.

A B C D 41. Why is it usually better to calculate a mode from grouped, rather than ungrouped, data?

(a) The ungrouped data tend to be bimodal.

(b) The mode for the grouped data will be the same, regardless of the skewness of the distribution.

(c) Extreme values have less effect on grouped data.

(d) The chance of an unrepresentative value being chosen as the mode is reduced.

A B C D 42. In which of these cases would the mode be most useful as an indicator of central tendency?

(a) Every value in a data set occurs exactly once.

(b) All but three values in a data set occur once; three values occur 100 times each.

(c) All values in a data set occur 100 times each.

(d) Every observation in a data set has the same value.

A B C D E 43. Which of the following is an example of a parameter?

(a) $\bar{x}$.

(b) n.

(c) μ.

(d) All of these.

(e) (b) and (c), but not (a).

A B C D E 44. Which of the following is NOT a measure of central tendency?

(a) Geometric mean.

(b) Median.

(c) Mode.

(d) Arithmetic mean.

(e) All these are measures of central tendency.

A B C D E F 45. When a distribution is symmetrical and has one mode, the highest point on the curve is called the

(a) Range.

(b) Mode.

(c) Median.

(d) Mean.

(e) All of these.

(f) (b), (c), and (d), but not (a).

A B C D E 46. When referring to a curve that tails off to the left end, you would call it

(a) Symmetrical.

(b) Skewed right.

(c) Positively skewed.

(d) All of these.

(e) None of these.

A B C D 47. Disadvantages of using the range as a measure of dispersion include all of the following except

(a) It is heavily influenced by extreme values.

(b) It can change drastically from one sample to the next.

(c) It is difficult to calculate.

(d) It is determined by only two points in the data set.

A B C D 48. Why is it necessary to square the differences from the mean when computing the population variance?
(a) So that extreme values will not affect the calculation.
(b) Because it is possible that N could be very small.
(c) Some of the differences will be positive and some will be negative.
(d) None of these.

A B C D 49. Assume that a population has $\mu = 100$ and $\sigma = 10$. If a particular observation has a standard score of 1, it can be concluded that
(a) Its value is 110.
(b) It lies between 90 and 110, but its exact value cannot be determined.
(c) Its value is greater than 110.
(d) Nothing can be determined without knowing N.

A B C D E 50. Assume that a population has $\mu = 100$, $\sigma = 10$, and $N = 1{,}000$. According to Chebyshev's theorem, which of the following situations is NOT possible?
(a) 150 values are greater than 130.
(b) 930 values lie between 100 and 108.
(c) 22 values lie between 120 and 125.
(d) 70 values are less than 90.
(e) All these situations are possible.

A B C D E 51. Which of the following is an example of a relative measure of dispersion?
(a) Standard deviation.
(b) Variance.
(c) Coefficient of variation.
(d) All of these.
(e) (a) and (b), but not (c).

A B C D 52. Which of the following is true?
(a) The variance can be calculated for grouped or ungrouped data.
(b) The standard deviation can be calculated for grouped or ungrouped data.
(c) The standard deviation can be calculated for grouped or ungrouped data, but the variance can be calculated only for ungrouped data.
(d) (a) and (b), but not (c).

A B C D E 53. If one were to divide the standard deviation of a population by the mean of the same population and multiply this value by 100, one would have calculated the
(a) Population standard score.
(b) Population variance.
(c) Population standard deviation.
(d) Population coefficient of variation.
(e) None of these.

A B C D E 54. How does the computation of a sample variance differ from the computation of a population variance?
(a) μ is replaced by $\bar{x}$.
(b) N is replaced by $n - 1$.
(c) N is replaced by n.
(d) (a) and (c), but not (b).
(e) (a) and (b), but not (c).

A B C D E F 55. The square of the variance of a distribution is the
(a) Standard deviation.
(b) Mean.
(c) Range.
(c) Absolute deviation.
(e) (a) and (d).
(f) None of these.

A B C D E 56. Chebyshev's theorem says that 99 percent of the values will lie within ± 3 standard deviations from the mean for
(a) Bell-shaped distributions.
(b) Positively skewed distributions.
(c) Left-tailed distributions.
(d) All distributions.
(e) No distributions.

57. If a curve can be divided into two equal parts that are mirror images, it is _____. If it cannot be divided in this way, it is _____.

58. The symbol $\bar{x}$ denotes the mean of a _____. μ denotes the mean of a _____.

59. Assigning small-value consecutive integers to midpoints during calculation of the mean is called _____.

60. When dealing with quantities that change over a period of time, it is better to calculate a _____ mean than a _____ mean.

61. If two values in a group of data occur more often than any others, the distribution of the data is said to be _____.

62. The extent to which values in a distribution are grouped together is a measure of _____.

63. In a frequency distribution the median is the 0.5 _____ because half of the data values are less than or equal to this value.

64. The difference between the values of the first and third quartiles is the _____ range.

65. The measure of the average squared distance between the mean and each item in the population is the _____. The positive square root of this value is the _____.

66. The expression of the standard deviation as a percentage of the mean is the _____.

67. The number of standard deviation units that an observation lies above or below the mean is called the _____

68. Fractiles that divide the data into 100 equal parts are called _____.