Statistical Methodologies

FREQUENCY DISTRIBUTION

(1)

1. Such things as test scores, class rank, weight, and income, are called variables. Income, for instance, is called a variable because different income values are possible. In general, things that vary in value from case to case or time to time are called ____________.

VARIABLES

2. The number of times a particular value of a variable occurs is referred to as the *frequency* of that value. If 17 students receive a score of 70 on a test, then the score of 70 has a ____________ of 17.

FREQUENCY

3. A distribution is a series of separate values such as scores which are ordered to magnitude. A group of ordered scores can make up a *distribution*. For example, a group of scores ranging from the lowest to the highest score is a ____________ (see table).

Scores
13
11
11
9
9
9
8
5

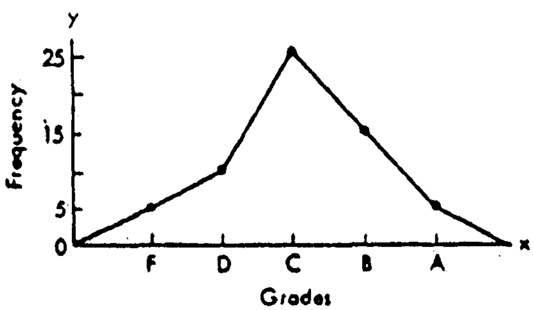
DISTRIBUTION

4. A set of ordered scores and their corresponding frequencies is called a *frequency distribution*. This can be represented in table or graph form. The table below shows the number of times a score occurs in its group. This table is a frequency ____________.

Scores	Frequency
13	I
11	II
9	III
8	I
5	I

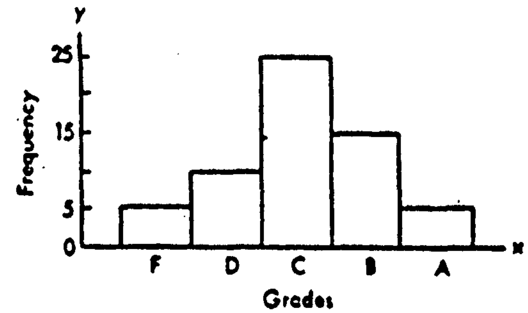
DISTRIBUTION

5. Frequency distributions can also be graphically illustrated. The two most common graphs used to illustrate frequency distributions are the *frequency polygon* and the *histogram*. If scores and their frequencies are illustrated with points connected by lines, it is called a *frequency polygon*. Because the illustration below shows the frequency of particular scores by the height of points that are connected by lines, it is called a frequency ____________



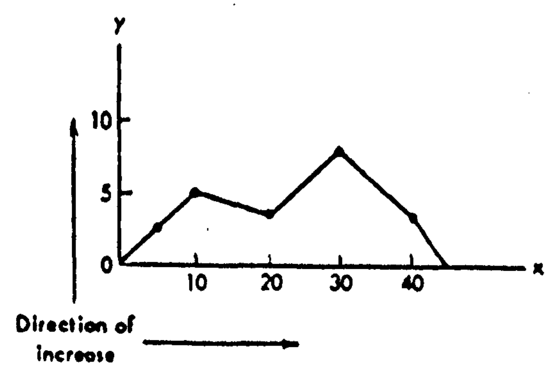
POLYGON

6. In a histogram frequency distribution, the scores and their frequencies are designated by the use of rectangular boxes. In the frequency distribution below, the height of the rectangular boxes indicates the frequency of students that received that particular score. It is called a ____________.



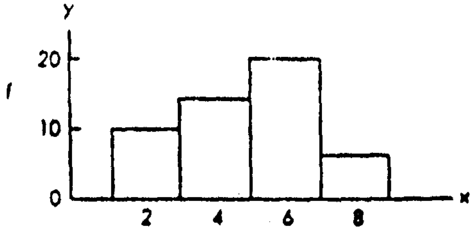
HISTOGRAM

7. It is the accepted practice for the vertical side of a graph, called the *ordinate* axis, to be used to designate the frequency. The horizontal side, called the *abscissa* axis, is used for the scores. Direction of increase is upward for the frequency on the ordinate axis. Direction of increase for the variable is from left to right on the ____________ axis.



ABSCISSA

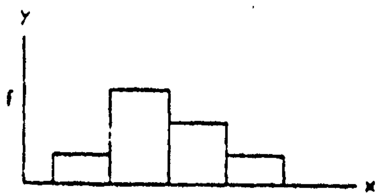
8. On this graph the *f*, which designates the frequency, is the ________ axis, and the *x*, designating the variable, is the ________ axis.



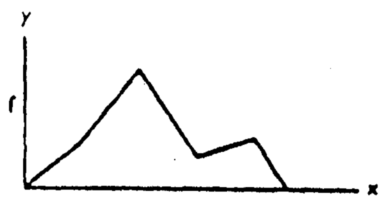
ORDINATE

ABSCISSA

9. The two most common graphs used to illustrate frequency distributions are the frequency polygon and the ________ ________. Graph A. is a ________ Graph B. is a ________ ________ ________



Graph A



Graph B

HISTOGRAM

HISTOGRAM

FREQUENCY POLYGON

AVERAGES

10. After scores have been tabulated into a frequency distribution, a measure of *central tendency*, or central position is often calculated. Central tendency gives us a concise description of the average or typical performance of the group as a whole. Measures of ________ tendency allow us to compare two or more groups in terms of typical performance.

CENTRAL

11. In statistics there are several "averages" or measures of ________ ________ in common use. Three of these are (a) the mean, (b) the median, and (c) the mode.

CENTRAL TENDENCY

12. The mean is generally the most familiar and most useful to us. The mean is computed by dividing the *sum of the scores* by the *total number of scores*. The formula for the mean would be

$$\text{Mean} = \frac{\text{sum of the scores}}{?}.$$

TOTAL NUMBER OF SCORES

13. Instead of stating that the mean is the sum of the scores divided by the total number of scores, it is easier to use the following symbols:
   a. Mean = $\bar{X}$ read "X bar") or *M*. (The symbols $\bar{X}$ or *M* are used when referring to the mean of a sample from the total population.)

b. Sum of the scores = ΣX (Σ = sum; X = each score).
c. Total number of scores = N.
Thus the formula for the mean would be $\bar{X} = ?/?$

$\Sigma X/N$

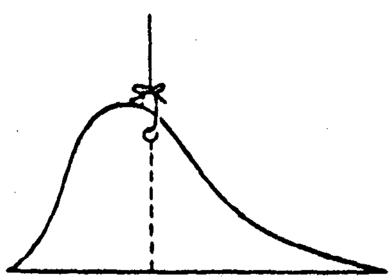
14. Compute the mean ($\bar{X} = \Sigma X/N$) from the given information:

Scores (X):	7
	3
	6
	4
Sum of scores (ΣX):	20
Number of scores (N):	4

$\bar{X} = ?/? = ?$

$20/4 = 5$

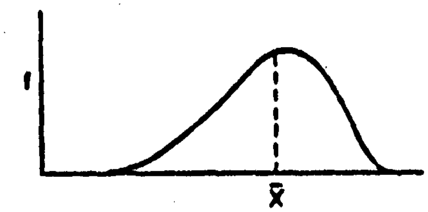
15. Finding the arithmetic mean of a distribution is analogous to finding the center of moment, or the balance point, in a solid block. If a distribution were suspended by the mean it would hang level or balanced. The mean, whose symbol is ________, is the center of moment in a frequency distribution.



Mean

$\bar{X}$ or M

16. Thus, if extremely high or extremely low scores are added to a distribution, the mean tends to shift toward those scores. If the center of balance of the distribution is shifted to one side or the other of the curve, the curve becomes "skewed." The following curve has a few extremely low scores. Consequently, this distribution is ________



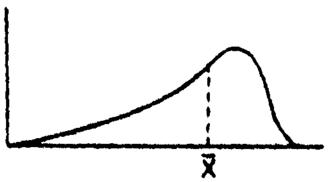
SKEWED

17. Extreme scores, either high or low, tend to ________ a distribution.

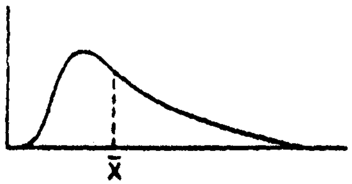
SKEW

18. If a distribution is massed so that the greatest number of scores is at the right end of the curve and a few scores are scattered at the left end, the curve is said to be *negatively* skewed. If the massing of scores is at the left end of the curve with the tail extending to the right end, then the curve is *positively* skewed.

Graph A illustrates ________________ skewness. Graph B illustrates ________________ skewness.



Graph A

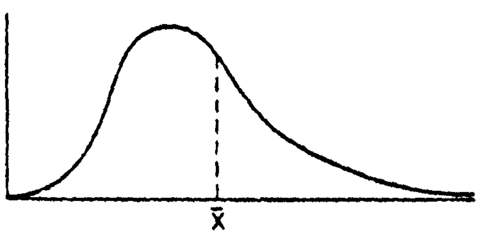


Graph B

NEGATIVE

POSITIVE

9. This graph's tail is extending to the right because of a few extremely high scores, therefore, it is ________________ skewed.



POSITIVELY

A curve is *symmetrical* when one half of the curve is a reproduction of the other half. If you folded a frequency polygon at the mean and the two halves were similar, then the frequency distribution represented by the polygon would be said to be ________________

SYMMETRICAL

21. According to the formula for computing the mean ($\bar{X} = \Sigma X/N$), we can define the mean as the arithmetic average of the scores in a distribution. If we added extreme scores to one end of a previously symmetrical curve, the mean would shift towards those extreme scores. Would the curve be symmetrical or not symmetrical? ________________

NOT SYMMETRICAL (OR ASYMMETRICAL)

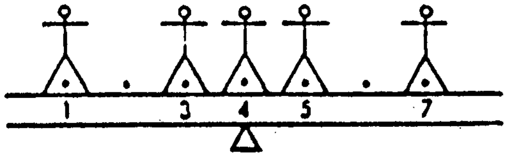
22. Regardless of whether the curve is symmetrical or asymmetrical, the mean is always the center of balance. Does this imply that the mean is centrally located in asymmetrical curves? ________________

NO

23. Let us illustrate this point by placing a distribution along an interval scale such as that below. Each figure represents one person. The scale would obviously balance if a fulcrum were

under the middle number, 4. Calculate the mean by the formula $\bar{X} = \Sigma X/N$ to verify this. Was this distribution symmetrical?

____

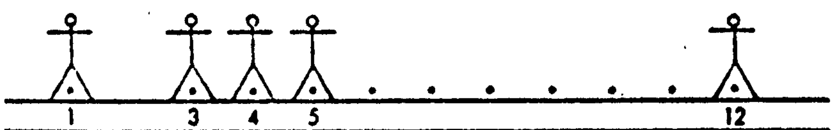


$\bar{X} = 20/5 = 4$

YES

1. If the person with a score of 7 had gotten 12, what would be the mean? ______________ Place a fulcrum (i.e., Δ) at the balance point of the scale below. Is the fulcrum centrally located? ______________ Is this distribution symmetrical?

____



5

FULCRUM SHOULD BE UNDER NUMBER 5

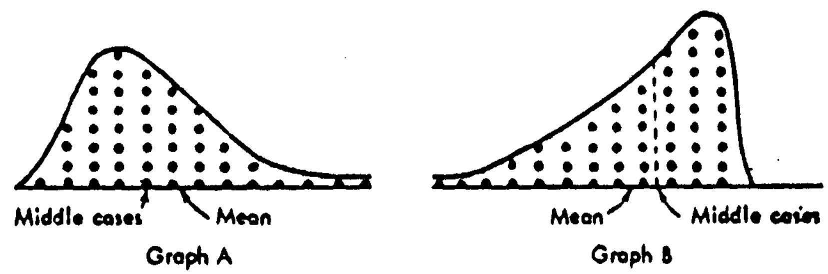
NO

NO

25. What would be the mean for the above distribution if the person who scored 12 had instead scored 22? ______________

7

26. When a curve is positively skewed (see graph A) the mean is located to the ______________ (right or left) of most of the cases. When a curve is negatively skewed (see graph B) the mean is located to the ______________ of most of the cases. (Each dot is one case.)



RIGHT

LEFT

27. If the measures 2, 2, 3, 3, 15 were a frequency distribution, the curve would be ______________ skewed.

POSITIVELY

28. The preceding distribution of 2, 2, 3, 3, 15 has a mean of 5. What will the mean be if 10 points are added to the score of 15 (making it a score of 25)? ______________

7

29. We saw that by adding 10 points to the score of 15, the mean of the distribution 2, 2, 3, 3, 15 was raised by 2 points. The reason for this is that the mean is an arithmetic average and each score contributes to its value. When 10 was added it was averaged or distributed equally among the five scores. This has the same effect as adding a constant of 2 points to each score (10 points/5 scores = 2 points per score). When 2 points are added to each score, the mean is raised by ______________ points (from 5 to 7).

2

30. When a constant is added to each score of a distribution, that constant is added to the previous mean to find the new mean. If each score of a distribution is *multiplied* by a constant, the new mean is found by multiplying the old mean by that ______________

CONSTANT

31. The distribution 0, 2, 2, 3, 13 has a mean of 4. What would the mean be if each score was *multiplied* by a constant of 2? ______________

8

MEDIAN

32. By adding or substituting an extreme score to a distribution, the mean no longer represents a centrally located score but represents a measure that is more typical of the extreme score.

This causes us to rely on another measure of central tendency which is called the median or the middle score. The median is abbreviated Md or Mdn. The measure of central tendency that is less affected by the addition of an extreme score is the ______________

MEDIAN

33. The median is a *point* on a scale of measurement above which are exactly half the cases and below which are the other half of the cases. The student should note that the median is defined as a *point* and not as a specific measurement, e.g., a score or a case. From the distribution 4, 6, 8, 10, 12, it is easy to see that 8 is the middle score. The score of 8 is at the *point* where there are two scores above and two scores below, hence, 8 is the median. What is the median of 11, 11, 14, 19, 19? ______________

14

34. To obtain the median, the measures are arranged in ascending order from the lowest to the highest measure. Then by count-

ing up this scale, the point is selected where there are an equal number of cases above this point and an equal number of cases below this point. The value of this point is the middle or the ________ case.

MEDIAN

35. The median of the distribution 3, 2, 0, 1, 6 can be found by first arranging the measures from the lowest to the ________ ________ number (0, 1, 2, 3, 6). Then we find the middle score or case, which is ________, and that is the median.

HIGHEST

2

36. It is not too difficult to determine the median of a distribution with an odd number of cases—if providing the score at the midpoint has a frequency of 1. In the distribution 3, 5, 7, 9, 11, 13, 13, the median score, which is ________, has a frequency of 1. The score of 13 has a frequency of ________ ________.

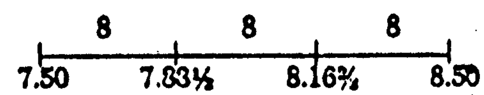
9

2

37. A distribution whose midpoint score has a frequency greater than 1 (e.g., 5, 6, 9, 9, 9, 11) presents a special problem. To overcome this, the student needs to know what is meant by "the interval of a score." For our purposes, the interval of a score ranges from .5 unit below a given score up to .5 unit above a given score. For example, the score of 9 includes all values within the limits of 8.5 up to 9.5. The exact midpoint of the interval whose lower and upper limits are 8.5 and 9.5, respectively, is 9. The score of 17 would represent the interval from 16.5 up to ________

17.5

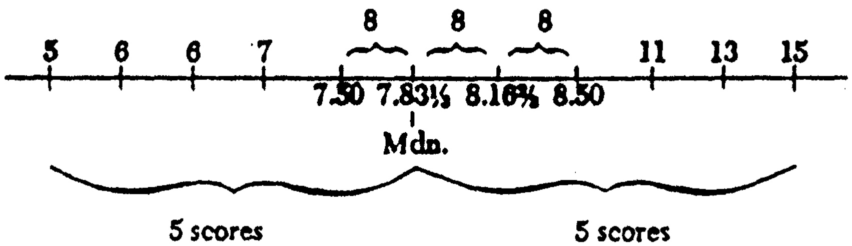
38. In the distribution of 5, 6, 6, 7, 8, 8, 8, 11, 13, 15, the score of 8 has a frequency of ________ The score of 8 has an interval range from 7.5 up to ________ It is assumed that the unit of three scores (8, 8, 8) is spread equally through the interval of 7.5 to 8.5. Each 8 occupies $\frac{1}{3}$ ($0.33\frac{1}{3}$) of a unit. For example:



3

8.5

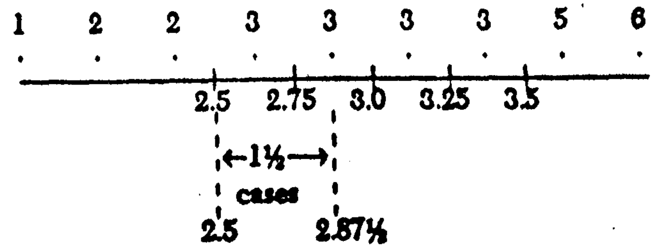
39. The midpoint of the distribution 5, 6, 6, 7, 8, 8, 8, 11, 13, 15, where half the scores are on one side and half the scores are on the other, is between the fifth and ________________ score. Below the interval 7.5 to 8.5 there are four scores, consequently the fifth score extends ⅓ of the way into the three score unit. Thus, the point between the fifth score and the sixth, which is the ________________, is found by adding ⅓ of the unit to 7.5 (7.5 + .33⅓ = 7.83⅓). Note the illustration:



SIXTH

MIDPOINT OR MEDIAN

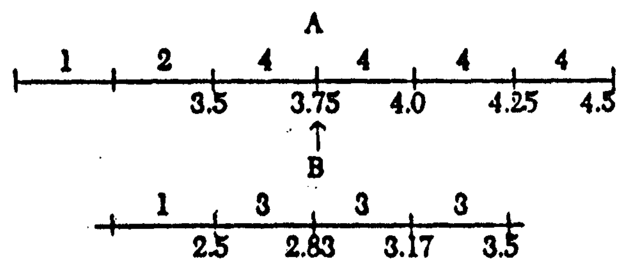
40. The distribution 1, 2, 2, 3, 3, 3, 3, 5, 6 has nine cases. The median of these nine cases is a point which has 4.5 cases below it and ________________ cases above it. This midpoint falls within a four-case unit. Below the lower limit of 3 (which is 2.5) there are three cases; therefore by extending one and one-half cases into the interval of 2.5 to 3.5 we can locate the median. Each 3 accounts for one-fourth of the four-case unit, hence one and one-half cases is equal to ¼ plus ⅛ of a unit (0.25 + 0.12½). The value of the median is 2.5 + 0.37½ = ________________. It is illustrated as follows:



4.5

2.87½

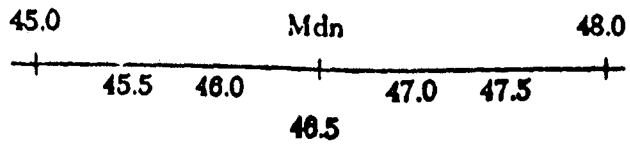
41. The same principle applies to distributions with even numbers of cases except that the median falls midway between the two middle cases. For example in distribution A, shown below, the arrow indicates the median. Draw an arrow to indicate the median of distribution B.



2.83

42. The distribution 40, 43, 44, 45, 48, 53, 56, 60 has eight cases. The median is a point midway between the fourth and ________

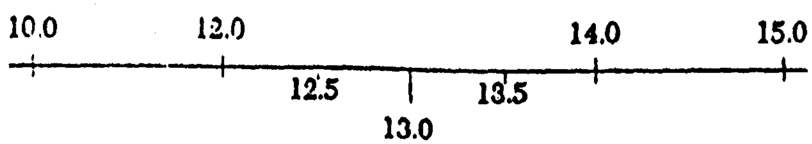
——— scores. The upper limit of 45 (the fourth score) is 45.5 and the lower limit of 48 (the fifth score) is ——————— and midway between these two limits is the point 46.5, which is the median (as noted in the illustration below)..



FIFTH

47.5

43. The distribution 10, 12, 14, 15 has four cases. The median is a point midway between the second and ——————— case. The upper limit of 12 is ———————, and the lower limit of 14 is ——————— and midway between these two limits is the point ———————, which is the median. Note the illustration:



THIRD

12.5

13.5

13

44. The values of the median as a method for obtaining the most typical measure of central tendency is a skewed distribution becomes even greater the more extreme the end score is. For example, the median of 0, 6, 9, 10, 10 is ——————— The mean of this distribution is 7 (i.e., 35/5). If one of the 10's is substituted by the extreme number 55 (the new distribution would be 0, 6, 9, 10, 55), the median remains ——————— but the mean is now ———————

9

9

16

45. The mean and median will *both* be affected if the score of 55 is *added* to the distribution 0, 6, 9, 10, 10. The new distribution would be 0, 6, 9, 10, 10, 55. The mean is 90/6 = 15. The addition of the extreme score shifted the value of the mean 8 points to the right. How far to the right was the median shifted? ——————— Which measure of central tendency was affected the least? ———————

.5 OR ONE-HALF OF A POINT

MEDIAN

46. When we want to minimize the effect of one or more extreme scores, we should use the ______________ to represent the average score of the distribution.

MEDIAN

47. The median, for both odd and even number of cases, is the point on a distribution where there are an equal number of cases above and ______________ that point.

BELOW

MODE

48. A third measure of central tendency is the mode. It may be defined as the one value or score which occurs with the most frequency. The mode of the series 2, 3, 4, 4, 4, 5, 5 is 4. The mode of the series 7, 8, 10, 10, 10, 11, 11 is ______________. The median is ______________

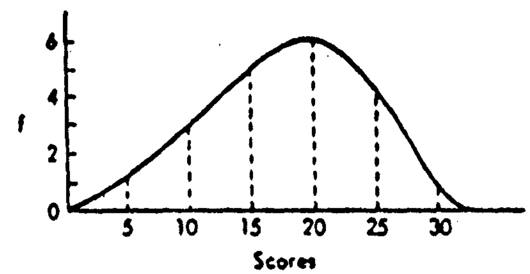
10

10

49. Is it possible for a distribution to have a median and a mode of the same value? ______________ (yes or no)

YES

50. The mode is used as a simple, inspectional "average" to show, in a hurry, the center of concentration of a frequency distribution. What is the mode or the rough average of the frequency polygon shown below? ______________



20

51. The mode is not generally used unless there are a large number of cases in a distribution. When the number of cases in a distribution is small, there is a good possibility of several scores having the same frequency. The frequency polygon shown below is an extreme example. It is evident that the mode is 10 but it does not give a close approximation of the average case. The mean is 25 ($\Sigma X/N = 125/5$). The cases, in ascending order, are

10, 10, 25, 35, 45, with the number 25 at the midpoint; thus ______________ is the median.

25

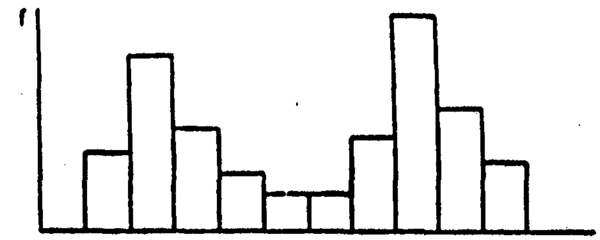
52. The mode is used, in preference to either the median or the mean, when a measure of the most characteristic value of a group is desired. What is meant by "the most characteristic value" can be exemplified by clothing fashions. The ______________ is what is being worn the most.

MODE

53. The mode is used also to be sure that the average you obtain exists in actuality. In finding the average size of automobile tire that is purchased, the mean or median size might be a tire that doesn't exist. Therefore, one would want to know the size of tire bought most often. This would be the ______________.

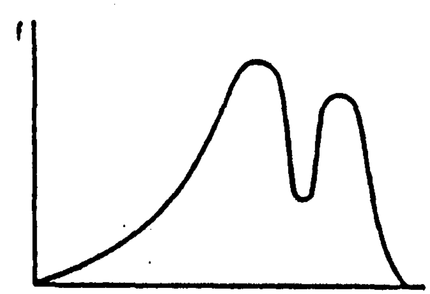
MODE

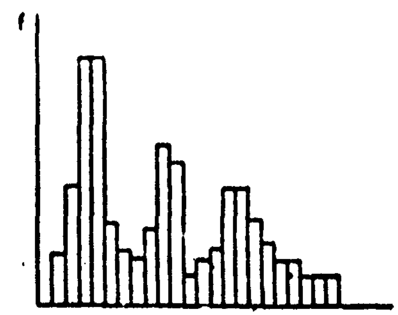
54. In addition to serving as a measure of central tendency, the concept of modality is useful in describing the shape of some distributions. If a histogram or a frequency distribution has two peaks, it is referred to as a *bimodal* distribution. If a distribution has more than two peaks, it is called *multimodal*. The following histogram appears to have two separate concentrations of frequencies; consequently it is called ______________.



BIMODAL

55. The distribution of the frequency polygon illustrated below is ______________. The distribution of the histogram is ______________. The frequency pologon is negatively skewed, whereas the histogram is ______________ skewed.



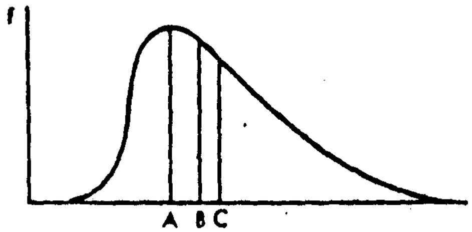


BIMODAL

MULTIMODAL

POSITIVELY

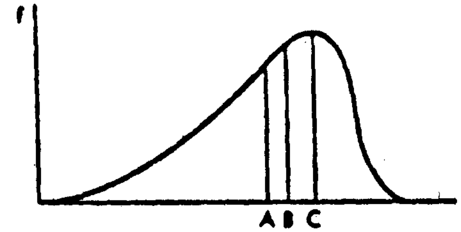
56. The score which occurs with the most frequency is the mode; hence the mode is totally uninfluenced by extreme scores. The mean is greatly influenced by extreme scores. On the basis of these two statements and the preceding exercises on the median, it is evident that line A is the mode (it is totally uninfluenced by the extreme scores). Line B is not affected as much as line C, thus it must be the ________________. Line C is the ________ ________; it was influenced the most by the extreme scores.



MEDIAN

MEAN

57. The frequency distribution below is ________________ skewed. Line A is the ________________. Line B is the ________ ________. Line C is the ________________. The mean of a negatively skewed distribution is located left of the center. The mean of a positively skewed distribution is located ________ ________ of the center.



NEGATIVELY

MEAN

MEDIAN

MODE

RIGHT

58. What are the mean, median, and mode of the distribution

1 2 2 4 6 ? ________________

3

2.25

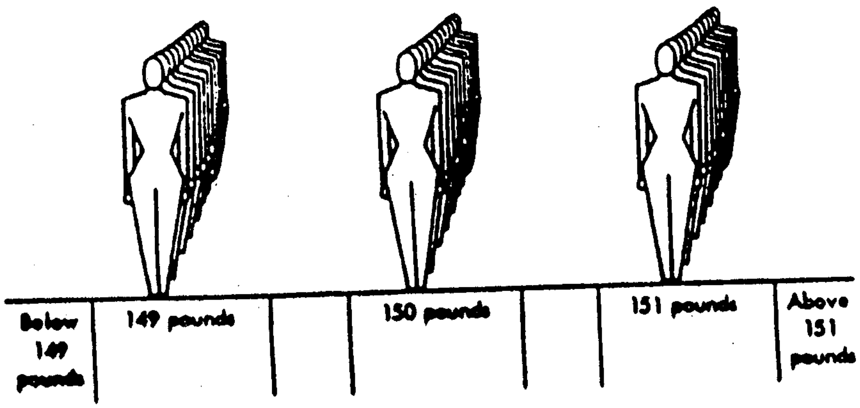
2

59. The mean, median, and mode have been discussed as averages. It should now be evident why these three statistical tools are called measures of ________________

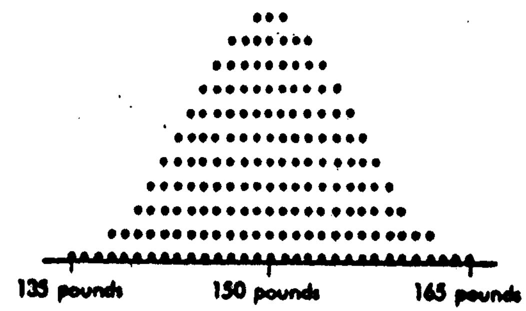
CENTRAL TENDENCY

THE NORMAL CURVE

60. Let us suppose that each of 260 students lined up in front of a sign that gave his weight. The sign running from left to right in order of increasing weight are from 135 to 165 pounds. The numbers of persons in any one line is the frequency of that weight. The number of 150 pounders is the ________________ of 150 pounders.

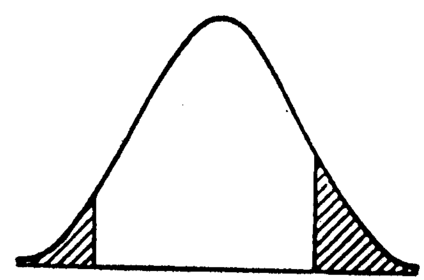


61. From an airplane, the place where this odd event was occurring might look like the diagram below. Each dot represents a ________________



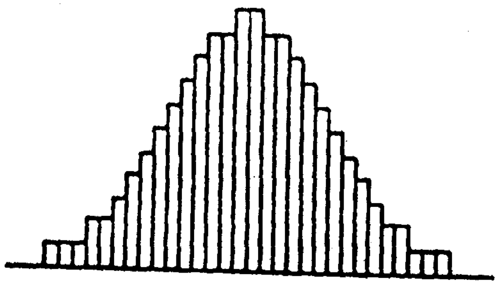
PERSON OR STUDENT

62. Assuming that the students are separated from one another by the same amount of space, the *number of cases would be indicated by the area.* For example with 260 cases, the 26 heaviest students would occupy the extreme right 10% of the crowd. The 13 lightest people would occupy the extreme left ________ ______% of the crowd.



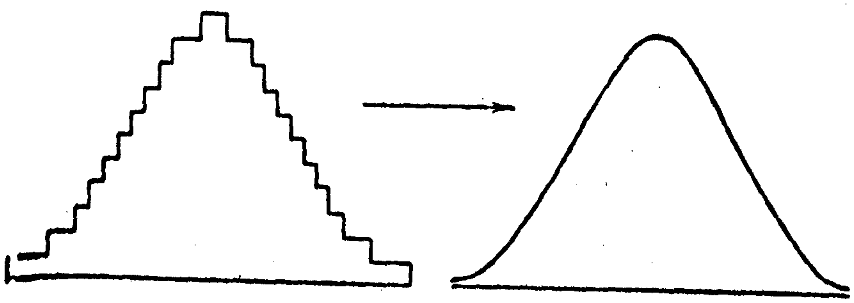
5%

63. If each column of students is represented by a rectangular box, we have our old friend the ____________



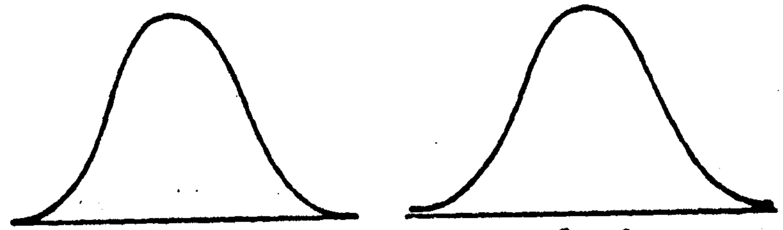
HISTOGRAM

64. If we have a very large number of people and use very small weight categories, the irregular step-like curve would become smooth and continuous. The resulting figure approaches a special type of curve called the normal curve. In frequency distributions normality is not associated with small groups of people but rather with very ____________ groups of people.



LARGE

65. In a normal curve (which by definition describes an infinite number of cases) the tails of the curve never touch the baseline. Which curve below could be a true normal curve? ____________



Curve A   Curve B

B

66. It has been found that quantitative data gathered from high-frequency random measurements of natural phenomena and of many mental and social traits, even though not precisely normal in distribution, can be closely described by the normal ____________

CURVE

67. The distributions of such diverse properties as achievement test scores, I. Q., and height and weight of people form approximately ____________ ____________

NORMAL CURVES

68. The end points of a normal curve remain open and recede indefinitely so as never to touch the abscissa or base line because the number of cases has to be ____________

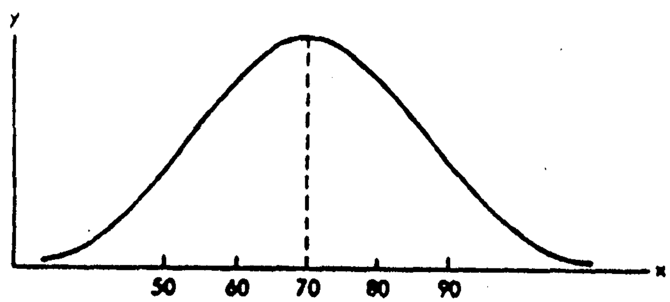
INFINITE

69. When a line approaches infinitely close to another line but does not touch that line, the lines are said to be asymptotic. The end points of a normal curve are ______________ to the base line.

ASYMPTOTIC

70. The bell-shaped curve illustrated below approximates what the statisticians call a normal curve. Note the following properties:

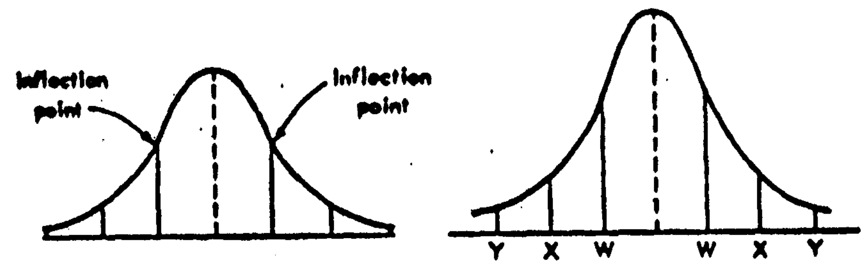
a. It is symmetrical.
b. The mean, median, and mode have the same value (in this instance, 70).
c. There are thus an equal number of scores on either side of the mean (central axis).
d. It is composed of infinitely large numbers of ______________.
e. The end points of the curve are ______________ to the abscissa (base line).



CASES

ASYMPTOTIC

71. Another identifying characteristic of the normal curve is its mathematical construction. There are two points on the normal curve where the curve changes direction from convex to concave. These points are points of inflection (see graph A). Are the inflection points on graph B at line W, line X, or line Y? ______________

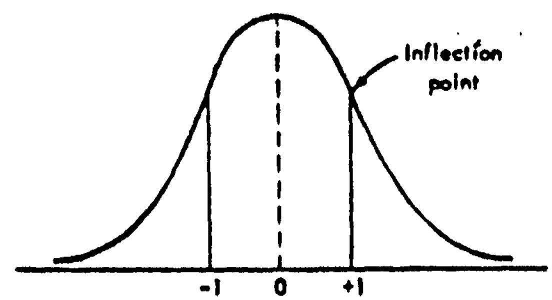


Graph A

Graph B

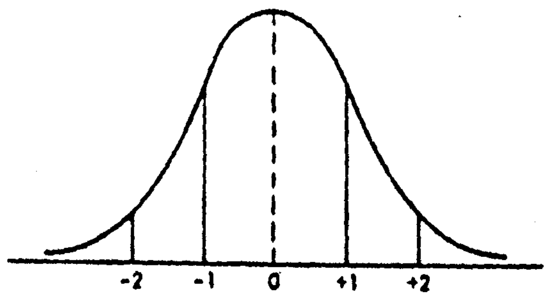
LINE W

72. By the methods of calculus it can be shown that a line drawn perpendicular from the point of inflection to the abscissa is *one unit* of distance or deviation from the central axis. If one uses this distance as a *standard*, a uniform method of dividing the base line into equal segments can be established. If the central axis is designated as zero, the line one unit of distance to the right would be plus and the line one unit of distance to the left would be ______________



MINUS

73. Mathematically, the lines −1 and +1 are situated one unit of distance or deviation from the central axis or values (the mean, median, and mode). These two lines are designated as ±1 (read as plus and minus one). Two units of distance or deviation from the central axis are labeled as +2 and ______________



oooooooooooooooooooooooooooooooo

−2

74. Using the unit of distance established by constructing a perpendicular line from the point of inflection to the abscissa as a standard, we can divide the base line into several equal segments. Since the normal curve is asymptotic with respect to the abscissa, one could divide the base line into equal parts indefinitely. All segments would be a uniform or standard distance. The unit of distance was established by constructing a perpendicular line from the point of ______________ to the abscissa.

oooooooooooooooooooooooooooooooo

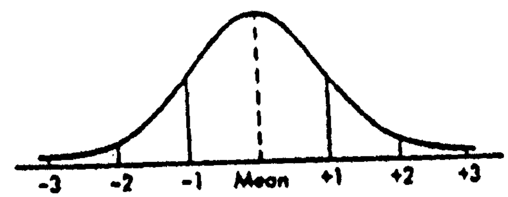
INFLECTION

75. The proportion of cases beyond ±3 units from the center of the normal curve is so small that they are generally ignored. It is thus common practice to illustrate only those cases contained between the arbitrary limits of +3 and ______________ units of deviation.

oooooooooooooooooooooooooooooooo

−3

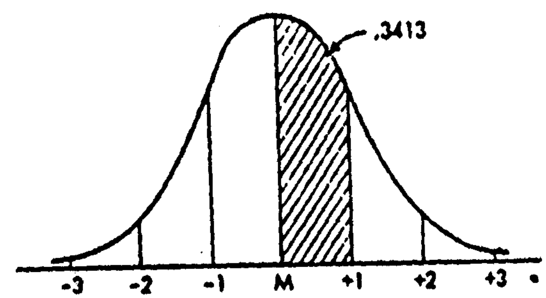
76. Notice that in the graph below each divided segment is equal to the distance from (or the deviation from) the mean to the perpendicular line drawn from the ______________ point.



oooooooooooooooooooooooooooooooo

INFLECTION

77. The total area under the normal curve may be set to equal 1 or unity. Between the mean and +1 unit of deviation above (to the right of) the mean is .3413 (about ⅓) of the total area, i.e., from the mean to +1 unit of deviation lies 34.13% of the total cases. Since −1 unit of deviation is proportionate to +1 unit of deviation, ______________% of the total cases lie between −1 unit of deviation and the mean.

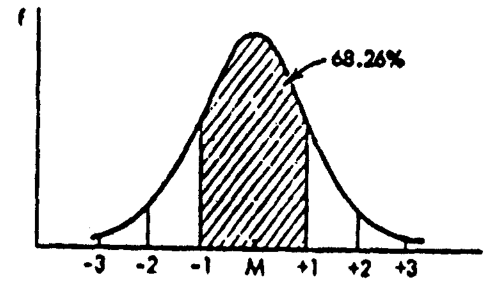


oooooooooooooooooooooooooooooooo

34.13%

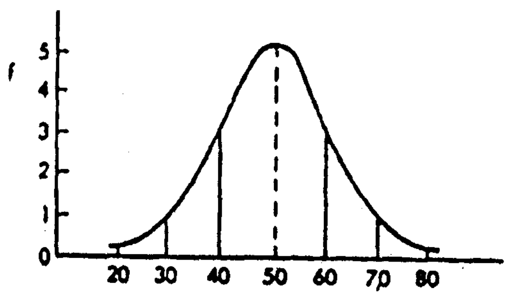
78. Because of the massing of scores around the central values, a little more than ⅔ (2 × 34.13% = 68.26%) of the total frequencies are between +1 and −1 deviations. If a normal distribution has

a total frequency of 1000 scores, approximately 341 scores (34.13% × 1000) are located between the mean and −1 unit of deviation and approximately 341 scores are located between the mean and +1 unit of deviation. How many scores are located between −1 deviation and +1 unit of deviation? ________



683 (682.6)

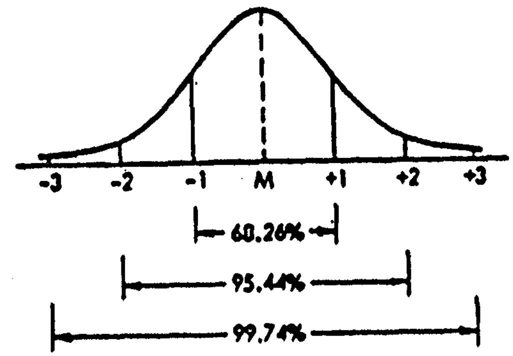
79. In this frequency distribution the deviation points, −1 and +1, mark off the middle ________% of the total scores. They occur at the scores of 40 and ________



68.26% or 68%

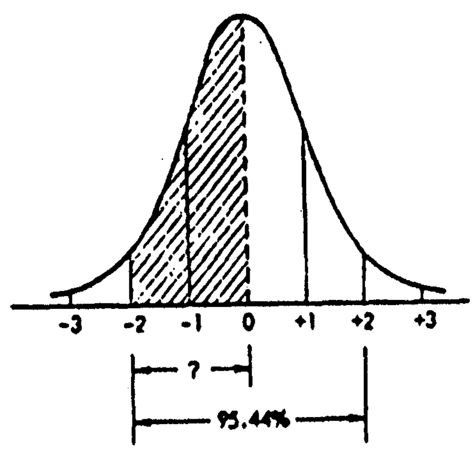
60

80. Although the normal curve extends indefinitely to the left and to the right, the end points of the curve approach the base line so closely that over 95.44% (see graph below) of the area or frequencies are included between the limits −2 and +2 and 99.74% of the cases are included between the limits −________ and + ________



−3 and +3

81. The percentage of cases contained between the mean (central axis) of a normal curve and +3 units of deviation is 49.87% (one-half of 99.74%). The percentage of cases between the central axis (the mean) of a normal curve and −2 units of deviation is ________%



47.72%

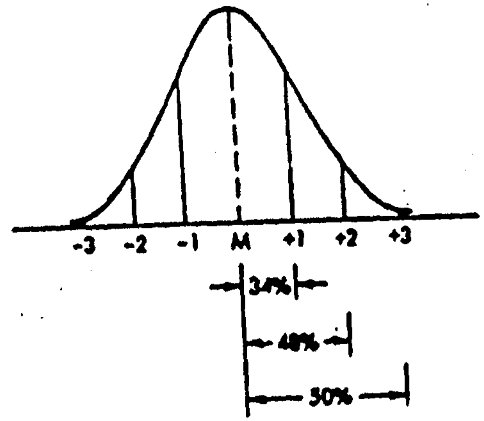
82. As we stated before, for practical purposes the limits of the frequencies of the normal curve rarely exceed those of ±3 units of deviation from the mean. The approximate twenty six hundredths of one percent (.0026) of the total cases existing outside the limits of ±3 are so slight in amount that the unity of the curve is generally assumed to be unaffected. Approximately thirteen hundredths of one percent (.0013) of the total cases extend beyond +3 and approximately ____________ hundredths of one percent of the total cases extend beyond −3.

THIRTEEN

83. Though the percentage of cases is very small and insignificant at a considerable distance from the mean (beyond ±3) the proportion of frequencies approaches zero but never equals zero. The reason is that the normal curve is ____________ with respect to the abscissa or base line.

ASYMPTOTIC

84. Since the proportion of the total cases that exist beyond the limits of ±3 is so slight, it may be plausible to treat data as if 100% of the total cases fall within ±3 deviations. If one makes this assumption, the percentage of cases is rounded off to the nearest whole percent. Thus (note graph below) the percentages of cases from the mean to +1, +2, and +3 are 34%, 48%, and 50%, respectively. The percentages of cases *from the mean* to −1, −2, and −3 deviations are ____________%, ____________%, and ____________%, respectively.



34%

48%

50%

85. The percentage of cases below the mean is ____________%.

50%

86. The percentage of cases between the mean and +1 is ____________% of the total cases.

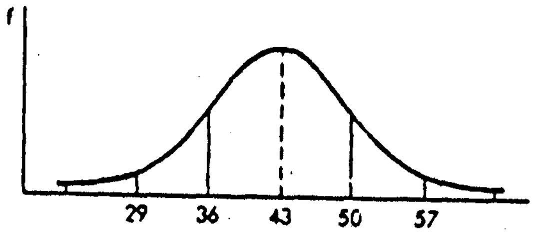
34%

87. The percentage of cases below +1 deviation is 84% (50% plus 34%) of the total cases and the percentage of cases above +1 deviation is ______________% of the total cases. The percentage of cases below −1 deviation is ______________%.

16% (100% − 84%)

16% (50% − 34%)

88. In relation to the scores on the graph below, about what percentage of cases lie below 43? ______________ Between 43 and 57? ______________ Below 57? ______________ Above 57? ______________ Below 29? ______________



50%

48%

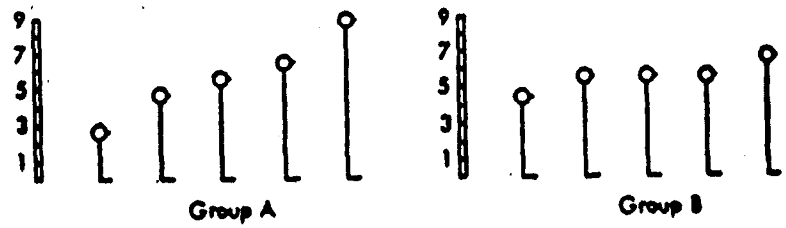
98%

2%

2%

VARIABILITY

89. Descriptions of groups by frequency distributions, central tendency, and normality have been discussed. Another way of describing a group is to have some index of how much variety exists. Consider the height of the two groups of people below. Both groups have a mean and median of 6 feet but the most variable is group ______________



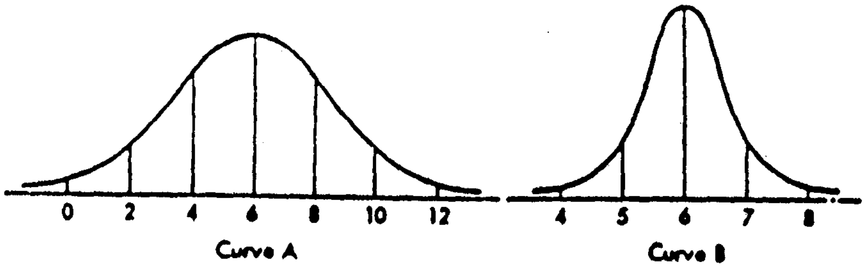
GROUP A

90. One common measure of variability is the range. The range of a set of scores is the distance between the midpoints of the lowest and highest scores. To find the range, subtract the lowest score from the highest score. The range of group A whose heights are 3, 5, 6, 7, 9 is 9 minus 3 or 6. The range for less variable group B whose heights in feet are 5, 6, 6, 6, 7 is ______________

2 (OR 7 MINUS 5)

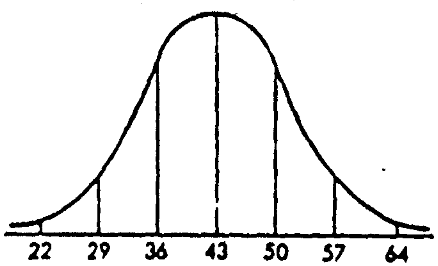
91. If the normal curves below, in which the vertical deviation lines are one standard unit apart, represent large populations,

which curve represents the most variablo group? ______



CURVE A

92. The distance from one deviation line to another is the other major index of variability. It is called a *standard deviation*. In the diagram below, 29 differs from 43 by two ______



STANDARD DEVIATIONS

93. When members of a group deviate very little from each other, the standard deviations are very small. The reverse is true for highly variable groups. Consequently, the variability or diversit of two groups can be compared by the relative size of thei ______ ______.

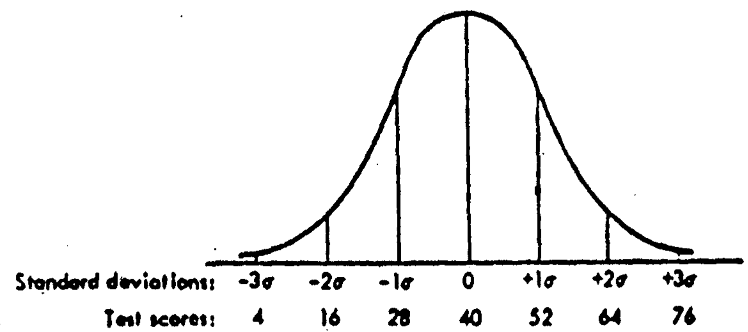
STANDARD DEVIATIONS

94. The capital letters S.D. are used when referring to the standa deviation of a sample of a population. It is common practic to symbolize the standard deviation by the small Greek lette sigma (σ) when referring to the population values. To abbreviat a sample's standard deviation, one could use the capital lette ______. To abbreviate the standard deviation for population, use the small Greek letter ______

S.D.

σ OR SIGMA

95. Suppose that a population of scores is distributed so that th mean is 40 and the distance of 12 points is 1σ (read as on standard deviation). The students one standard deviation abov the mean received the score of ______



52

96. What is the difference, in score points, between the scores of $+1\sigma$ and -2σ of the above illustration? __________

36 (3 × 12)

97. The standard deviation is a kind of average of all the deviations from the mean score. The amount a score (X) deviates from the mean ($\bar{X}$) is symbolized by the small letter x, that is, $X - \bar{X} = x$. Give the symbols for the following:

Raw score __________

Mean __________

Deviation __________

X

$\bar{X}$

x

98. To calculate a standard deviation, the deviations from the mean have to be squared. To square a number, you multiply it by itself. For example, 5 squared, or 5^2, $= 5 \times 5 = 25$; $2^2 =$ _____

4

99. A minus times a minus equals a plus, therefore, $(-4)^2$ or $-4 \times -4 = 16$. Complete the following:

$$(-7)^2 = \_\_\_\_\_\_$$

$$(-3)^2 + (-1)^2 + (-1)^2 + 5^2 = \_\_\_\_\_\_$$

49

36

100. The opposite of squaring a number is taking a square root. The square root of 25 or $\sqrt{25} = 5$; $\sqrt{16} =$ __________

4

101. When some arithmetic occurs inside a square-root sign, work the arithmetic before taking the square root.

$$\sqrt{1+3} = \sqrt{4} = 2$$

$$\sqrt{\frac{5^2+7}{2}} = \sqrt{\frac{32}{2}} = \sqrt{16} = \_\_\_\_\_\_$$

$$\sqrt{\frac{(-3)^2+(-1)^2+(-1)^2+5^2}{4}} = \sqrt{\frac{?}{4}} = \sqrt{?} = \_\_\_\_\_\_$$

Note: When we take the square root of a number, both positive and negative roots occur; but we are concerned only with the positive roots.

4

$\sqrt{36/4} = \sqrt{9} = 3$

102. Squaring, dividing and taking the square root are used in solving the formula for the standard deviation: S.D. $= \sqrt{\Sigma x^2/N}$. The symbol $\sqrt{}$ directs a person to take the square __________

ROOT

103. The formula for the standard deviation, $\sqrt{\Sigma x^2/N}$, can be executed in six steps:

1. Compute the mean ($\bar{X}$).
2. Subtract the mean from each score to find the deviations from the mean ($X - \bar{X} = x$).
3. Square each deviation from the mean (x^2).
4. Add the squares of the deviations (Σx^2).
5. Divide the Σx^2 by N ($\Sigma x^2/N$).
6. Take the square root of $\Sigma x^2/N$ ($\sqrt{\Sigma x^2/N}$).

The steps above illustrate why the standard deviation is defined as the square root of the average of the squared deviations from the ______________

MEAN

104. To see how to calculate a standard deviation, let's assume a distribution of 0, 2, 2, 8. The mean ($\bar{X}$) is 3.

$X - \bar{X}$ or x (deviations) = (0–3), (2–3), (2–3), (8–3)

= –3, –1, –1, ____

x^2 (squared deviations) = 9, 1, 1, ____

Σx^2 (sum of squared deviations) = 9 + 1 + 1 + ____ = ____

5

25

25 = 36

105. We have found the sum of the squared deviations (Σx^2) to be 36. With an N of 4, the average of the squared deviations from the mean is $\Sigma x^2/N = 36/4 = 9$. The square root of the average of the squared deviations from the mean equals 1 standard deviation. That is, S.D. $= \sqrt{\Sigma x^2/N} = \sqrt{36/4} =$ __________

3

106. Calculate the standard deviation for the distribution 1, 3, 4, 5, 7.

a. $\bar{X} = 4$

b. $(X - \bar{X})$ or $x =$ (1–4), (3–_), (_–_), (_ _), (_ _)

c. $x^2 =$ __, __, __, __, __

d. $\Sigma x^2 =$ __ + __ + __ + __ + __

e. $\Sigma x^2/N =$ ____

f. S.D. $= \sqrt{\Sigma x^2/N} =$ ____

4, (4–4), (5–4), (7–4)

–3, –1, 0, 1, 3

9, 1, 0, 1, 9

4

2

107. Subtracting a constant from or adding a constant to all the raw scores of a distribution does not change the value of the stand-

ard deviation. If the standard deviation of 50, 52, 52, 58 is 3, the standard deviation of 50–50, 52–50, 52–50, 58–50 or 0, 2, 2, 8 is ____________

Note: The formula used here for computing standard deviations is based directly on the definition of standard deviation. This formula was chosen because it best helps one understand the basic concept of standard deviation. If you have to calculate standard deviations of various data, consult a statistical methods book for the standard deviation formula most appropriate to your data.

3

108. The range is easier to understand and easier to calculate than the standard deviation but it has some serious disadvantages. Not much else can be done with the range. The standard deviation (and its square called the variance) is the basis of a whole branch of statistics. The measure of variation having the greater versatility is the ____________.

STANDARD DEVIATION

109. The size of the range depends a good deal upon the size of the sample. There is more chance of simultaneously drawing a very high score and a very low score when the sample is larger. Consequently, range generally increases with an increase in the size of the ____________

SAMPLE

110. Because all the scores are used in computing the standard deviation while only two scores (the highest and lowest) are used in computing the range, the standard deviation is much more stable than the range. The most stable measure of variability is the ____________

STANDARD DEVIATION

111. For example, a sample of 20 scores could be drawn at random from a population of 200 scores. The standard deviation and the range could now be calculated and the 20 scores returned to the population pile. If this process were repeated many times, the standard deviations would vary in size much less than would the ____________.

RANGE

INTERPRETING TEST SCORES

112. The number of correct answers that a person acquires on a test is called his raw score. Assuming that each question on a test counted one point, a raw score of 12 would mean that an individual answered ____________ questions of the test correctly.

NOMINAL, ORDINAL, AND INTERVAL SCALES

Often the methods used in analyzing or interpreting experimental data differ according to the measurement scale used. The nominal scale merely names categories, such as male-female, or success-failure, or Democrat-Republican-Independent. There may be any number of categories, but every item must be uniquely classified as belonging to one or another of them. If numbers are assigned, such as 1 = male and 2 = female, the numbers have no meaning as an actual numerical measure; in other words, it makes no sense to talk about 1.5 as being half-male and half-female. The ordinal scale is a ranking measure, such as private-corporal-sergeant, or first-second-third place in a race, or A-B-C-D-E in a grading scale. There still may be categories named, but with ordinal measure the categories now are ranked as being "better" or "worse," "higher" or "lower," etc. The interval scale is a true numerical measurement scale in which numerical differences have meaning, such as degrees for temperature, pounds for weight, feet for height, or counting measure for the number of redheads in a class. Sometimes it is difficult to classify certain observations as to measurement scale, and sometimes the classification will vary according to the context of the experiment. For example, the group assignment of children to reading groups would be nominal if the choice of any particular child for Group A, Group B, and so on were made arbitrarily. However, if the children were assigned according to their reading ability, it would be ordinal, since it would rank them. If, after the children are assigned, we wished to count the number of children in each group, our counts would involve the interval scale.