Chapter Concepts Test

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

T F 1. Regression analysis is used to describe how well an estimating equation describes the relationship being studied.

T F 2. Given that the equation for a line is $Y = 26 - 24X$, we may say that the relationship of Y to X is direct and linear.

T F 3. An r^2 value close to zero indicates a strong correlation between X and Y.

T F 4. Regression and correlation analyses are used to determine cause-and-effect relationships.

T F 5. The sample coefficient of correlation, r, is nothing more than $\sqrt{r^2}$, and we cannot interpret its meaning directly as a percentage of some kind.

T F 6. The standard error of estimate measures the variability of the observed values around the regression equation.

T F 7. The regression line is derived from a sample, not the entire population.

T F 8. We may interpret the sample coefficient of determination as the amount of the variation in Y that is explained by the regression line.

T F 9. Lines drawn on either side of the regression line at ± 1, ± 2, and ± 3 times the value of the standard error of estimate are called confidence lines.

T F 10. The estimating equation is valid over only the same range as that given by the original sample data on which it was developed.

T F 11. In the equation $Y = a + bX$ for dependent variable Y and independent variable X, the Y-intercept is b.

T F 12. If a line is fitted to a set of points by the method of least squares, the individual positive and negative errors from the line sum to zero.

T F 13. If $s_e = 0$ for an estimating equation, it must perfectly estimate the dependent variable at the observed points.

T F 14. Suppose the slope of an estimating equation is positive. Then the value of r must be the positive square root of r^2.

T F 15. If $r = 0.8$, then the regression equation explains 80 percent of the total variation in the dependent variable.

T F 16. The coefficient of correlation is the percentage of the total variation of the dependent variable that is explained by the regression.

T F 17. The standard error of estimate is measured perpendicularly from the regression line rather than on the Y-axis.

T F 18. By squaring individual errors, the least-squares method magnifies all deviations from the estimated regression line.

T F 19. A regression equation may not be valid when extended outside the sample range of the independent variable.

T F 20. An r^2 value measures only the strength of a linear relationship between the two variables X and Y.

T F 21. A small value of r^2 implies that there is not a significant cause-and-effect relationship between X and Y.

A B C D E 22. Suppose that we know the height of a student but do not know her weight. We use an estimating equation to determine an estimate of her weight based on her height. We can therefore surmise that:

(a) Weight is the independent variable.

(b) Height is the dependent variable.

(c) The relationship between weight and height is an inverse one.

(d) None of these.

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

A B C D 23. Suppose you are told that there is a direct relationship between the price of artichokes and the amount of rain that fell during the growing season. It can be calculated that:

(a) Prices tend to be high when rainfall is high.

(b) Prices tend to be low when rainfall is high.

(c) A large amount of rain causes prices to rise.

(d) A lack of rain causes prices to rise.

A B C D 24. Suppose it is calculated that a is 4 and b is 2 for a particular estimating line with one independent variable. If the independent variable has a value of 2, what value should be expected for the dependent variable?

(a) 8.

(b) 10.

(c) -1.

(d) 0.

A B C D E 25. Suppose the estimating equation $\hat{Y} = 5 - 2X$ has been calculated for a set of data. Which of the following is true for this situation?

(a) The Y-intercept of the line is 2.

(b) The slope of the line is negative.

(c) The line represents an inverse relationship.

(d) All of these.

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

A B C D 26. We know that the standard error is the same at all points on a regression line because we assumed that:

(a) Observed values for Y are normally distributed around each estimated value of $\hat{Y}$.

(b) The variances of the distribution around each possible value of Y are the same.

(c) All available data were taken into account when the regression line was calculated.

(d) None of these.

A B C D 27. The variation of the Y values around the regression line is best expressed as:

(a) $\Sigma(Y + \overline{Y})^2$.

(b) $\Sigma(Y - \overline{Y})^2$.

(c) $\Sigma(Y - \hat{Y})^2$.

(d) $\Sigma(Y + \hat{Y})^2$.

A B C D 28. The value of r^2 for a particular situation is 0.49. What is the coefficient of correlation?

(a) 0.49.

(b) 0.7.

(c) 0.07.

(d) Cannot be determined from the information given.

A B C D 29. The fraction $\Sigma(Y - \hat{Y})^2 / \Sigma(Y - \overline{Y})^2$ represents:

(a) The fraction of total variation in Y that is unexplained.

(b) The fraction of total variation in Y that is explained.

(c) The fraction of total variation in Y that was caused by changes in X.

(d) None of these.

A B C D 30. In the equation $Y = A + BX + e$, the e represents:
(a) The X-intercept of the observed data.
(b) The value of Y to which others are compared to determine the best fit.
(c) Random disturbances from the population regression line.
(d) None of these.

A B C D 31. Suppose you wish to compare the hypothesized value of B to a sample value of b that has been calculated. Which of the following *must* be calculated before the others?
(a) s_b.
(b) s_e.
(c) s_p.
(d) Calculations can be made in any order.

A B C D E 32. For the estimating equation to be a perfect estimator of the dependent variable, which of these would have to be true?
(a) The standard error of the estimate is zero.
(b) All the data points are on the regression line.
(c) The coefficient of determination is -1.
(d) (a) and (b) but not (c).
(e) All of these.

A B C D 33. If the dependent variable increases as the independent variable increases in an estimating equation, the coefficient of correlation will be in the range:
(a) 0 to -1.
(b) 0 to -0.05
(c) 0 to -2.
(d) None of these.

A B C D 34. Suppose the fraction of variation in Y that is unexplained by the independent variable X is $\frac{1}{4}$. Then r^2 is:
(a) $\frac{1}{4}$.
(b) $\frac{3}{4}$.
(c) $\frac{15}{16}$.
(d) None of these.

A B C D E 35. The sample coefficient of determination is developed from the variation of the observed Y values around:
(a) The mean of the observed independent variables.
(b) The mean of the observed dependent variables.
(c) The fitted regression line.
(d) (b) and (c) but not (a).
(e) (a), (b), and (c).

A B C D E 36. If $Y = a + bX$, the sample regression line, and $Y = A + BX$, the true unknown population regression equation, are equivalent, then the following must be true:
(a) The estimating equation is a perfect estimator of the dependent variable.
(b) All the data points are on the regression line.
(c) $r^2 = 1$.
(d) All of the above.
(e) None of the above.

37. If the dependent variable in a relationship decreases as the independent variable increases, the relationship is _____

38. An association between two variables that is described by a curved line is a _____ one.

39. Every straight line has a _____, which represents how much each unit change of the independent variable changes the dependent variable.

40. The extent to which observed values differ from their predicted values on the regression line is measured by the _____.

41. _____ is a measure of the proportion of variation in the dependent variable that is explained by the regression line.

42. If 75 percent of the variation in the dependent variable is explained by the regression line, then the value of r will be about _____.

43. _____ is used to measure how well the regression line explains the variation of the dependent variable.

44. The sign of r indicates the _____ of the relationship between the two variables X and Y.

45. The method of least squares finds the best-fit line through a set of points, that is, the line that _____ the error between the observed points and the estimated points on the line.