CORRELATION & REGRESSION

MULTIPLE CHOICE QUESTIONS

In the following multiple-choice questions, select the best answer.

1. The correlation coefficient is used to determine:
   a. A specific value of the y-variable given a specific value of the x-variable
   b. A specific value of the x-variable given a specific value of the y-variable
   c. The strength of the relationship between the x and y variables
   d. None of these

2. If there is a very strong correlation between two variables then the correlation coefficient must be
   a. any value larger than 1
   b. much smaller than 0, if the correlation is negative
   c. much larger than 0, regardless of whether the correlation is negative or positive
   d. None of these alternatives is correct.

3. In regression, the equation that describes how the response variable (y) is related to the explanatory variable (x) is:
   a. the correlation model
   b. the regression model
   c. used to compute the correlation coefficient
   d. None of these alternatives is correct.

4. The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study:

   $\hat{y} = -0.0127 + 0.0180x$

   The above equation implies that:
   a. each beer consumed increases blood alcohol by 1.27%
   b. on average it takes 1.8 beers to increase blood alcohol content by 1%
   c. each beer consumed increases blood alcohol by an average of amount of 1.8%
   d. each beer consumed increases blood alcohol by exactly 0.018

5. SSE can never be
   a. larger than SST
   b. smaller than SST
   c. equal to 1
   d. equal to zero

6. Regression modeling is a statistical framework for developing a mathematical equation that describes how
   a. one explanatory and one or more response variables are related
   b. several explanatory and several response variables response are related
   c. one response and one or more explanatory variables are related
   d. All of these are correct.

7. In regression analysis, the variable that is being predicted is the
   a. response, or dependent, variable
   b. independent variable
   c. intervening variable
   d. is usually x

8. Regression analysis was applied to return rates of sparrowhawk colonies. Regression analysis was used to study the relationship between return rate (***x***: % of birds that return to the colony in a given year) and immigration rate (***y***: % of new adults that join the colony per year). The following regression equation was obtained.

   $\hat{y} = 31.9 - 0.34x$

   Based on the above estimated regression equation, if the return rate were to decrease by 10% the rate of immigration to the colony would:
   a. increase by 34%
   b. increase by 3.4%
   c. decrease by 0.34%
   d. decrease by 3.4%

9. In least squares regression, which of the following is **not** a required assumption about the error term ε?
   a. The expected value of the error term is one.
   b. The variance of the error term is the same for all values of x.
   c. The values of the error term are independent.
   d. The error term is normally distributed.

10. Larger values of r^2 (R^2) imply that the observations are more closely grouped about the
   a. average value of the independent variables
   b. average value of the dependent variable
   c. least squares line
   d. origin

11. In a regression analysis if $r^2 = 1$, then
   a. SSE must also be equal to one
   b. SSE must be equal to zero
   c. SSE can be any positive value
   d. SSE must be negative

12. The coefficient of correlation
    a. is the square of the coefficient of determination
    b. is the square root of the coefficient of determination
    c. is the same as r-square
    d. can never be negative

13. In regression analysis, the variable that is used to explain the change in the outcome of an experiment, or some natural process, is called
    a. the x-variable
    b. the independent variable
    c. the predictor variable
    d. the explanatory variable
    e. all of the above (a-d) are correct
    f. none are correct

14. In the case of an algebraic model for a straight line, if a value for the x variable is specified, then
    a. the exact value of the response variable can be computed
    b. the computed response to the independent value will always give a minimal residual
    c. the computed value of y will always be the best estimate of the mean response
    d. none of these alternatives is correct.

15. A regression analysis between sales (in $1000) and price (in dollars) resulted in the following equation:

    $\hat{y} = 50{,}000 - 8X$

    The above equation implies that an
    a. increase of $1 in price is associated with a decrease of $8 in sales
    b. increase of $8 in price is associated with an increase of $8,000 in sales
    c. increase of $1 in price is associated with a decrease of $42,000 in sales
    d. increase of $1 in price is associated with a decrease of $8000 in sales

16. In a regression and correlation analysis if $r^2 = 1$, then
    a. SSE = SST
    b. SSE = 1
    c. SSR = SSE
    d. SSR = SST

17. If the coefficient of determination is a positive value, then the regression equation
    a. must have a positive slope
    b. must have a negative slope
    c. could have either a positive or a negative slope
    d. must have a positive y intercept

18. If two variables, x and y, have a very strong linear relationship, then
    a. there is evidence that x causes a change in y
    b. there is evidence that y causes a change in x
    c. there might not be any causal relationship between x and y
    d. None of these alternatives is correct.

19. If the coefficient of determination is equal to 1, then the correlation coefficient
    a. must also be equal to 1
    b. can be either -1 or +1
    c. can be any value between -1 to +1
    d. must be -1

20. In regression analysis, if the independent variable is measured in kilograms, the dependent variable
    a. must also be in kilograms
    b. must be in some unit of weight
    c. cannot be in kilograms
    d. can be any units

21. The data are the same as for question 4 above. The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study:

    $\hat{y} = -0.0127 + 0.0180x$

    Suppose that the legal limit to drive is a blood alcohol content of 0.08. If Ricky consumed 5 beers the model would predict that he would be:
    a. 0.09 above the legal limit
    b. 0.0027 below the legal limit
    c. 0.0027 above the legal limit
    d. 0.0733 above the legal limit

22. In a regression analysis if SSE = 200 and SSR = 300, then the coefficient of determination is
    a. 0.6667
    b. 0.6000
    c. 0.4000
    d. 1.5000

23. If the correlation coefficient is 0.8, the percentage of variation in the response variable explained by the variation in the explanatory variable is
    a. 0.80%
    b. 80%
    c. 0.64%
    d. 64%

24. If the correlation coefficient is a positive value, then the slope of the regression line
    a. must also be positive
    b. can be either negative or positive
    c. can be zero
    d. can not be zero

25. If the coefficient of determination is 0.81, the correlation coefficient
    a. is 0.6561
    b. could be either + 0.9 or - 0.9
    c. must be positive
    d. must be negative

26. A fitted least squares regression line
    a. may be used to predict a value of y if the corresponding x value is given
    b. is evidence for a cause-effect relationship between x and y
    c. can only be computed if a strong linear relationship exists between x and y
    d. None of these alternatives is correct.

27. Regression analysis was applied between $ sales (***y***) and $ advertising (***x***) across all the branches of a major international corporation. The following regression function was obtained.

    $\hat{y} = 5000 + 7.25x$

    If the advertising budgets of two branches of the corporation differ by $30,000, then what will be the predicted difference in their sales?
    a. $217,500
    b. $222,500
    c. $5000
    d. $7.25

28. Suppose the correlation coefficient between height (as measured in feet) versus weight (as measured in pounds) is 0.40. What is the correlation coefficient of height measured in inches versus weight measured in ounces? [12 inches = one foot; 16 ounces = one pound]
    a. 0.40
    b. 0.30
    c. 0.533
    d. cannot be determined from information given
    e. none of these

29. Assume the same variables as in question 28 above; height is measured in feet and weight is measured in pounds. Now, suppose that the units of both variables are converted to metric (meters and kilograms). The impact on the slope is:
    a. the sign of the slope will change
    b. the magnitude of the slope will change
    c. both a and b are correct
    d. neither a nor b are correct

30. Suppose that you have carried out a regression analysis where the total variance in the response is 133452 and the correlation coefficient was 0.85. The residual sums of squares is:
a. 37032.92
b. 20017.8
c. 113434.2
d. 96419.07
e. 15%
f. 0.15

31. This question is related to questions 4 and 21 above. The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study:

$\hat{y}$= -0.0127 + 0.0180x

Another guy, his name Dudley, has the regression equation written on a scrap of paper in his pocket. Dudley goes out drinking and has 4 beers. He calculates that he is under the legal limit (0.08) so he decides to drive to another bar. Unfortunately Dudley gets pulled over and confidently submits to a road-side blood alcohol test. He scores a blood alcohol of 0.085 and gets himself arrested. Obviously, Dudley skipped the lecture about residual variation. Dudley's residual is:

a. +0.005
b. -0.005
c. +0.0257
d. -0.0257

32. You have carried out a regression analysis; but, after thinking about the relationship between variables, you have decided you must swap the explanatory and the response variables. After refitting the regression model to the data you expect that:
a. the value of the correlation coefficient will change
b. the value of SSE will change
c. the value of the coefficient of determination will change
d. the sign of the slope will change
e. nothing changes

33. Suppose you use regression to predict the height of a woman's current boyfriend by using her own height as the explanatory variable. Height was measured in feet from a sample of 100 women undergraduates, and their boyfriends, at Dalhousie University. Now, suppose that the height of both the women and the men are converted to centimeters. The impact of this conversion on the slope is:
a. the sign of the slope will change
b. the magnitude of the slope will change
c. both a and b are correct
d. neither a nor b are correct

34. A residual plot:
   a. displays residuals of the explanatory variable versus residuals of the response variable.
   b. displays residuals of the explanatory variable versus the response variable.
   c. displays explanatory variable versus residuals of the response variable.
   d. displays the explanatory variable versus the response variable.
   e. displays the explanatory variable on the x axis versus the response variable on the y axis.

35. When the error terms have a constant variance, a plot of the residuals versus the independent variable *x* has a pattern that
   a. fans out
   b. funnels in
   c. fans out, but then funnels in
   d. forms a horizontal band pattern
   e. forms a linear pattern that can be positive or negative

36. You studied the impact of the dose of a new drug treatment for high blood pressure. You think that the drug might be more effective in people with very high blood pressure. Because you expect a bigger change in those patients who start the treatment with high blood pressure, you use regression to analyze the relationship between the initial blood pressure of a patient (*x*) and the change in blood pressure after treatment with the new drug (*y*). If you find a very strong positive association between these variables, then:
   a. there is evidence that the higher the patients initial blood pressure, the bigger the impact of the new drug.
   b. there is evidence that the higher the patients initial blood pressure, the smaller the impact of the new drug.
   c. there is evidence for an association of some kind between the patients initial blood pressure and the impact of the new drug on the patients blood pressure
   d. none of these are correct, this is a case of regression fallacy

Question 37:
A variety of summary statistics were collected for a small sample (10) of bivariate data, where the dependent variable was y and an independent variable was x.

$\Sigma X = 90$ $\quad \Sigma\left(Y - \overline{Y}\right)\left(X - \overline{X}\right) = 466$

$\Sigma Y = 170$ $\quad \Sigma\left(X - \overline{X}\right)^2 = 234$

$n = 10$ $\quad \Sigma\left(Y - \overline{Y}\right)^2 = 1434$

SSE = 505.98

37.1 Use the formula to the right to compute the sample correlation coefficient:
   a. 0.8045
   b. -0.8045
   c. 0
   d. 1

$$r = \frac{\sum_{i=1}^{n}\left(\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}}$$

37.2 The least squares estimate of b_1 equals
a. 0.923
b. 1.991
c. -1.991
d. -0.923

37.3 The least squares estimate of b_0 equals
a. 0.923
b. 1.991
c. -1.991
d. -0.923

37.4 The sum of squares due to regression (SSR) is
a. 1434
b. 505.98
c. 50.598
d. 928.02

37.5 The coefficient of determination equals
a. 0.6471
b. -0.6471
c. 0
d. 1

37.6 The point estimate of y when x = 0.55 is
a. 0.17205
b. 2.018
c. 1.0905
d. -2.018
e. -0.17205

MULTIPLE CHOICE ANSWERS

1. c	11. b	21. b	31. c	37.5 a
2. b	12. b	22. b	32. b	37.6 a
3. b	13. e	23. d	33. d	
4. c	14. a	24. a	34. c	
5. a	15. d	25. b	35. d	
6. c	16. d	26. a	36. d	
7. a	17. c	27. a	37.1 a	
8. b	18. c	28. a	37.2 b	
9. a	19. b	29. b	37.3 d	
10. c	20. d	30. a	37.4 d	