SIMPLE LINEAR REGRESSION:

The HR manager at ACME food services wants to examine the relationship between a workers income and their years of experience on the job. He randomly selects a sample of 10 employees and records their income (in $1,000s) and their length of employment (in years) with the results presented here.

x (experience)	y (income)
13	34
11	36
12	28
15	45
9	23
10	34
12	39
10	27
18	56
12	31

a)   Find the least squares regression line.

b)   Interpret the coefficients

c)   Find and interpret s_ε for the data and interpret what it tells you about how well the regression line fits the data.

d)   Is there sufficient evidence at the 5% significance level to indicate that x and y are linearly related?

e)   Calculate the coefficient of determination for the data and interpret the result.

f)   Determine a 90% prediction interval for the income of an individual who has had 12 years of experience.

g)   Determine a 90% confidence interval for the mean income of all individuals who have had 12 years of experience.

MULTIPLE CHOICE QUESTIONS

1. The regression line $\hat{y} = 3 + 2x$ has been fitted to the data points (4, 8), (2, 5), and (1, 2). The sum of the squared residuals will be:
    a. 7
    b. 15
    c. 8
    d. 22

2. If an estimated regression line has a *y*-intercept of 10 and a slope of 4, then when $x = 2$ the actual value of *y* is:
    a. 18
    b. 15
    c. 14
    d. unknown

3. Given the least squares regression line $\hat{y} = 5 - 2x$:
    a. the relationship between *x* and *y* is positive
    b. the relationship between *x* and *y* is negative
    c. as *x* increases, so does *y*
    d. as *x* decreases, so does *y*

4. A regression analysis between weight (*y* in pounds) and height (*x* in inches) resulted in the following least squares line: $\hat{y} = 120 + 5x$. This implies that if the height is increased by 1 inch, the weight, on average, is expected to:
    a. increase by 1 pound
    b. decrease by 1 pound
    c. increase by 5 pounds
    d. increase by 24 pounds

5. A regression analysis between sales (in $1000) and advertising (in $100) resulted in the following least squares line: $\hat{y} = 75 + 6x$. This implies that if advertising is $800, then the predicted amount of sales (in dollars) is:
   a. $4875
   b. $123,000
   c. $487,500
   d. $12,300

6. A regression analysis between sales (in $1000) and advertising (in $) resulted in the following least squares line: $\hat{y} = 80,000 + 5x$. This implies that an:
   a. increase of $1 in advertising is expected, on average, to result in an increase of $5 in sales
   b. increase of $5 in advertising is expected, on average, to result in an increase of $5,000 in sales
   c. increase of $1 in advertising is expected, on average, to result in an increase of $80,005 in sales
   d. increase of $1 in advertising is expected, on average, to result in an increase of $5,000 in sales

7. Which of the following techniques is used to predict the value of one variable on the basis of other variables?
   a. Correlation analysis
   b. Coefficient of correlation
   c. Covariance
   d. Regression analysis

8. The residual is defined as the difference between:
   a. the actual value of y and the estimated value of y
   b. the actual value of x and the estimated value of x
   c. the actual value of y and the estimated value of x
   d. the actual value of x and the estimated value of y

9. In the simple linear regression model, the *y*-intercept represents the:
    a. change in *y* per unit change in *x*
    b. change in *x* per unit change in *y*
    c. value of *y* when *x* = 0
    d. value of *x* when *y* = 0

10. In the first order linear regression model, the population parameters of the *y*-intercept and the slope are estimated respectively, by:
    a. b_0 and b_1
    b. b_0 and β_1
    c. β_0 and b_1
    d. β_0 and β_1

11. In the simple linear regression model, the slope represents the:
    a. value of *y* when *x* = 0
    b. average change in *y* per unit change in *x*
    c. value of *x* when *y* = 0
    d. average change in *x* per unit change in *y*

12. In regression analysis, the residuals represent the:
    a. difference between the actual *y* values and their predicted values
    b. difference between the actual *x* values and their predicted values
    c. square root of the slope of the regression line
    d. change in *y* per unit change in *x*

13. In the first-order linear regression model, the population parameters of the *y*-intercept and the slope are, respectively,
    a. b_0 and b_1
    b. b_0 and β_1
    c. β_0 and b_1
    d. β_0 and β_1

14. In a simple linear regression problem, the following statistics are calculated from a sample of 10 observations: $\sum(x-\bar{x})(y-\bar{y}) = 2250$, $s_x = 10, \sum x = 50$, $\sum y = 75$. The least squares estimates of the slope and *y*-intercept are respectively:
    a. 1.5 and 0.5
    b. 2.5 and 1.5
    c. 1.5 and 2.5
    d. 2.5 and –5.0

15. If a simple linear regression model has no *y*-intercept, then:
    a. all values of *x* are zero
    b. all values of *y* are zero
    c. when *y* = 0 so does *x*
    d. when *x* = 0 so does *y*

16. In the least squares regression line $\hat{y} = 3 - 2x$, the predicted value of *y* equals:
    a. 1.0 when *x* = -1.0
    b. 2.0 when *x* = 1.0
    c. 2.0 when *x* = -1.0
    d. 1.0 when *x* = 1.0

17. The least squares method for determining the best fit minimizes:
    a. total variation in the dependent variable
    b. sum of squares for error
    c. sum of squares for regression
    d. All of the above

18. In order to predict with 90% confidence the expected value of *y* for a given value of *x* in a simple linear regression problem, a random sample of 10 observations is taken. Which of the following *t*-table values listed below would be used?
    a. 2.228
    b. 2.306
    c. 1.860
    d. 1.812

19. If the coefficient of correlation is 0.90, then the percentage of the variation in the dependent variable *y* that is explained by the variation in the independent variable *x* is:
    a. 90%
    b. 81%
    c. 95%
    d. None of these choices.

20. In simple linear regression, the coefficient of correlation *r* and the least squares estimate b_1 of the population slope β_1:
    a. must be equal.
    b. must have the same sign.
    c. are not related.
    d. None of these choices.

21. If the sum of squared residuals is zero, then the:
    a. coefficient of determination must be 1.0.
    b. coefficient of correlation must be 1.0.
    c. linear relationship between *x* and *y* is perfect.
    d. All of these choices are true.

Chapter 16 MCQ Answer Key

1 d	2 d	3 b	4c	5 b	6 d	7 d	8 a	9 c	10 a
11 b	12 a	13 d	14 d	15 d	16 d	17 b	18 c	19 b	20 b
21 d									