Name: _____ **Class:** _____ **Date:** _____ **ID: A**

Regression

Multiple Choice
Identify the choice that best completes the statement or answers the question.

____ 1. 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 decreases, so does y.
   d. None of these choices.

____ 2. A regression analysis between sales (in $1,000) and advertising (in $1,000) resulted in the following least squares line: $\hat{y} = 80 + 5x$. This implies that:
   a. as advertising increases by $1,000, sales increases by $5,000.
   b. as advertising increases by $1,000, sales increases by $80,000.
   c. as advertising increases by $5, sales increases by $80.
   d. None of these choices.

____ 3. 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

____ 4. 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

____ 5. Testing whether the slope of the population regression line could be zero is equivalent to testing whether the:
   a. sample coefficient of correlation could be zero
   b. standard error of estimate could be zero
   c. population coefficient of correlation could be zero
   d. sum of squares for error could be zero

____ 6. A regression line using 25 observations produced SSR = 118.68 and SSE = 56.32. The standard error of estimate was:
   a. 2.11
   b. 1.56
   c. 2.44
   d. None of these choices.

____ 7. If all the points in a scatter diagram lie on the least squares regression line, then the coefficient of correlation must be:
   a. 1.0
   b. −1.0
   c. either 1.0 or −1.0
   d. 0.0

____ 8. If the coefficient of correlation is −0.60, then the coefficient of determination is:
a. −0.60
b. −0.36
c. 0.36
d. 0.77

____ 9. The standard error of estimate s_ε is given by:
a. $\text{SSE} / (n-2)$
b. $\sqrt{\text{SSE} / (n-2)}$
c. $\sqrt{\text{SSE} / (n-2)}$
d. $\text{SSE} / \sqrt{n-2}$

____ 10. If the standard error of estimate $s_\varepsilon = 20$ and $n = 10$, then the sum of squares for error, SSE, is:
a. 400
b. 3,200
c. 4,000
d. 40,000

____ 11. In regression analysis, the coefficient of determination R^2 measures the amount of variation in y that is:
a. caused by the variation in x.
b. explained by the variation in x.
c. unexplained by the variation in x.
d. None of these choices.

____ 12. In a regression problem, if the coefficient of determination is 0.95, this means that:
a. 95% of the y values are positive.
b. 95% of the variation in y can be explained by the variation in x.
c. 95% of the y values are predicted correctly by the model.
d. None of these choices.

____ 13. The sample correlation coefficient between x and y is 0.375. It has been found out that the p-value is 0.256 when testing H_0: $\rho = 0$ against the two-sided alternative H_1: $\rho \neq 0$. To test H_0: $\rho = 0$ against the one-sided alternative H_1: $\rho > 0$ at a significant level of 0.193, the p-value will be equal to
a. 0.128
b. 0.512
c. 0.744
d. 0.872

____ 14. In simple linear regression, which of the following statements indicates there is no linear relationship between the variables x and y?
a. Coefficient of determination is −1.0.
b. Coefficient of correlation is 0.0.
c. Sum of squares for error is 0.0.
d. None of these choices.

____ 15. 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.

____ 16. 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.

____ 17. The width of the confidence interval estimate for the predicted value of y depends on
a. the standard error of the estimate
b. the value of x for which the prediction is being made
c. the sample size
d. All of these choices are true.

____ 18. In a multiple regression model, the mean of the probability distribution of the error variable ε is assumed to be:
a. 1.0
b. 0.0
c. k, where k is the number of independent variables included in the model.
d. None of these choices.

____ 19. In a multiple regression analysis involving 6 independent variables, the total variation in y is 900 and SSR = 600. What is the value of SSE?
a. 300
b. 1.50
c. 0.67
d. None of these choices.

____ 20. For a multiple regression model the following statistics are given: Total variation in y = 250, SSE = 50, k = 4, and n = 20. Then, the coefficient of determination adjusted for the degrees of freedom is:
a. 0.800
b. 0.747
c. 0.840
d. 0.775

____ 21. A multiple regression model has the form: $\hat{y} = 5.25 + 2x_1 + 6x_2$. As x_2 increases by one unit, holding x_1 constant, then the value of y will increase by:
a. 2 units
b. 7.25 units
c. 6 units on average
d. None of these choices

____ 22. If all the points for a multiple regression model with two independent variables were right on the regression plane, then the coefficient of determination would equal:
a. 0.
b. 1.
c. 2, since there are two independent variables.
d. None of these choices.

____ 23. For a multiple regression model, the total variation in *y* can be expressed as:
a. SSR + SSE.
b. SSR − SSE.
c. SSE − SSR.
d. SSR / SSE.

____ 24. A multiple regression equation includes 5 independent variables, and the coefficient of determination is 0.81. The percentage of the variation in *y* that is explained by the regression equation is:
a. 81%
b. 90%
c. 86%
d. about 16%

____ 25. In a multiple regression model, the value of the coefficient of determination has to fall between
a. −1 and +1.
b. 0 and +1.
c. −1 and 0.
d. None of these choices.

True/False
Indicate whether the statement is true or false.

____ 26. In a simple linear regression problem, the least squares line is $\hat{y} = -3.75 + 1.25x$, and the coefficient of determination is 0.81. The coefficient of correlation must be −0.90.

____ 27. A zero correlation coefficient between a pair of random variables means that there is no linear relationship between the random variables.

____ 28. In reference to the equation $\hat{y} = -0.80 + 0.12x_1 + 0.08x_2$, the value −0.80 is the *y*-intercept.

____ 29. In multiple regression, the standard error of estimate is defined by $s_\varepsilon = \sqrt{\text{SSE}/(n-k)}$, where *n* is the sample size and *k* is the number of independent variables.

____ 30. One of the consequences of multicollinearity in multiple regression is biased estimates on the slope coefficients.

____ 31. Multicollinearity is present when there is a high degree of correlation between the independent variables included in the regression model.

Short Answer

Car Speed and Gas Mileage

An economist wanted to analyze the relationship between the speed of a car (x) and its gas mileage (y). As an experiment a car is operated at several different speeds and for each speed the gas mileage is measured. These data are shown below.

Speed	25	35	45	50	60	65	70
Gas Mileage	40	39	37	33	30	27	25

32. {Car Speed and Gas Mileage Narrative} Estimate the gas mileage of a car traveling 70 mph.

33. A scatter diagram includes the following data points:

x	3	2	5	4	5
y	8	6	12	10	14

Two regression models are proposed: (1) $\hat{y} = 1.2 + 2.5x$, and (2) $\hat{y} = 4.0x$. Using the least squares method, which of these regression models provides the better fit to the data? Why?

Sunshine and Skin Cancer

A medical statistician wanted to examine the relationship between the amount of sunshine (x) in hours, and incidence of skin cancer (y). As an experiment he found the number of skin cancer cases detected per 100,000 of population and the average daily sunshine in eight counties around the country. These data are shown below.

Average Daily Sunshine	5	7	6	7	8	6	4	3
Skin Cancer per 100,000	7	11	9	12	15	10	7	5

34. {Sunshine and Skin Cancer Narrative} Draw a scatter diagram of the data and plot the least squares regression line on it.

35. {Sunshine and Skin Cancer Narrative} Estimate the number of skin cancer cases per 100,000 people who live in a state that gets 6 hours of sunshine on average.

36. {Sunshine and Skin Cancer Narrative} Can we conclude at the 1% significance level that there is a linear relationship between sunshine and skin cancer?

Game Winnings & Education

An ardent fan of television game shows has observed that, in general, the more educated the contestant, the less money he or she wins. To test her belief she gathers data about the last eight winners of her favorite game show. She records their winnings in dollars and the number of years of education. The results are as follows.

Contestant	Years of Education	Winnings
1	11	750
2	15	400
3	12	600
4	16	350
5	11	800
6	16	300
7	13	650
8	14	400

37. {Game Winnings & Education Narrative} Draw a scatter diagram of the data. Comment on whether it appears that a linear model might be appropriate.

38. {Game Winnings & Education Narrative} Interpret the value of the slope of the regression line.

39. {Game Winnings & Education Narrative} Conduct a test of the population coefficient of correlation to determine at the 5% significance level whether a negative linear relationship exists between years of education and TV game shows' winnings.

Oil Quality and Price

Quality of oil is measured in API gravity degrees--the higher the degrees API, the higher the quality. The table shown below is produced by an expert in the field who believes that there is a relationship between quality and price per barrel.

Oil degrees API	Price per barrel (in $)
27.0	12.02
28.5	12.04
30.8	12.32
31.3	12.27
31.9	12.49
34.5	12.70
34.0	12.80
34.7	13.00
37.0	13.00
41.0	13.17
41.0	13.19
38.8	13.22
39.3	13.27

A partial Minitab output follows:

Descriptive Statistics

Variable	N	Mean	StDev	SE Mean
Degrees	13	34.60	4.613	1.280
Price	13	12.730	0.457	0.127

Covariances

	Degrees	Price
Degrees	21.281667	
Price	2.026750	0.208833

Regression Analysis

Predictor	Coef	StDev	T	P
Constant	9.4349	0.2867	32.91	0.000
Degrees	0.095235	0.008220	11.59	0.000

S = 0.1314 R-Sq = 92.46% R-Sq(adj) = 91.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	2.3162	2.3162	134.24	0.000
Residual Error	11	0.1898	0.0173		
Total	12	2.5060			

40. {Oil Quality and Price Narrative} Draw a scatter diagram of the data. Comment on whether it appears that a linear model might be appropriate to describe the relationship between the quality of oil and price per barrel.

41. The following 10 observations of variables x and y were collected.

x	1	2	3	4	5	6	7	8	9	10
y	25	22	21	19	14	15	12	10	6	2

a. Calculate the standard error of estimate.
b. Test to determine if there is enough evidence at the 5% significance level to indicate that x and y are negatively linearly related.
c. Calculate the coefficient of correlation, and describe what this statistic tells you about the regression line.

Sales and Experience

The general manager of a chain of furniture stores believes that experience is the most important factor in determining the level of success of a salesperson. To examine this belief she records last month's sales (in $1,000s) and the years of experience of 10 randomly selected salespeople. These data are listed below.

Salesperson	Years of Experience	Sales
1	0	7
2	2	9
3	10	20
4	3	15
5	8	18
6	5	14
7	12	20
8	7	17
9	20	30
10	15	25

42. (Sales and Experience Narrative} Determine the coefficient of determination and discuss what its value tells you about the two variables.

43. {Sales and Experience Narrative} Estimate with 95% confidence the average monthly sales of all salespersons with 10 years of experience.

44. {Sales and Experience Narrative} Compute the standardized residuals.

Life Expectancy

An actuary wanted to develop a model to predict how long individuals will live. After consulting a number of physicians, she collected the age at death (y), the average number of hours of exercise per week (x_1), the cholesterol level (x_2), and the number of points that the individual's blood pressure exceeded the recommended value (x_3). A random sample of 40 individuals was selected. The computer output of the multiple regression model is shown below.

THE REGRESSION EQUATION IS

$y = 55.8 + 1.79x_1 - 0.021x_2 - 0.061x_3$

Predictor	*Coef*	*StDev*	*T*
Constant	55.8	11.8	4.729
x_1	1.79	0.44	4.068
x_2	–0.021	0.011	–1.909
x_3	–0.016	0.014	–1.143

S = 9.47 R–Sq = 22.5%

ANALYSIS OF VARIANCE

Source of Variation	*df*	*SS*	*MS*	*F*
Regression	3	936	312	3.477
Error	36	3230	89.722	
Total	39	4166		

45. {Life Expectancy Narrative} Is there sufficient evidence at the 5% significance level to infer that the number of points that the individual's blood pressure exceeded the recommended value and the age at death are negatively linearly related?

Regression
Answer Section

MULTIPLE CHOICE

1.	ANS: B	PTS: 1	REF: SECTION 16.1-16.2
2.	ANS: A	PTS: 1	REF: SECTION 16.1-16.2
3.	ANS: D	PTS: 1	REF: SECTION 16.1-16.2
4.	ANS: A	PTS: 1	REF: SECTION 16.1-16.2
5.	ANS: C	PTS: 1	REF: SECTION 16.3-16.4
6.	ANS: B	PTS: 1	REF: SECTION 16.3-16.4
7.	ANS: C	PTS: 1	REF: SECTION 16.3-16.4
8.	ANS: C	PTS: 1	REF: SECTION 16.3-16.4
9.	ANS: C	PTS: 1	REF: SECTION 16.3-16.4
10.	ANS: B	PTS: 1	REF: SECTION 16.3-16.4
11.	ANS: B	PTS: 1	REF: SECTION 16.3-16.4
12.	ANS: B	PTS: 1	REF: SECTION 16.3-16.4
13.	ANS: A	PTS: 1	REF: SECTION 16.3-16.4
14.	ANS: B	PTS: 1	REF: SECTION 16.3-16.4
15.	ANS: B	PTS: 1	REF: SECTION 16.3-16.4
16.	ANS: B	PTS: 1	REF: SECTION 16.3-16.4
17.	ANS: D	PTS: 1	REF: SECTION 16.5
18.	ANS: B	PTS: 1	REF: SECTION 17.1-17.2
19.	ANS: A	PTS: 1	REF: SECTION 17.1-17.2
20.	ANS: B	PTS: 1	REF: SECTION 17.1-17.2
21.	ANS: C	PTS: 1	REF: SECTION 17.1-17.2
22.	ANS: B	PTS: 1	REF: SECTION 17.1-17.2
23.	ANS: A	PTS: 1	REF: SECTION 17.1-17.2
24.	ANS: A	PTS: 1	REF: SECTION 17.1-17.2
25.	ANS: B	PTS: 1	REF: SECTION 17.1-17.2

TRUE/FALSE

26.	ANS: F	PTS: 1	REF: SECTION 16.3-16.4
27.	ANS: T	PTS: 1	REF: SECTION 16.3-16.4
28.	ANS: T	PTS: 1	REF: SECTION 17.1-17.2
29.	ANS: F	PTS: 1	REF: SECTION 17.1-17.2
30.	ANS: F	PTS: 1	REF: SECTION 17.3
31.	ANS: T	PTS: 1	REF: SECTION 17.3

SHORT ANSWER

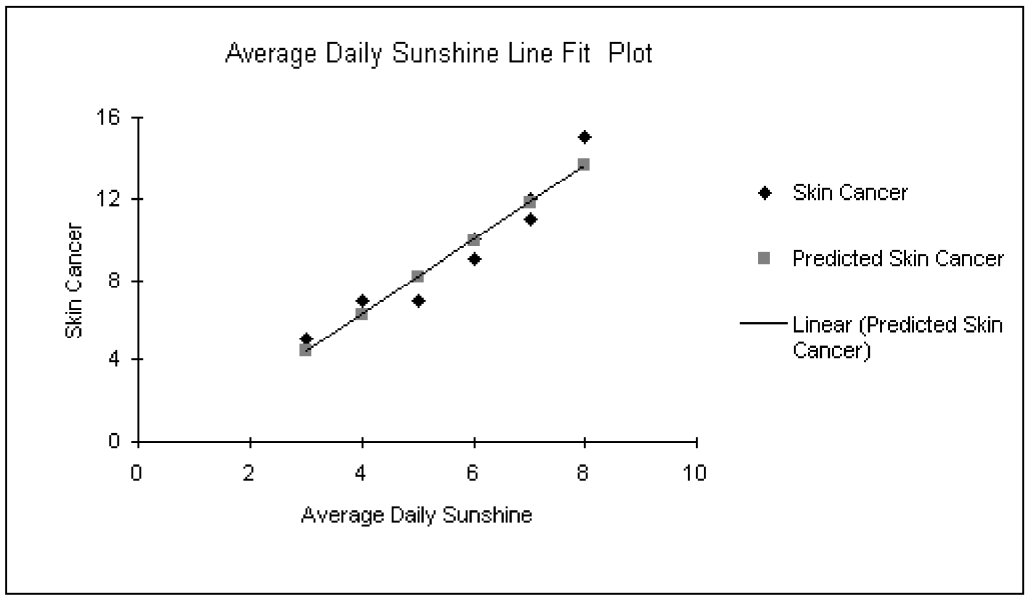
32. ANS:
When $x = 70$, $\hat{y} = 25.9393$ mpg.

PTS: 1 REF: SECTION 16.1-16.2

33. ANS:
SSE = 4.95 and 126.25 for models 1 and 2, respectively. Therefore, model (1) fits the data better than model (2).

PTS: 1 REF: SECTION 16.1-16.2

34. ANS:



PTS: 1 REF: SECTION 16.1-16.2

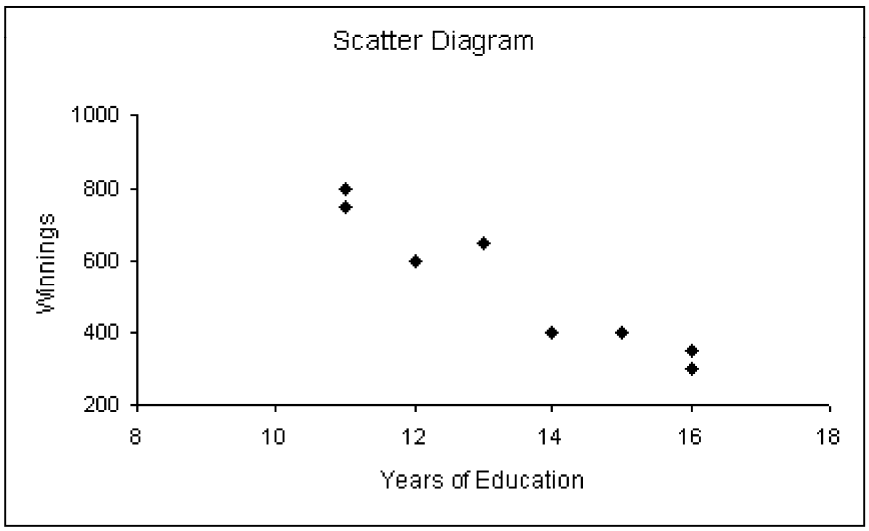
35. ANS:
When $x = 6$, $\hat{y} = 9.961$.

PTS: 1 REF: SECTION 16.1-16.2

36. ANS:
H_0: $\rho = 0$ vs. H_1: $\rho \neq 0$
Rejection region: $| t | > t_{0.005,6} = 3.707$
Test statistic: $t = 8.485$
Conclusion: Reject the null hypothesis. We conclude at the 1% significance level that there is a linear relationship between sunshine and skin cancer, according to this data. The relationship is positive, indicating that more sunshine is associated with more skin cancer cases.

PTS: 1 REF: SECTION 16.3-16.4

37. ANS:



It appears that a linear model is appropriate; further analysis is needed.

PTS: 1 REF: SECTION 16.1-16.2

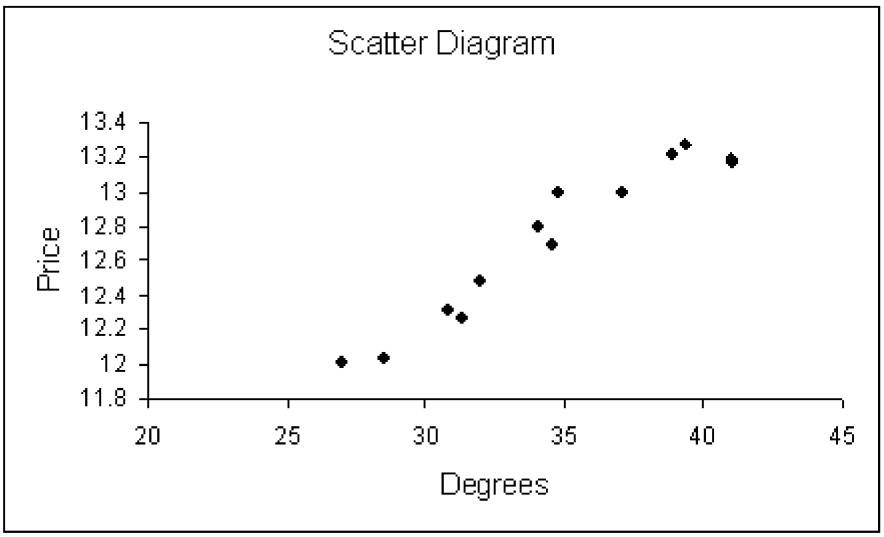
38. ANS:
For each additional year of education a contestant has, his or her winnings on TV game shows decreases by an average of approximately $89.20.

PTS: 1 REF: SECTION 16.1-16.2

39. ANS:
H_0: $\rho = 0$ vs. H_1: $\rho < 0$
Rejection region: $t < -t_{0.05,6} = -1.943$
Test statistic: $t = -8.2227$
Conclusion: Reject the null hypothesis. A negative linear relationship exists between years of education and TV game shows' winnings, according to this data.

PTS: 1 REF: SECTION 16.3-16.4

40. ANS:



A linear model might be appropriate to describe the relationship between the quality of oil and price per barrel.

PTS: 1 REF: SECTION 16.1-16.2

41. ANS:

a. $s_\varepsilon = 1.322$

b. H_0: $\beta_1 = 0$ vs. H_0: $\beta_1 < 0$
Rejection region: $t < t_{0.05,8} = -1.86$
Test statistic: $t = -16.402$
Conclusion: Reject the null hypothesis. There is enough evidence at the 5% significance level to indicate that x and y have a negative linear relationship, according to this data. As speed increases, gas mileage decreases.

c. $r = -0.9854$. This indicates a very strong negative linear relationship between the two variables.

PTS: 1 REF: SECTION 16.3-16.4

42. ANS:
$R^2 = 0.9536$, which means that 95.36% of the variation in sales is explained by the variation in years of experience of the salesperson.

PTS: 1 REF: SECTION 16.3-16.4

43. ANS:
19.447 ± 1.199. Thus LCL = 18.248 (in $1000s), and UCL = 20.646 (in $1000s).

PTS: 1 REF: SECTION 16.5

44. ANS:
−1.100, −1.210, 0.373, 2.108, 0.483, −0.026, −1.086, 0.538, −0.178, and 0.097

PTS: 1 REF: SECTION 16.6

45. ANS:
$H_0: \beta_3 = 0$ vs. $H_1: \beta_3 < 0$
Rejection region: $t < -t_{0.05,36} \approx -1.69$
Test statistic: $t = -1.143$
Conclusion: Don't reject the null hypothesis. No, sufficient evidence at the 5% significance level to infer that the number of points that the individual's blood pressure exceeded the recommended value and the age at death are negatively linearly related.

PTS: 1 REF: SECTION 17.1-17.2