

## Part 2 Barrons Questions

## Questions on Topic Four: Exploring Bivariate Data

### Multiple-Choice Questions

**Directions:** The questions or incomplete statements that follow are each followed by five suggested answers or completions. Choose the response that best answers the question or completes the statement.

1. A study collects data on average combined SAT scores (math, critical reading, and writing) and percentage of students who took the exam at 100 randomly selected high schools. Following is part of the computer printout for regression:

Variable	Coefficient	s.e. of coeff	t-ratio	prob
Constant	1576.32	12.65	124.6	$\leq 0.0001$
SAT	-2.84276	0.2461	-11.55	$\leq 0.0001$
R-squared = 76.5%    R-squared (adj) = 76.1%				

Which of the following is a correct conclusion?

- (A) SAT in the variable column indicates that SAT is the dependent (response) variable.
  - (B) The correlation is  $\pm 0.875$ , but the sign cannot be determined.
  - (C) The  $y$ -intercept indicates the mean combined SAT score if percent of students taking the exam has no effect on combined SAT scores.
  - (D) The  $R^2$  value indicates that the residual plot does not show a strong pattern.
  - (E) Schools with lower percentages of students taking the exam tend to have higher average combined SAT scores.
2. A simple random sample of 35 world-ranked chess players provides the following statistics:

Number of hours of study per day:  $\bar{x} = 6.2$ ,  $s_x = 1.3$

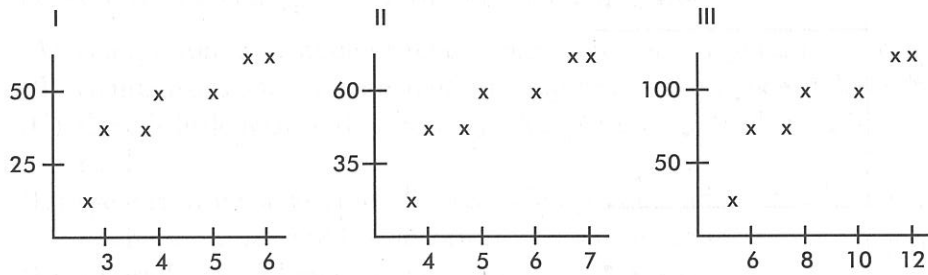
Yearly winnings:  $\bar{y} = \$208,000$ ,  $s_y = \$42,000$

Correlation  $r = .15$

Based on this data, what is the resulting linear regression equation?

- (A)  $\widehat{\text{Winnings}} = 178,000 + 4850 \text{ Hours}$
- (B)  $\widehat{\text{Winnings}} = 169,000 + 6300 \text{ Hours}$
- (C)  $\widehat{\text{Winnings}} = 14,550 + 31,200 \text{ Hours}$
- (D)  $\widehat{\text{Winnings}} = 7750 + 32,300 \text{ Hours}$
- (E)  $\widehat{\text{Winnings}} = -52,400 + 42,000 \text{ Hours}$

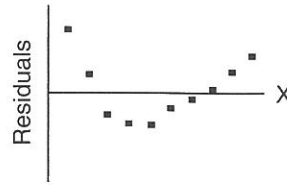
3. A rural college is considering constructing a windmill to generate electricity but is concerned over noise levels. A study is performed measuring noise levels (in decibels) at various distances (in feet) from the campus library, and a least squares regression line is calculated with a correlation of 0.74. Which of the following is a proper and most informative conclusion for an observation with a negative residual?
- The measured noise level is 0.74 times the predicted noise level.
  - The predicted noise level is 0.74 times the measured noise level.
  - The measured noise level is greater than the predicted noise level.
  - The predicted noise level is greater than the measured noise level.
  - The slope of the regression line at that point must also be negative.
4. Consider the following three scatterplots:



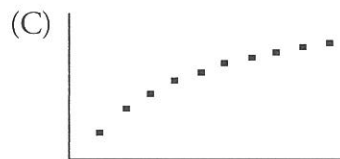
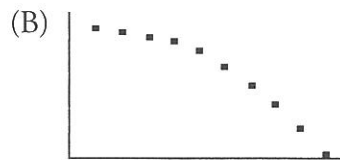
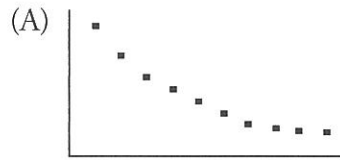
Which has the greatest correlation coefficient?

- I
  - II
  - III
  - They all have the same correlation coefficient.
  - This question cannot be answered without additional information.
5. Suppose the correlation is negative. Given two points from the scatterplot, which of the following is possible?
- The first point has a larger  $x$ -value and a smaller  $y$ -value than the second point.
  - The first point has a larger  $x$ -value and a larger  $y$ -value than the second point.
  - The first point has a smaller  $x$ -value and a larger  $y$ -value than the second point.
- I only
  - II only
  - III only
  - I and III
  - I, II, and III

6. Consider the following residual plot:



Which of the following scatterplots could have resulted in the above residual plot? (The  $y$ -axis scales are not the same in the scatterplots as in the residual plot.)



(E) None of these could result in the given residual plot.

7. Suppose the regression line for a set of data,  $\hat{y} = 3x + b$ , passes through the point  $(2, 5)$ . If  $\bar{x}$  and  $\bar{y}$  are the sample means of the  $x$ - and  $y$ -values, respectively, then  $\bar{y} =$

- (A)  $\bar{x}$ .  
 (B)  $\bar{x} - 2$ .  
 (C)  $\bar{x} + 5$ .  
 (D)  $3\bar{x}$ .  
 (E)  $3\bar{x} - 1$ .

8. Suppose a study finds that the correlation coefficient relating family income to SAT scores is  $r = +1$ . Which of the following are proper conclusions?
- I. Poverty causes low SAT scores.
  - II. Wealth causes high SAT scores.
  - III. There is a very strong association between family income and SAT scores.
- (A) I only  
 (B) II only  
 (C) III only  
 (D) I and II  
 (E) I, II, and III
9. A study of department chairperson ratings and student ratings of the performance of high school statistics teachers reports a correlation of  $r = 1.15$  between the two ratings. From this information we can conclude that
- (A) chairpersons and students tend to agree on who is a good teacher.
  - (B) chairpersons and students tend to disagree on who is a good teacher.
  - (C) there is little relationship between chairperson and student ratings of teachers.
  - (D) there is strong association between chairperson and student ratings of teachers, but it would be incorrect to infer causation.
  - (E) a mistake in arithmetic has been made.
10. Which of the following statements about correlation  $r$  is true?
- (A) A correlation of .2 mean that 20% of the points are highly correlated.
  - (B) Perfect correlation, that is, when the points lie exactly on a straight line, results in  $r = 0$ .
  - (C) Correlation is not affected by which variable is called  $x$  and which is called  $y$ .
  - (D) Correlation is not affected by extreme values.
  - (E) A correlation of .75 indicates a relationship that is 3 times as linear as one for which the correlation is only .25.

Questions 11–13 refer to the following:

The relationship between winning game proportions when facing the sun and when the sun is on one's back is analyzed for a random sample of 10 professional players. The computer printout for regression is below:

Predictor	Coef	SE Coef	T	P
Constant	0.05590	0.02368	2.36	0.046
Facing	0.92003	0.03902	23.58	0.000

S = 0.0242922    R-Sq = 98.6%    R-Sq(adj) = 98.4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	0.32812	0.32812	556.03	0.000
Residual Error	8	0.00472	0.00059		
Total	9	0.33284			

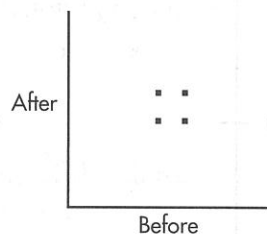
11. What is the equation of the regression line, where *face* and *back* are the winning game proportions when facing the sun and with back to the sun, respectively?
- (A)  $\widehat{face} = 0.056 + 0.920 \text{ back}$   
(B)  $\widehat{back} = 0.056 + 0.920 \text{ facing}$   
(C)  $\widehat{face} = 0.920 + 0.056 \text{ back}$   
(D)  $\widehat{back} = 0.920 + 0.056 \text{ facing}$   
(E)  $\widehat{face} = 0.024 + 0.039 \text{ back}$
12. What is the *correlation*?
- (A)  $-.984$   
(B)  $-.986$   
(C)  $.984$   
(D)  $.986$   
(E)  $.993$
13. For one player, the winning game proportions were 0.55 and 0.59 for *facing* and *back*, respectively. What was the associated *residual*?
- (A)  $-0.028$   
(B)  $0.028$   
(C)  $-0.0488$   
(D)  $0.0488$   
(E)  $0.3608$
14. Which of the following statements about residuals are true?
- I. The mean of the residuals is always zero.  
II. The regression line for a residual plot is a horizontal line.  
III. A definite pattern in the residual plot is an indication that a nonlinear model will show a better fit to the data than the straight regression line.
- (A) I and II  
(B) I and III  
(C) II and III  
(D) I, II, and III  
(E) None of the above gives the complete set of true responses.

15. Data are obtained for a group of college freshmen examining their SAT scores (math plus writing plus critical reading) from their senior year of high school and their GPAs during their first year of college. The resulting regression equation is

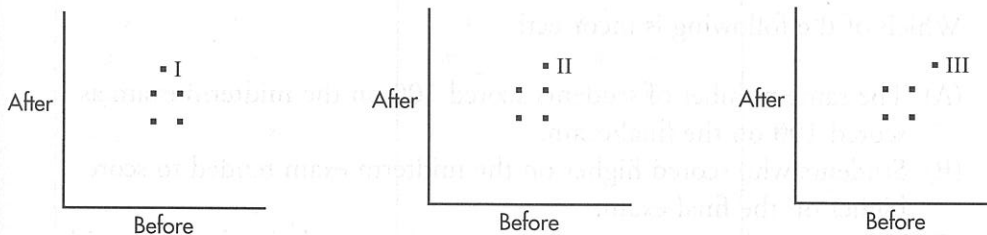
$$\widehat{\text{GPA}} = 0.55 + 0.00161 (\text{SAT total}) \quad \text{with} \quad r = .632$$

What percentage of the variation in GPAs can be explained by looking at SAT scores?

- (A) 0.161%  
 (B) 16.1%  
 (C) 39.9%  
 (D) 63.2%  
 (E) This value cannot be computed from the information given.
16. In a study of whether the structure of the adult human brain changes when a new skill is learned, the gray matter volume of four individuals was measured before and after learning a new cognitive skill. The resulting scatterplot was:



The correlation above is 0. Three researchers each run the experiment on a new subject and each obtain an additional data point:



Match the above scatterplots with their new correlations.

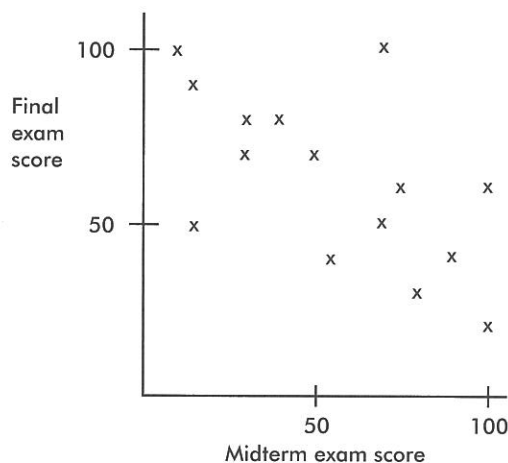
- (A) I: -0.33    II: 0    III: 0.33  
 (B) I: 0    II: 0.33    III: 0.64  
 (C) I: 0    II: 0.33    III: 1.0  
 (D) I: -0.33    II: 0    III: 1.0  
 (E) I: 0    II: 0.50    III: 1.0

17. In a study of winning percentage in home games versus average home attendance for professional baseball teams, the resulting regression line is:

$$\overline{\text{Winning percentage}} = 44 + 0.0003$$

What is the residual if a team has a winning percentage of 55% with an average attendance of 34,000?

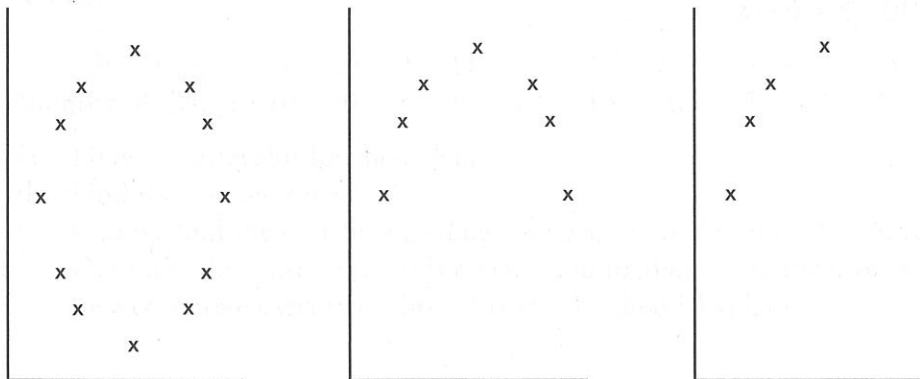
- (A) -11.0  
 (B) -0.8  
 (C) 0.8  
 (D) 11.0  
 (E) 23.0
18. Consider the following scatterplot of midterm and final exam scores for a class of 15 students.



Which of the following is incorrect?

- (A) The same number of students scored 100 on the midterm exam as scored 100 on the final exam.  
 (B) Students who scored higher on the midterm exam tended to score higher on the final exam.  
 (C) The scatterplot shows a moderate negative correlation between midterm and final exam scores.  
 (D) The coefficient of determination here is positive.  
 (E) No one scored 90 or above on both exams.
19. If every woman married a man who was exactly 2 inches taller than she, what would the correlation between the heights of married men and women be?
- (A) Somewhat negative  
 (B) 0  
 (C) Somewhat positive  
 (D) Nearly 1  
 (E) 1

20. Suppose the correlation between two variables is  $r = .23$ . What will the new correlation be if .14 is added to all values of the  $x$ -variable, every value of the  $y$ -variable is doubled, and the two variables are interchanged?
- (A) .23  
(B) .37  
(C) .74  
(D)  $-.23$   
(E)  $-.74$
21. Suppose the correlation between two variables is  $-.57$ . If each of the  $y$ -scores is multiplied by  $-1$ , which of the following is true about the new scatterplot?
- (A) It slopes up to the right, and the correlation is  $-.57$ .  
(B) It slopes up to the right, and the correlation is  $+.57$ .  
(C) It slopes down to the right, and the correlation is  $-.57$ .  
(D) It slopes down to the right, and the correlation is  $+.57$ .  
(E) None of the above is true.
22. Consider the set of points  $\{(2, 5), (3, 7), (4, 9), (5, 12), (10, n)\}$ . What should  $n$  be so that the correlation between the  $x$ - and  $y$ -values is 1?
- (A) 21  
(B) 24  
(C) 25  
(D) A value different from any of the above.  
(E) No value for  $n$  can make  $r = 1$ .
23. Consider the following three scatterplots:



Which of the following is a true statement about the correlations for the three scatterplots?

- (A) None are 0.  
(B) One is 0, one is negative, and one is positive.  
(C) One is 0, and both of the others are positive.  
(D) Two are 0, and the other is 1.  
(E) Two are 0, and the other is close to 1.



24. Consider the three points (2, 11), (3, 17), and (4, 29). Given any straight line, we can calculate the sum of the squares of the three vertical distances from these points to the line. What is the smallest possible value this sum can be?
- (A) 6  
(B) 9  
(C) 29  
(D) 57  
(E) None of these values
25. Suppose that the scatterplot of  $\log X$  and  $\log Y$  shows a strong positive correlation close to 1. Which of the following is true?
- (A) The variables  $X$  and  $Y$  also have a correlation close to 1.  
(B) A scatterplot of the variables  $X$  and  $Y$  shows a strong nonlinear pattern.  
(C) The residual plot of the variables  $X$  and  $Y$  shows a random pattern.  
(D) A scatterplot of  $X$  and  $\log Y$  shows a strong linear pattern.  
(E) A cause-and-effect relationship can be concluded between  $\log X$  and  $\log Y$ .
26. Consider  $n$  pairs of numbers. Suppose  $\bar{x} = 2$ ,  $s_x = 3$ ,  $\bar{y} = 4$ , and  $s_y = 5$ . Of the following, which could be the least squares line?
- (A)  $\hat{y} = -2 + x$   
(B)  $\hat{y} = 2x$   
(C)  $\hat{y} = -2 + 3x$   
(D)  $\hat{y} = \frac{5}{3} - x$   
(E)  $\hat{y} = 6 - x$