

# CHAPTER 5

## BIVARIATE DATA

### PRACTICE MULTIPLE CHOICE (MC)

NAME: \_\_\_\_\_

DATE: \_\_\_\_\_ PERIOD: \_\_\_\_\_

#### DATA ANALYSIS 7

A study is conducted relating AP Statistics exam scores to the total number of study hours for the AP Statistics class put in by students during the academic year, and the correlation is found to be .6. Which of the following are true statements.

- ☒ I. On the average, a 40 percent increase in study time results in a 24 percent increase in exam score.
- ☒ II. Sixty percent of a student's exam score can be explained by the number of study hours.
- ☒ III. Higher exam scores tend to be associated with higher numbers of study hours. **TRUE**

- (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. **E**

#### DATA ANALYSIS 20

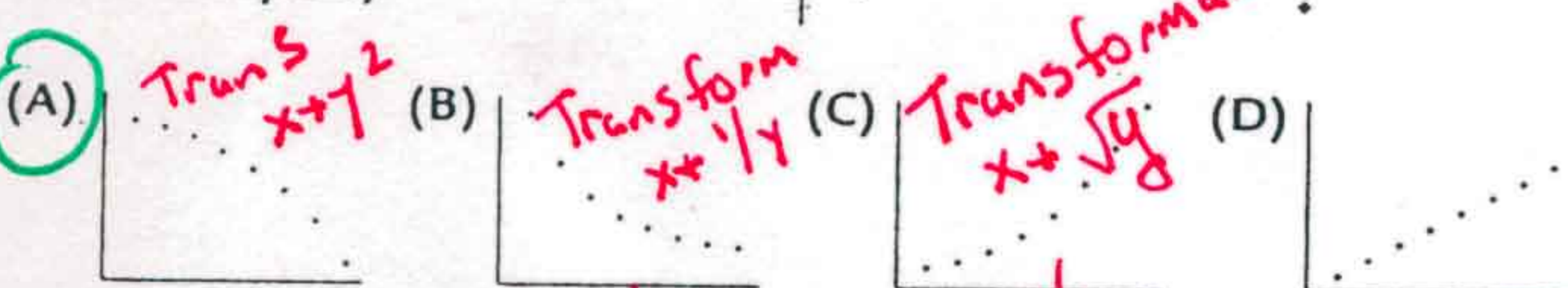
Which of the following statements about residuals are true?

- ☒ I. The mean of the residuals is always zero.
- ☒ II. The regression line for a residual plot is a horizontal line.
- ☒ III. The standard deviation of the residuals gives a measure of how the points in the scatterplot are spread around the 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.

#### DATA ANALYSIS 6

Which of the following scatterplots could have resulted in this 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.

#### DATA ANALYSIS 17

Data are obtained from a random sample of adult women with regard to their ages and their monthly expenditures on health products. The resulting regression equation is:  $\text{Expenditure} = 43 + 0.23(\text{Age})$  with  $r = .27$ . What percentage of the variation in expenditures can be explained by looking at ages?

- (A) 0.23 percent  
(B) 23 percent  
**(C) 7.29 percent**  
(D) 27 percent  
(E) 52.0 percent

$(.27)^2 = .0729$

### DATA ANALYSIS 10

Which of the following are true statements about the correlation coefficient  $r$ ?

- ~~I. A correlation of .3 means that 30 percent of the points are highly correlated.~~
- ~~II. The square of the correlation measures the proportion of the y-variance that is predictable from a knowledge of x.~~
- ~~III. Perfect correlation, that is, when the points lie exactly on a straight line, results in  $r = 0$ .~~

- (A) I only (B) II only (C) III only  
(D) None of these statements is true.  
(E) None of the above gives the complete set of true responses.

### DATA ANALYSIS 27

Suppose the regression line for a set of data,  $\hat{y} = a + 4x$ , passes through the point  $(1, 3)$ . If  $\bar{x}$  and  $\bar{y}$  are the sample means of the x- and y-values, respectively, then  $\bar{y} =$

- (A)  $\bar{x}$   
(B)  $4\bar{x}$   
(C)  $3 + 4\bar{x}$   
(D)  $2 + \bar{x}$   
(E)  $-1 + 4\bar{x}$

$$\begin{aligned} \hat{y} &= a + 4x & b_1 &= 4 \\ \hat{y} &= b_0 + b_1x \\ (1, 3) & \quad 3 = b_0 + 4(1) & b_0 &= \bar{y} - b_1\bar{x} \\ & \quad b_0 = -1 & \bar{y} &= b_0 + b_1\bar{x} \\ & \quad \bar{y} &= -1 + 4\bar{x} \end{aligned}$$

### DATA ANALYSIS 30

Which of the following statements about the correlation  $r$  are true?

- ~~I. The correlation and the slope of the regression line always have the same sign.~~
- ~~II. A correlation of  $-.32$  and a correlation of  $+.32$  show the same degree of clustering around the regression line.~~
- ~~III. A correlation of  $.78$  indicates a relationship that is 3 times as linear as one for which the correlation is  $.26$ .~~

- (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.

### DATA ANALYSIS 77

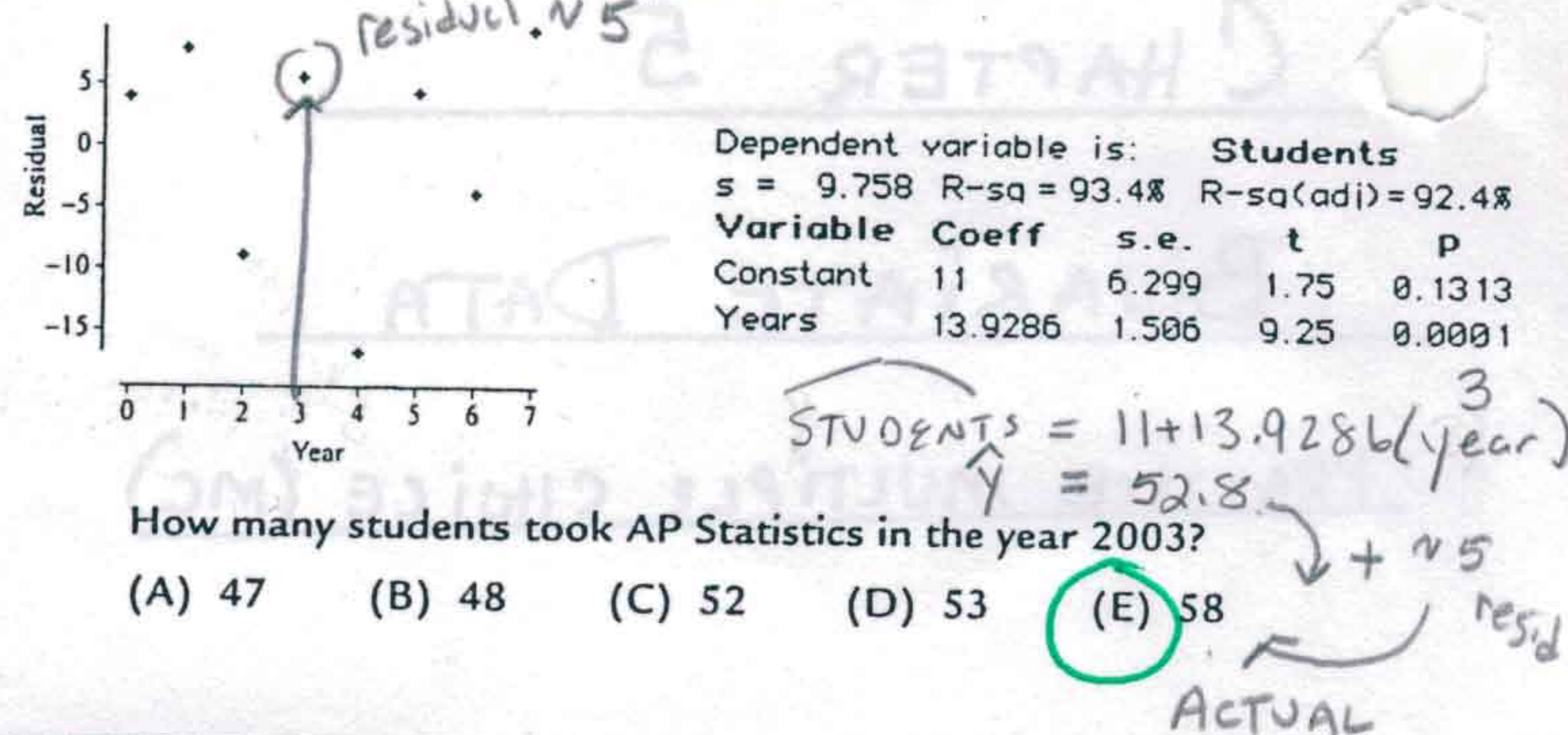
Four pairs of data are used in determining a regression line  $y = -2 + 6x$ . If the four values of the independent variable are 37, 52, 18, and 23, respectively, what is the mean of the four values of the dependent variable?

- (A) 32.5 (B) 193 (C) 195 (D) 778  
(E) The mean cannot be determined from the given information.

x	y
37	220
52	310
18	106
23	136
<hr/>	
772/4 = 193	

### DATA ANALYSIS 26

The number of students taking AP Statistics at a high school during the years 2000–2007 is fitted with a least squares regression line. The graph of the residuals and some computer output is as follows.



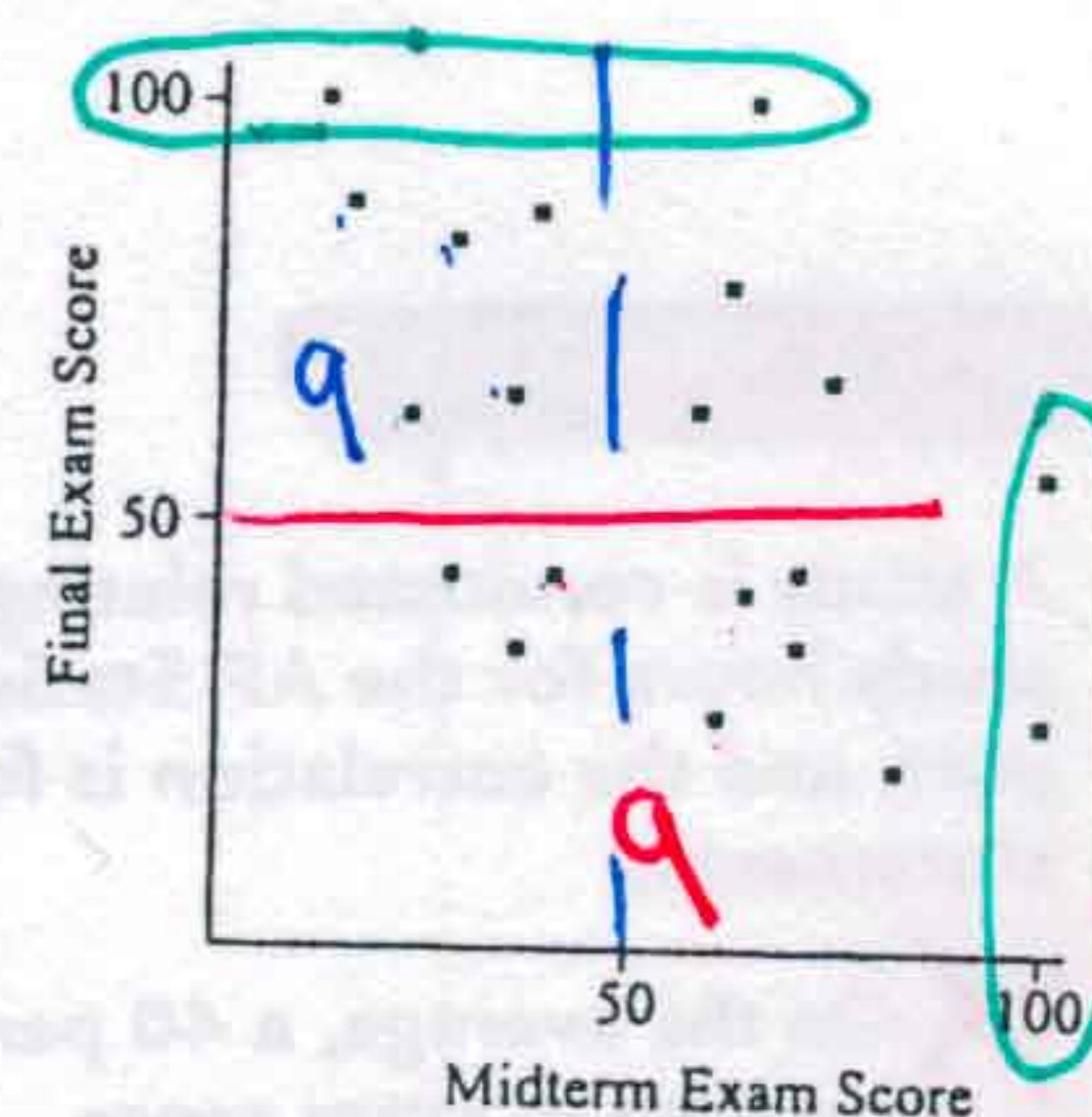
### DATA ANALYSIS 28

Consider the scatterplot of midterm and final exam scores for a class of 20 students.

Which of the following statements are true?

- ~~I. The same number of students scored 100 on the midterm as did on the final.~~
- ~~II. The same percentage of students scored below 50 on each exam.~~
- ~~III. Students who scored higher on the midterm tended to score higher on the final.~~

- (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.



### DATA ANALYSIS 42

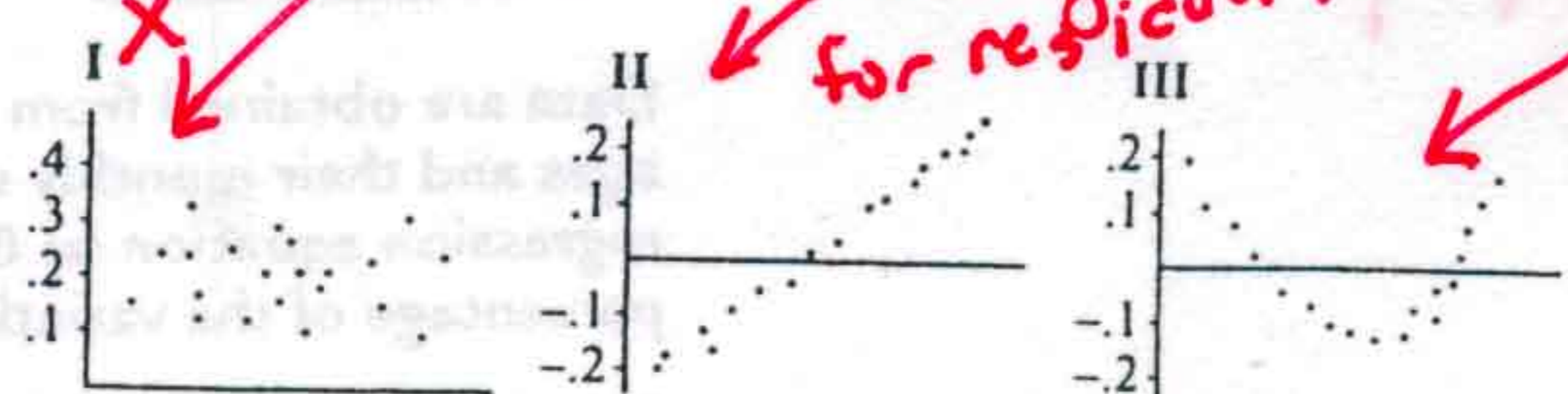
Suppose a study finds that the correlation coefficient relating job satisfaction to salary is  $r = +1$ . Which of the following are proper conclusions?

- ~~I. High salary causes high job satisfaction.~~
- ~~II. Low salary causes low job satisfaction.~~
- ~~III. There is a very strong association between salary and job satisfaction.~~

- (A) I only (B) II only (C) III only (D) I and II (E) I, II, and III

### DATA ANALYSIS 46

Which of the following are possible residual plots?



- (A) I only (B) II only (C) III only (D) I and II (E) I, II, and III

The mean of residuals = 0  
 The regression line horizontal  
 for residual plot is NONLINEAR  
 RESIDUAL PLOT

# DATA ANALYSIS 57

Consider  $n$  pairs of numbers. Suppose  $\bar{x} = 4$ ,  $s_x = 3$ ,  $\bar{y} = 2$ , and  $s_y = 5$ .

Of the following which could be the least squares line?

- (A)  $y = 2 + x$
- (B)  $y = -6 + 2x$
- (C)  $y = -10 + 3x$
- (D)  $y = 5/3 - x$
- (E)  $y = 6 - x$

$$b_1 = r \frac{s_y}{s_x} = \frac{5}{3}r$$

$$-1 \leq r \leq 1$$

$$-5/3 \leq b_1 \leq 5/3$$

$(\bar{x}, \bar{y})$  is on line

$$(4, 2) \quad \hat{y} = 6 - x$$

$$2 = 6 - 4 \checkmark$$

# DATA ANALYSIS 47

A scatterplot of a company's revenues versus time indicates a possible exponential relationship. A linear regression on  $y = \log(\text{revenue in } \$1,000)$  against  $x = \text{years since 2005}$  gives  $\hat{y} = 0.75 + 0.63x$  with  $r = .68$ . Which of the following are valid conclusions?

- I. On the average, revenue goes up 0.63 thousand dollars per year.
  - II. The predicted revenue for year 2009 is approximately 1,862 thousand dollars.
  - III. Forty-six percent of the variation in revenue can be explained by variation in time.
- (A) I only    (B) II only    (C) III only    (D) I and III    (E) None of the above are valid conclusions.

$$\log(\text{revenue in } \$1,000) = .75 + .63(4) = 3.27$$

$$\text{Revenue} = 10^{3.27} = 1,862 \text{ thousand } \$'s$$

# DATA ANALYSIS 60

Which of the following are true statements about the correlation coefficient  $r$ ?

- I. It is not affected by changes in the measurement units of the variables.
  - II. It is not affected by which variable is called  $x$  and which is called  $y$ .
  - III. It is not affected by extreme values.
- (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.

# DATA ANALYSIS 66

To the right is a scatterplot with one point labeled X. Suppose you find the least squares regression line. Which of the following are true statements?

- I. X has the largest residual, in absolute value, of any point on the scatterplot.
  - II. X is an influential point.
  - III. The residual plot will show a curved pattern.
- (A) I only    (B) II only    (C) III only    (D) I and III    (E) II and III



# DATA ANALYSIS 56

A linear regression analysis is performed on the data from two scatterplots, A and B, resulting in identical least squares regression lines with positive slopes. Which of the following statements is true?

- (A) The sum of the squares of the residuals in A equals the sum of the squares of the residuals in B.
- (B) The correlation in A equals the correlation in B.
- (C) If the sum of the squares of the residuals in A is greater than the sum of the squares of the residuals in B, then the correlation in A will be greater than the correlation in B.
- (D) If the sum of the squares of the residuals in A is greater than the sum of the squares of the residuals in B, then the correlation in A will be less than the correlation in B.
- (E) None of the above are true statements.

# DATA ANALYSIS 52

A study of weekly hours of television watched and SAT scores reports a correlation of  $r = -1.18$ . From this information, we can conclude that

- (A) students who watch more TV tend to have lower SAT scores.
- (B) the fewer the hours in front of a TV, the higher a student's SAT scores.
- (C) there is little relationship between weekly hours of television watched and SAT scores.
- (D) there is strong negative association between weekly hours of television watched and SAT scores, but it would be wrong to conclude causation.
- (E) a mistake in arithmetic has been made.

Correlation is between  $-1$  and  $1$

# DATA ANALYSIS 62

Suppose the correlation between two variables is .85. If each of the  $y$ -values 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  $-.85$ .
- (B) It slopes down to the right, and the correlation is  $-.85$ .
- (C) It slopes up to the right, and the correlation is .85.
- (D) It slopes down to the right, and the correlation is .85.
- (E) None of the above is true.



# DATA ANALYSIS 67

Data on the number of cancer deaths among Americans (in 1,000s) and years (since 2001) result in the regression line: Deaths = 550 - 6.05(Years) with  $r = .863$ . What is the correct interpretation of the slope?

- (A) The number of cancer deaths among Americans has been dropping by an average of 6,050 per year since 2001.
- (B) The baseline number of cancer deaths among Americans is 550,000.
- (C) The regression line explains 74.5 percent of the variation in cancer deaths among Americans over the years since 2001.
- (D) The regression line explains 86.3 percent of the variation in cancer deaths among Americans over the years since 2001.
- (E) Cancer will be cured in the year 2092.

# DATA ANALYSIS 70

Which of the following statements about influential points are true?

- I. Looking at a residual plot is an excellent way of picking out influential points.
- II. Removal of an influential point sharply affects the regression line.
- III. Determining a regression model with and without a point is an excellent way of picking out influential points.

- (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.

Influential pts can have small or large residuals

# DATA ANALYSIS 74

Suppose that the scatterplot of log X and log Y shows a strong positive correlation close to 1. Which of the following is true?

- I. The variables X and Y will also have a correlation close to 1.
- II. A scatterplot of the variables X and Y will show a strong nonlinear pattern.
- III. The residual plot of the variables X and Y will show a random pattern.

- (A) I only (B) II only (C) III only (D) I and II (E) I, II, and III

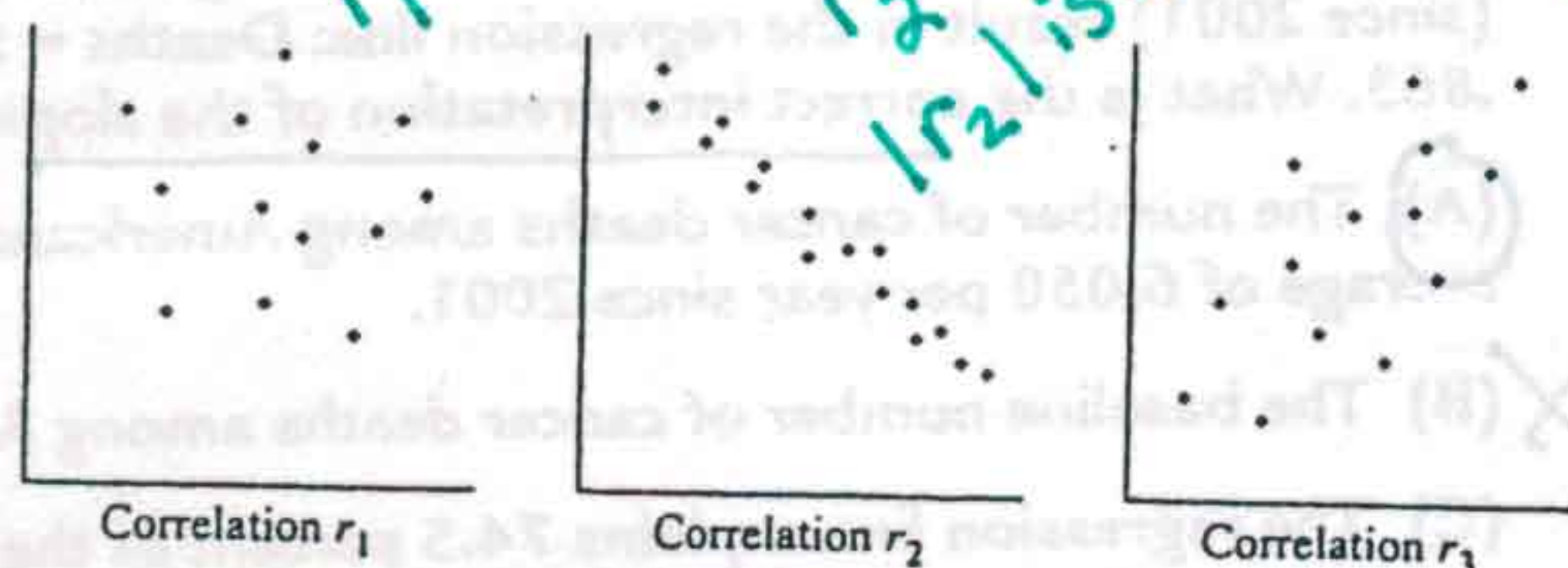
# DATA ANALYSIS 72

Consider the points  $(-1, 4)$ ,  $(2, 10)$ ,  $(4, 15)$ ,  $(7, 21)$ ,  $(10, n)$ . What should  $n$  be so that the correlation between the  $x$  and  $y$  values is  $r = 1$ ?

- (A) 26 (B) 27 (C) 28  
(D) A value different from any of the above.  
(E) No value for  $n$  can make  $r = 1$ .



# DATA ANALYSIS 76



Which of the following gives a proper ordering?

- (A)  $r_1 < r_3 < r_2$  (D)  $r_3 < r_1 < |r_2|$   
(B)  $r_2 < r_3 < r_1$  (E)  $r_1 < r_2 < r_3 < |r_2|$   
(C)  $r_2 < r_1 < r_3 < |r_2|$

neg 0 x 1/7/73