

# DATA ANALYSIS 40

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

- (A) When  $r = 0$ , there is no relationship between the variables. *might be curved*
- (B) When  $r = -2$ , 20 percent of the variables are closely related. *wupc*
- (C) When  $r = 1$ , there is a perfect cause-and-effect relationship between the variables. *never cause effect*
- (D) A correlation close to 1 means that a linear model will give the best fit to the data. *curved data could be close to 1*
- (E) All the statements are false.

# 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.

KEY



# DATA ANALYSIS 20

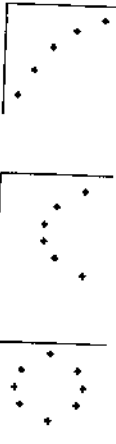
Which of the following statements about residuals is incorrect?

- (A) The mean of the residuals is always zero.
- (B) The sum of the residuals is always zero.
- (C) The regression line for a residual plot is a horizontal line.
- (D) The standard deviation of the residuals gives a measure of how the points in the scatterplot are spread around the regression line.
- (E) A residual equals the predicted  $y$  minus the observed  $y$ .

*other way around*

# DATA ANALYSIS 16

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 negative.
- (D) Two are 0, and the other is  $-1$ .
- (E) Two are 0, and the other is close to  $-1$ .

# DATA ANALYSIS 4

Consider the three points (4, 33), (5, 27), and (6, 15). 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) 2.45
- (B) 6
- (C) 8.66
- (D) 36
- (E) None of these values.

$$(-1)^2 + (2)^2 + (-1)^2 = 1 + 4 + 1 = 6$$

|                     |                |
|---------------------|----------------|
| $\hat{y} = 70 - 9x$ |                |
| plus in             |                |
| $x = 4$             | $\hat{y} = 34$ |
| $x = 5$             | $\hat{y} = 25$ |
| $x = 6$             | $\hat{y} = 16$ |
|                     | residual       |
|                     | -1             |
|                     | 2              |
|                     | -1             |

# DATA ANALYSIS 12

Data on ages (in years) and prices (in \$100) for ten cars of a specific model result in the regression line:  $\text{Price} = 250 - 30(\text{Age})$ . Given that 64 percent of the variation in price is explainable by variation in age, what is the value of the correlation coefficient  $r$ ?

- (A)  $-.64$
- (B)  $-.80$
- (C)  $.64$
- (D)  $.80$
- (E) There is insufficient information to answer this question.

*negative slope*

$$r = \sqrt{.64} = .8 = -.80$$

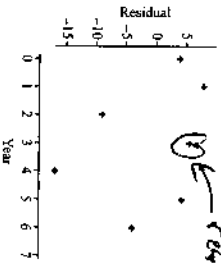
DATA ANALYSIS 60

Which of the following statements about the correlation coefficient  $r$  is *incorrect*?

- (A) It is not affected by changes in the measurement units of the variables.
- (B) It is not affected by which variable is called  $x$  and which is called  $y$ .
- (C) It is not affected by extreme values.
- (D) It gives information about a linear relationship, not about causation.
- (E) It always takes values between  $-1$  and  $1$ , even if the association is nonlinear.

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.



Dependent variable is: Students

| Variable | Coeff   | s.e.  | t    | p      |
|----------|---------|-------|------|--------|
| Constant | 11      | 6.299 | 1.75 | 0.1313 |
| Years    | 13.9286 | 1.506 | 9.25 | 0.0001 |

How many students took AP Statistics in the year 2003?

- (A) 47
- (B) 48
- (C) 52
- (D) 53
- (E) 58

$$\hat{y} = 13.9286(3) + 11 = 52.78 + 5 = 57.78$$

DATA ANALYSIS 37

A simple random sample of 25 world-ranked tennis players provides the following statistics: Number of hours of practice per day:  $\bar{x} = 7.3$ ,  $s_x = 1.2$ . Yearly winnings:  $\bar{y} = \$1,820,000$ ,  $s_y = \$310,000$ . Correlation  $r = .23$ . Based on this data, what is the resulting linear regression equation?

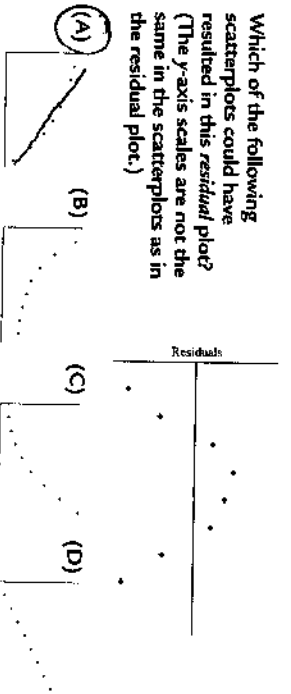
- (A)  $\widehat{\text{Winnings}} = 1,390,000 + 59,400 \text{ hours}$
- (B)  $\widehat{\text{Winnings}} = 1,300,000 + 71,300 \text{ hours}$
- (C)  $\widehat{\text{Winnings}} = -63,400 + 258,000 \text{ hours}$
- (D)  $\widehat{\text{Winnings}} = -443,000 + 310,000 \text{ hours}$
- (E)  $\widehat{\text{Winnings}} = -10,000,000 + 1,620,000 \text{ hours}$

$$m = r \frac{s_y}{s_x}$$

$$b = \bar{y} - m\bar{x}$$

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.)



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

DATA ANALYSIS 32

Suppose the correlation between two variables is  $r = .28$ . What will the new correlation be if  $.17$  is added to all values of the  $x$ -variable, every value of the  $y$ -variable is doubled, and the two variables are interchanged?

- (A) .28
- (B) .45
- (C) .56
- (D) .90
- (E)  $-.28$

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:  $\widehat{\text{Expenditure}} = 43 + 0.23(\text{Age})$  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

$$(.27)^2 = .0729$$