Answer Key

1. **E**	7. **E**	12. **E**	17. **C**	22. **E**
2. **A**	8. **C**	13. **B**	18. **B**	23. **E**
3. **D**	9. **E**	14. **D**	19. **E**	24. **A**
4. **D**	10. **C**	15. **C**	20. **A**	25. **B**
5. **E**	11. **B**	16. **B**	21. **B**	26. **E**
6. **A**				

Answers Explained

Multiple-Choice

1. **(E)** The variable column indicates the independent (explanatory) variable. The sign of the correlation is the same as the sign of the slope (negative here). In this example, the y-intercept is meaningless (predicted SAT result if no students take the exam). There can be a strong linear relation, with high R^2 value, but still a distinct pattern in the residual plot indicating that a non-linear fit may be even stronger. The negative value of the slope (–2.84276) gives that the predicted combined SAT score of a school is 2.84 points lower for each one unit higher in the percentage of students taking the exam, on average.

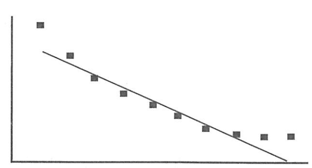
2. **(A)** Slope = $.15(\frac{42,000}{1.3}) \approx 4850$ and intercept = $208,000 - 4850(6.2) \approx 178,000$.

3. **(D)** Residual = Measured – Predicted, so if the residual is negative, the predicted must be greater than the measured (observed).

4. **(D)** The correlation coefficient is not changed by adding the same number to each value of one of the variables or by multiplying each value of one of the variables by the same positive number.

5. **(E)** A negative correlation shows a tendency for higher values of one variable to be associated with lower values of the other; however, given any two points, anything is possible.

6. **(A)** This is the only scatterplot in which the residuals go from positive to negative and back to positive.



7. **(E)** Since (2, 5) is on the line $y = 3x + b$, we have $5 = 6 + b$ and $b = -1$. Thus the regression line is $y = 3x - 1$. The point $(\bar{x}, \bar{y})$ is always on the regression line, and so we have $\bar{y} = 3\bar{x} - 1$.

8. **(C)** The correlation r measures association, not causation.

9. **(E)** The correlation r cannot take a value greater than 1.

10. **(C)** If the points lie on a straight line, $r = \pm 1$. Correlation has the formula $r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$ so x and y are interchangeable, and r does not depend on which variable is called x or y. However, since means and standard deviations can be strongly influenced by outliers, r too can be strongly affected by extreme values. While $r = .75$ indicates a better fit with a linear model than $r = .25$ does, we cannot say that the linearity is threefold.

11. **(B)** The "Predictor" column indicates the independent variable with its coefficient to the right.

12. **(E)** $r = \sqrt{.986} = .993$

13. **(B)** $\widehat{back} = 0.056 + 0.920\,(0.55) = 0.562$ and so the *residual* $= 0.59 - 0.562 = 0.028$

14. **(D)** The sum and thus the mean of the residuals are always zero. In a good straight-line fit, the residuals show a random pattern.

15. **(C)** The coefficient of determination r^2 gives the proportion of the y-variance that is predictable from a knowledge of x. In this case $r^2 = (.632)^2 = .399$ or 39.9%.

16. **(B)** The point I doesn't contribute to a line with negative or positive slope. In none of the scatterplots do the points fall on a straight line, so none of them have correlation 1.0.

17. **(C)** Predicted winning percentage = 44 + 0.0003(34,000) = 54.2, and

    Residual = Observed – Predicted = 55 – 54.2 = 0.8.

18. **(B)** On each exam, two students had scores of 100. There is a general negative slope to the data showing a moderate negative correlation. The coefficient of determination, r^2, is always ≥ 0. While several students scored 90 or above on one or the other exam, no student did so on both exams.

19. **(E)** On the scatterplot all the points lie perfectly on a line sloping up to the right, and so $r = 1$.

20. **(A)** The correlation is not changed by adding the same number to every value of one of the variables, by multiplying every value of one of the variables by the same positive number, or by interchanging the x- and y-variables.

21. **(B)** The slope and the correlation coefficient have the same sign. Multiplying every y-value by –1 changes this sign.

22. **(E)** A scatterplot readily shows that while the first three points lie on a straight line, the fourth point does not lie on this line. Thus no matter what the fifth point is, all the points cannot lie on a straight line, and so r cannot be 1.

23. **(E)** All three scatterplots show very strong nonlinear patterns; however, the correlation r measures the strength of only a linear association. Thus $r = 0$ in the first two scatterplots and is close to 1 in the third.

24. **(A)** Using your calculator, find the regression line to be $\hat{y} = 9x - 8$. The regression line, also called the least squares regression line, minimizes the sum of the squares of the vertical distances between the points and the line. In this case (2, 10), (3, 19), and (4, 28) are on the line, and so the minimum sum is $(10-11)^2 + (19-17)^2 + (28-29)^2 = 6$.

25. **(B)** When transforming the variables leads to a linear relationship, the original variables have a nonlinear relationship, their correlation (which measures linearity) is not close to 1, and the residuals do not show a random pattern. While r close to 1 indicates strong association, it does not indicate cause and effect.

26. **(E)** The least squares line passes through $(\bar{x}, \bar{y}) = (2, 4)$, and the slope b satisfies $b = r\frac{s_y}{s_x} = \frac{5r}{3}$. Since $-1 \le r \le 1$, we have $-\frac{5}{3} \le b \le \frac{5}{3}$.

Free-Response

1. (a) A calculator gives $\widehat{\text{Attendance}} = 12{,}416 + 180.4\,(\text{Wins})$.
   (b) Each additional home win raises the average attendance by about 180 people, on average.
   (c) 12,416 + 180.4(25) = 16,926
   (d) 17,000 = 12,416 + 180.4(Wins) gives Wins = 25.4 so 26 wins needed to average at least 17,000 average attendance.
   (e) With 34 wins, the predicted average attendance is 12,416 + 180.4(34) = 18,550 so the residual is 18,997 – 18,550 = 447.

2. (a)