Key to Linear Regression Multiple Choice

1. C — Distractors: B is r^2; D is outlier; E is correlation coefficient r
2. B — $\hat{y} = -2.29 + (1.70)(5) = 6.21$; residual $= 6 - 6.21 = -0.21$
3. C — Tall fathers generally have tall sons, so r should be positive. But r cannot be over 1, and r will not be exactly 1 unless every data point falls on the best fit line, which should not be true for this data.
4. A — slope $= r\frac{s_y}{s_x} = (.63)\frac{1.9}{2.5} = .4788$
5. E — Distractors: A and C switch slope and intercept; B treats slope as a % instead of the ratio of \$1000/year
6. D — Distractors: A and E include categorical variables instead of numerical variables. In C, r cannot be over 1. In B, r should not have units.
7. E — $\ln \hat{h} = -1.2 + 1.4\ln(60) = 4.532$; $\hat{h} = e^{4.532} = 92.95$ ft
8. C
9. A — $r = \pm\sqrt{.64} = \pm 0.80$; Without the graph the direction of the association cannot be determined.
10. B — Too many of the residuals are positive; since residuals are $y - \hat{y}$, that means the actual values are larger than the predicted values.

Rubric for Linear Regression Free Response

1. **Solution**

Part (a):

The slope is 325.39 magazines per year. For each year since 1988, the number of magazines published in the US increases by about 325, on average.

Part (b):

The y-intercept is 13549.9 magazines. The predicted number of magazines published in the US in 1988 (year 0) is 13550 magazines.

Part (c):

1999 is year 11 (because 1999-1988 = 11).

$\text{magazines} = 13549.9 + 325.39\ \text{year} = 13549.9 + (325.39)(11) = 17129$

We predict that there were 17129 magazines published in the US in 1999.

Part (d):

Since the slope is positive, the correlation coefficient is the positive square root of 0.848: $r = +\sqrt{0.848} = 0.921$

Since the correlation coefficient is +0.921, there is a strong, positive linear relationship between the number of magazines published in the US and the year.

Scoring

All parts can be essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is correct if 1) the numerical value is correct, 2) correct units are given for the slope, 3) the interpretation is correct and in context, and 4) the interpretation distinguishes between the model and the data by using words like about, approximately, or on average. The slope value may be rounded in the interpretation.

Part (a) is partially correct if the student correctly does 2 or 3 of the items listed above.

Part (a) is incorrect if the student correctly does 0 or 1 of the items listed above.

Part (b) is correct if 1) the numerical value is correct, 2) correct units are given for the y-intercept, 3) the interpretation is correct and in context, and 4) the interpretation distinguishes between the model and the data by using words like about, approximately, or predicted. The y-intercept value may be rounded in the interpretation.

Part (b) is partially correct if the student correctly does 2 or 3 of the items listed above.

Part (b) is incorrect if the student correctly does 0 or 1 of the items listed above.

Part (c) is essentially correct if the student identifies 1999 as year 11 and uses that value in a correct regression equation to find the number of magazines in 1999.

Part (c) is partially correct if the student correctly identifies 1999 as year 11 but doesn't use that value in a correct regression equation

OR

Has a correct regression equation but uses the wrong year

Part (d) is essentially correct if the value for r is correct and the interpretation is correct and in context.

Part (d) is partially correct if only one of the value for r or the interpretation in context is correct.

Each essentially correct response is worth 1 point; each partially correct response is worth half a point.

4 Complete Response

3 Substantial Response

2 Developing Response

1 Minimal Response

If a response is between two scores (for example, 2.5 points), use a holistic approach to determine whether to score up or down depending on the strength of the response and communication.

2. **Solution**

Part (a):

Yes, a linear model is appropriate. The residual plot shows no pattern and a test for slope shows that there is a relationship ($H_0 : \beta_1 = 0$; $H_a : \beta_1 > 0$; where β_1 is the slope of the weight vs. height graph, $df = 4$, $t = 8.987$; p-value $= 0.0004$).

Part (b):

$\hat{y} = 3.6956x - 13.430$ where x = height in inches and $\hat{y}$ = predicted height in pounds

Part (c):

Dakota's predicted weight is $\hat{y} = (3.6956)(23.5) - 13.430 = 73.4$ pounds.

Dakota's residual (read from the graph) is approximately 1.55 to 1.6.

$$\text{residual} = y - \hat{y}$$

$$1.6 = y - 73.4$$

$$y = 73.4 + 1.6 = 75 \text{ pounds}$$

Dakota's actual weight is 75 pounds.

Scoring

Part (a) can be essentially correct (E) or incorrect (I). Parts (b) and (c) can be essentially correct (E), partially correct (P), or incorrect (I).

Part (a) can be essentially correct even if it fails to mention the linear regression t test as long as the residual graph is discussed.

Part (a) is incorrect if the only evidence for linearity given is the value of the correlation coefficient, $r = 0.976$.

Part (b) is essentially correct if the correct numerical values of both the slope and y-intercept are present in the equation **and** both variables are identified. Note: variable names (height and weight) may be used in the equation in place of x and y for full credit.

Part (b) is partially correct if the correct numerical values of both the slope and y-intercept are present in the equation, but the variables are not identified

OR

both variables are identified, but the numerical values of the slope and y-intercept are incorrect

OR

Only one numerical value is correct and only one variable is identified.

Part (c) is essentially correct if 1) the predicted weight is correct, 2) an appropriate residual (between 1.5 and 1.7) is read from the graph, and 3) the actual weight is computed correctly with work shown. The weights must be distinguished by labels (actual, predicted) or symbols (y for actual weight, $\hat{y}$ for predicted weight).

Part (c) is partially correct if two of the three tasks are completed correctly. The computation of the actual weight can be considered to be correct if an incorrect residual is substituted correctly into the residual formula.

Pact (c) is incorrect if zero or one of the tasks is completed correctly.

4 Complete Response

All parts essentially correct.

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and two parts partially correct

1 Minimal Response

One part essentially correct and either zero or one part partially correct

OR

No parts essentially correct and two parts partially correct