AP Stat Workday on Chapter 3 & 4.1 Name Key

Multiple Choice Questions on Linear Regression

C 1. Residuals are
(A) possible models not explored by the researcher.
(B) variation in the response variable that is explained by the model. r^2
(C) the difference between the observed response and the values predicted by the model. $y - \hat{y}$
(D) data collected from individuals that is not consistent with the rest of the group. outliers
(E) a measure of the strength of the linear relationship between x and y r

B 2. Data was collected on two variables x and y and a least squares regression line was fitted to the data. The resulting equation is $\hat{y} = -2.29 + 1.70x$. What is the residual for point (5, 6)? $\hat{y} = -2.29 + 1.70(5) = 6.21$
(A) −2.91 (B) −0.21 (C) 0.21 (D) 6.21 (E) 7.91 Resid = 6 − 6.21 = −0.21

C 3. Child development researchers studying growth patterns of children collect data on the heights of fathers and sons. The correlation between the fathers' heights and the heights of their 16-year-old sons is most likely to be Father ↑ Son ↑ = + Father ↓ Son ↓ = +
(A) near −1.0 (B) near 0 (C) near +0.7 (D) exactly +1.0 (E) somewhat greater than +1.0 Not possible

A 4. Given a set of ordered pairs (x, y) with $s_x = 2.5$, $s_y = 1.9$, $r = 0.63$, what is the slope of the regression line of y on x? $b = 0.63\left(\frac{1.9}{2.5}\right) = .4788$
(A) 0.48 (B) 0.65 (C) 1.32 (D) 1.90 (E) 2.63

E 5. The relation between the selling price of a car (in $1,000) and its age (in years) is estimated from a random sample of cars of a specific model. The relation is given by the following formula:

$$pred(\text{SellingPrice}) = 24.2 - 1.182(\text{Age})$$

Which of the following can be concluded from this equation?
(A) For every year the car gets older, the selling price drops by approximately $2420. False
(B) For every year the car gets older, the selling price goes down by approximately 11.82 percent. False
(C) On average, a new car costs about $11,820. False
(D) On average, a new car costs about $23,018. False 24.2 − 1.182 ??
(E) For every year the car gets older, the selling price drops by approximately $1182. True

D 6. All but one of these statements is false. Which one could be **true**?
(A) The correlation between a football player's weight and the position he plays is 0.54. → Qualitative
(B) The correlation between a car's length and its fuel efficiency is 0.71 miles per gallon. No units
(C) There is a high correlation (1.09) between height of a corn stalk and its age in weeks. Not possible
(D) The correlation between the amounts of fertilizer used and quantity of beans harvested is 0.42.
(E) There is a correlation of 0.63 between gender and political party. Qualitative

C 7. Which is true?
I. Random scatter in the residuals indicates a linear model.
II. If two variables are very strongly associated, then the correlation between them will be near +1.0 or −1.0.
III. Changing the units of measurement for x or y changes the correlation coefficient. False

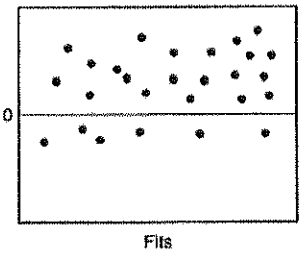
(A) I only (B) II only (C) I and II only (D) II and III only (E) I, II, and II

E 8. If the coefficient of determination r^2 is calculated as 0.49, then the correlation coefficient

(A) is 0.7599 (B) is − 0.70 (C) is 0.2401
(D) is 0.70 (E) cannot be determined without the data

B 9. Which of the following is a correct conclusion based on the residual plot displayed?

(A) The line overestimates the data.
(B) The line underestimates the data.
(C) It is not appropriate to fit a line to these data since there is clearly no correlation.
(D) The data are not related.
(E) There is a nonlinear relationship between the variables.



E 10. It is easy to measure the circumference of a tree's trunk, but not so easy to measure its height. Foresters developed a model for ponderosa pines that they use to predict tree's height (in feet) from the circumference of its trunk (in inches): $\ln \hat{h} = -1.2 + 1.4(\ln C)$ A lumberjack finds a tree with a circumference of 60 inches, how tall does this model estimate the tree to be?

(A) 5 ft (B) 11 ft (C) 19 ft (D) 83 ft (E) 93 ft

Free Response Questions on Linear Regression: Answer on your own notebook paper.

1. The National Directory of Magazines tracks the number of magazines published in the United States each year. An analysis of data from 1988 to 2007 gives the following computer output. The dates were recorded as years since 1988. Thus, the year 1988 was recorded as year 0. A residual plot (not shown) showed no pattern.

Predictor	Coef	StDev	T	P
Constant	13549.9	2.731	7.79	0.000
Years	325.39	0.1950	10.0	0.000
S = 836.2	R-Sq = 84.8%	R-Sq (adj) = 80.6%		

(a) What is the value of the slope of the least squares regression line? Interpret the slope in the context of this situation.

(b) What is the value of the y-intercept of the least squares regression line? Interpret the y-intercept in the context of this situation.

(c) Predict the number of magazines published in the United States in 1999.

(d) What is the value of the correlation coefficient for number of magazines published in the US and years since 1988? Interpret this correlation.

(2) a) The correlation coefficient of 0.976 indicates a strong, positive linear association between height and weight of Labrador Retrievers.
The coefficient of determination of 0.953 indicates this linear model is a good fit for our data.
The residual plot shows no pattern indicating a linear model would be appropriate for this data.

b) $\text{pred}\left(\begin{array}{c}\text{weight in pounds of} \\ \text{Labrador Retrievers}\end{array}\right) = -13.430 + 3.6956\left(\begin{array}{c}\text{height in} \\ \text{inches of} \\ \text{Labrador Retrievers}\end{array}\right)$

c) Dakota's residual is where height = 23.5.
This residual is approximately 1.6.

$$\text{Resid} = y - \hat{y}$$
$$\text{pred}(\text{weight}) = -13.430 + 3.6956(23.5)$$
$$= 73.42$$

$$1.6 = y - 73.42$$
$$75.02 = y$$

Dakota's predicted weight is 73.42 lbs.
Dakota's actual weight is 75.02 lbs.

[Doesn't ask you to discuss model. If it did, you would say "This model underestimates the weight of a Lab weighing 75 lbs."]

(3) a) Using Model 1:

$$\text{pred(resale price)} = 29784 - 343.58(35)$$
$$= 17758.70$$

The predicted resale price of a car of this make and model which has been driven 35,000 miles is $17,758.70.

b) Model 1 is not appropriate because the scatterplot shows a non-linear trend and the residual plot shows a pattern.
Model 3 is best because the scatterplot looks the straightest (most linear) and the residual plot shows less of a pattern. Model 2 has a curved scatterplot and more of a pattern in the residual plots.

c) Using Model 3:

$$\text{pred(resale price)} = 43254 - 17153 \log(35)$$
$$= 16768.60$$

The predicted resale price of a car of this make and model which has been driven 35,000 miles is $16,768.60.