AP Stat Workday on Chapter 3 & 4.1 Name _____

Multiple Choice Questions on Linear Regression

_____1. Residuals are
(A) possible models not explored by the researcher.
(B) variation in the response variable that is explained by the model.
(C) the difference between the observed response and the values predicted by the model.
(D) data collected from individuals that is not consistent with the rest of the group.
(E) a measure of the strength of the linear relationship between x and y

_____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)?
(A) −2.91 (B) −0.21 (C) 0.21 (D) 6.21 (E) 7.91

_____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
(A) near −1.0 (B) near 0 (C) near +0.7 (D) exactly +1.0 (E) somewhat greater than +1.0

_____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?
(A) 0.48 (B) 0.65 (C) 1.32 (D) 1.90 (E) 2.63

_____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.
(B) For every year the car gets older, the selling price goes down by approximately 11.82 percent.
(C) On average, a new car costs about \$11,820.
(D) On average, a new car costs about \$23,018.
(E) For every year the car gets older, the selling price drops by approximately \$1182.

_____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.
(B) The correlation between a car's length and its fuel efficiency is 0.71 miles per gallon.
(C) There is a high correlation (1.09) between height of a corn stalk and its age in weeks.
(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.

_____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.

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

_____ 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

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

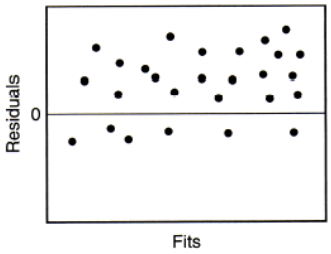
(A) The line overestimates the data.

(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.



_____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.

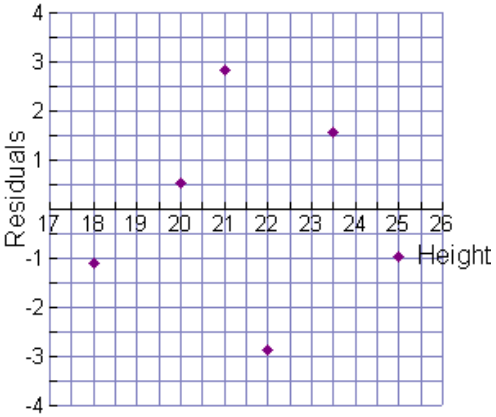
(b) What is the value of the *y*-intercept of the least squares regression line? Interpret the *y*-intercept in the context of this situation.

(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. The heights (in inches) and weights (in pounds) of six male Labrador Retrievers were measured. The height of a dog is measured at the shoulder. A linear regression analysis was done, and the residual plot and computer output are given below.

Predictor	Coef	StDev	T	P
Constant	−13.430	1.724	7.792	0.0000
Height	3.6956	0.4112	8.987	0.0004
S = 2.297	R-Sq = 95.3%	R-Sq (adj) = 90.6%		



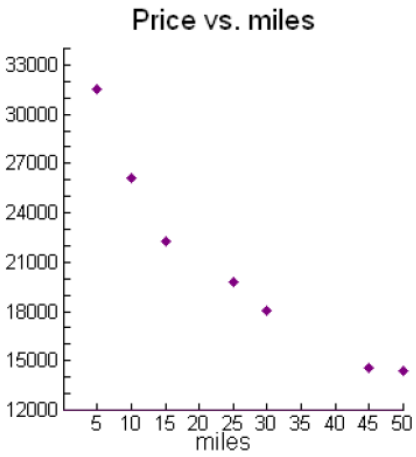
(a) Is a line an appropriate model to use for these data? What information tells you this?

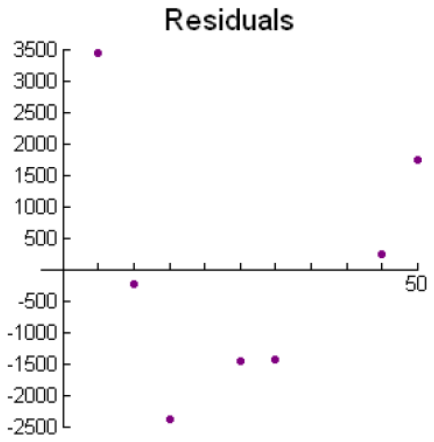
(b) Write the equation of the least squares regression line. Identify any variables used in this equation.

(c) Dakota, a male Labrador, was one of the dogs measured for this study. His height is 23.5 inches. Find Dakota's predicted weight **and** Dakota's actual weight.

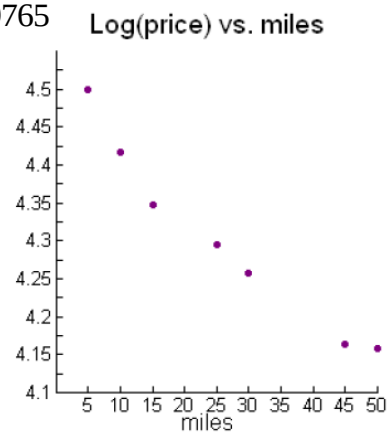
3. As more miles are driven in a car, the resale value of the car generally declines. This is called depreciation. For a certain make and model of car, information is gathered on the resale price in dollars and the number of miles driven (in thousands of miles). The scatterplot of price (y) versus miles (x), the residual plot, and the least squares regression line is shown for this data. In addition, the scatterplot, residual plot, and the accompanying best fit lines are shown for two other models using the common logarithm.

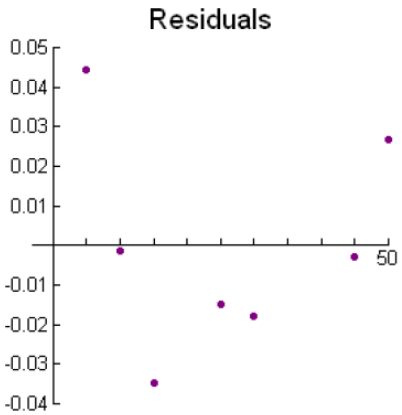
Model 1: $\hat{y} = 29784 - 343.58x \quad r = -0.9452$



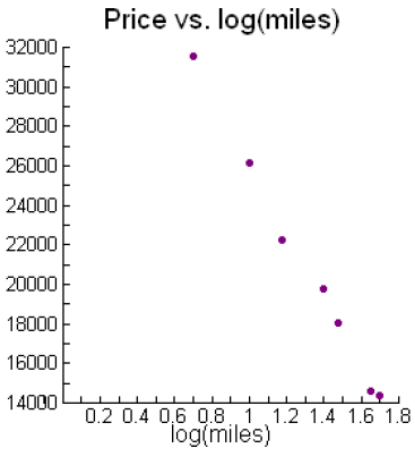


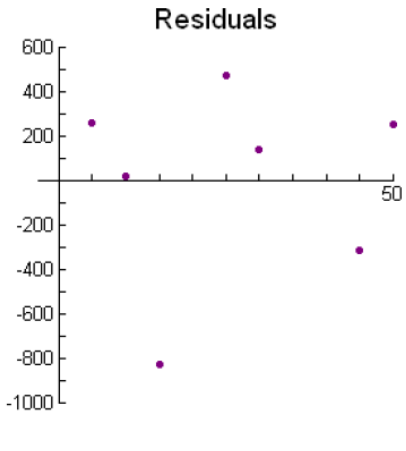
Model 2: $\log \hat{y} = 4.4901 - 0.0071910x \quad r = -0.9765$





Model 3: $\hat{y} = 43254 - 17153 \log x \quad r = -0.9975$





(a) Using Model 1, estimate a resale price for a car of this make and model which has been driven 35,000 miles.

(b) Model 1 is not the most appropriate to use to compute an estimated resale price. Explain why it is not appropriate, and determine whether Model 2 or Model 3 is better.

(c) Use the model you chose in part (b) to estimate a resale price for a car of this make and model that has been driven 35,000 miles.