AP Statistics
Unit 02 REDO – HW #1

Name Key 2018
Period______

Multiple Choice

1. In a statistics class, a linear regression equation was computed to predict the final exam score from the score on the first test. The equation was $\hat{y} = 10 + 0.9x$ where y is the final exam score and x is the score on the first test. Carla scored 95 on the first test. What is the predicted value of her score on the final exam?
   (a) 95
   (b) 85.5
   (c) 90
   (d) 95.5
   (e) None of the above

D

$\hat{y} = 10 + 0.9(95) = 95.5$

2. Which of the following are true statements about the correlation coefficient r?
   ~~I.~~ A correlation of 0.3 means that 30 percent of the points are highly correlated.
   II. The square of the correlation measures the proportion of the y-variance that is predictable from a knowledge of the regression of y on x.
   ~~III.~~ Perfect correlation, that is, when the points lie exactly on a straight line, results in r = 0.
   (a) I only
   (b) II only
   (c) III only
   (d) None of these statements are true
   (e) None of the above gives a complete set of true responses

B

3. In regression, the residuals are which of the following?
   (a) Those factors unexplained by the data
   (b) The difference between the observed responses & the values predicted by the regression line
   (c) Those data points, which were recorded after the formal investigation was completed
   (d) Possible models unexplored by the investigator
   (e) None of the above

B

4. What does the square of the correlation (r^2) measure?
   (a) The slope of the least squares regression line
   (b) The intercept of the least squares regression line
   (c) The extent to which cause and effect is present in the data
   (d) The fraction of the variation in the values of *y* that is explained by least-squares regression of y on x
   (e) The fraction of the variation in the values of *x* that is explained by least-squares regression of y on x

D

5. Which of the following statements about correlation r are true?
   I. When r = 0, there is no relationship between the variables
   ~~II.~~ When r = .2, 20 percent of the variables are closely related
   ~~III.~~ When r = 1, there is a perfect cause-and-effect relationship between the variables
   (a) I only
   (b) II only
   (c) III only
   (d) I, II, III
   (e) All the statements are false

A

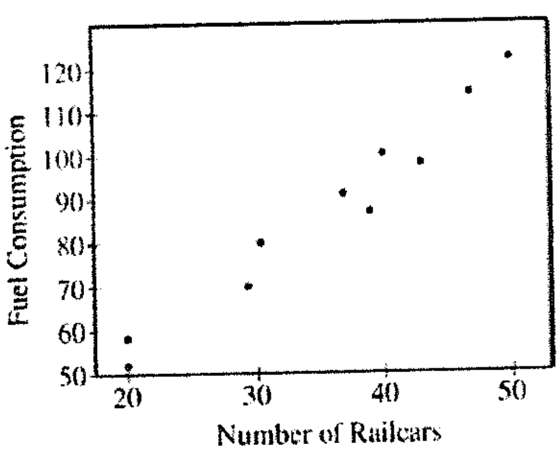
Free Response:

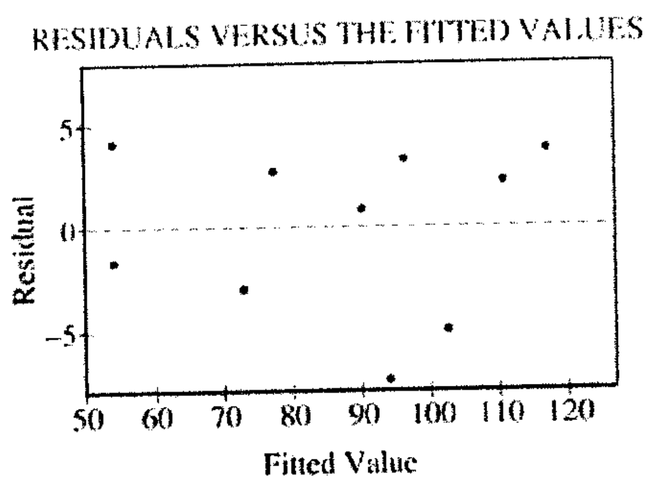
1. The Great Plains Railroad is interested in studying how fuel consumption is related to the number of railcars for its trains on a certain route between Oklahoma City and Omaha.

   A random sample of 10 trains on this route has yielded the data in the table below.

x Number of Railcars	y Fuel Consumption (units/mile)
20	58
20	52
37	91
31	80
47	114
43	98
39	87
50	122
40	100
29	70

A scatterplot, a residual plot, and the output from the regression analysis for these data are shown below.





The regression equation is:

$$\widehat{\text{Fuel Consumption}} = 10.7 + 2.15 \text{ Railcars}$$

And the value of R^2 is 96.7%

a. Is a linear model appropriate for modeling these data? Clearly explain your reasoning. Yes, the linear model is appropriate for these data. The scatterplot shows a strong, positive, linear association between the number of railcars and fuel consumption, and the residual plot shows a reasonably random scatter of points above and below zero.

b. Suppose the fuel consumption cost is $25 per unit. Give a point estimate (single value) for the change in the average cost of fuel per mile for each additional railcar attached to a train. Show your work.

The change in fuel consumption (per mile) for each additional railcar is represented by the slope: 2.15 units of fuel per railcar per mile. If each unit of fuel costs $25, then each additional railcar costs $25·(2.15) = $53.75 per mile.

c. Calculate the residual value for a train with 37 railcars. Show all work.

residual = $y - \hat{y}$ $y = 91$ units/mile

$\hat{y} = 10.7 + 2.15(37) = 90.25$ units/mile

residual = $91 - 90.25 = 0.75$

The predicted fuel consumption for a train with 37 railcars is 0.75 units/mile less than the actual fuel consumption.

d. Interpret the value of R^2 in the context of this problem.

96.7% of the variation in the fuel consumption values can be explained by the LSRL of fuel consumption on number of railcars.

e. Would it be reasonable to use the fitted regression equation to predict the fuel consumption for a train on this route if the train had 65 railcars? Explain.

NO, because our model was created based on data gathered from trains with up to 50 railcars but not more. A train with 65 cars is outside of the domain of values used to create the model, so this is extrapolation, which is not reasonable!