

Name: _____

Date: _____

AP Statistics Assignment #3.4 Extra Practice Questions on Regression Models

1. The gas mileage of an automobile first increases and then decreases as the speed increases. Suppose that this relationship is very regular, as shown by the following data on speed (miles per hour) and mileage (miles per gallon):

| | | | | | |
|---------|----|----|----|----|----|
| Speed | 20 | 30 | 40 | 50 | 60 |
| Mileage | 25 | 29 | 31 | 28 | 24 |

- Make a scatterplot of mileage versus speed
 - Show that the correlation between speed and mileage is $r=0$. Explain why the correlation is 0 even though there is a strong relationship between speed and mileage?
2. Data on the IQ test scores and reading test scores for a group of fifth-grade children give the regression line $[\text{Reading score} = -33.4 + 0.882 (\text{IQ score})]$ for predicting reading score from IQ score
- Explain what the slope of this line tells you
 - Find the predicted reading scores for two children with IQ scores of 90 and 130, respectively
 - Draw a graph of the regression line for IQ's between 90 and 130.
 - Interpret the y-intercept of this line. Why doesn't this make any sense?
3. Ecologists sometimes find rather strange relationships in our environment, where one study suggests that beavers benefit beetles. Researchers laid out 23 circular plots, each four meters in diameter, in an area where beavers were cutting down cottonwood trees. In each plot, they counted the number of stumps from trees cut by beavers and the number of clusters of beetle larvae.

| | | | | | | | | | | | | | | | | | | | | | | | |
|---------------|----|----|----|----|----|----|----|----|----|----|----|----|----|---|----|----|----|---|----|---|----|----|----|
| Stumps | 2 | 2 | 1 | 3 | 3 | 4 | 3 | 1 | 2 | 5 | 1 | 3 | 2 | 1 | 2 | 2 | 1 | 1 | 4 | 1 | 2 | 1 | 4 |
| Beetle Larvae | 10 | 30 | 12 | 24 | 36 | 40 | 43 | 11 | 27 | 56 | 18 | 40 | 25 | 8 | 21 | 14 | 16 | 6 | 54 | 9 | 13 | 14 | 50 |

- Make a scatterplot that shows how the number of beaver-caused stumps influences the number of beetle larvae clusters. What does your plot show?
- Find the least-squares regression line and draw it on your plot
- Construct a residual plot. How well does the linear model fit the data?
- Interpret the r^2 -value in the context of this problem

4. There is a linear relationship between the number of chirps made by the striped ground cricket and the air temperature. A least squares fit of some data collected gives the model:

$\hat{y} = 25.2 + 3.3x$ for $9 < x < 25$ where 'x' is the number of chirps per minute and \hat{y} is the estimated temperature in degrees Fahrenheit. What is the estimated increase in temperature that corresponds to an increase of 5 chirps per minute?

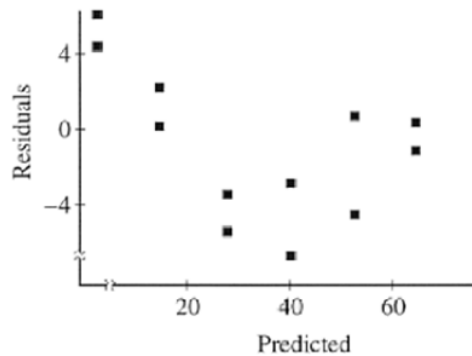
5. In a study of the application of a certain type of weed killer, 14 fields containing large number of weeds were treated. The weed killer was prepared at seven different strengths by adding 1, 1.5, 2, 2.5, 3, 3.5, or 4 teaspoons to a gallon of water. Two randomly selected fields were treated with each strength of weed killer. After a few days, the percentage of weeds killed on each field was measured. The computer output obtained from fitting a least squares regression line to the data is shown below. A plot of the residuals is provided as well:

Dependent variable is: percent killed

R squared = 97.2% R squared (adjusted) = 96.9%

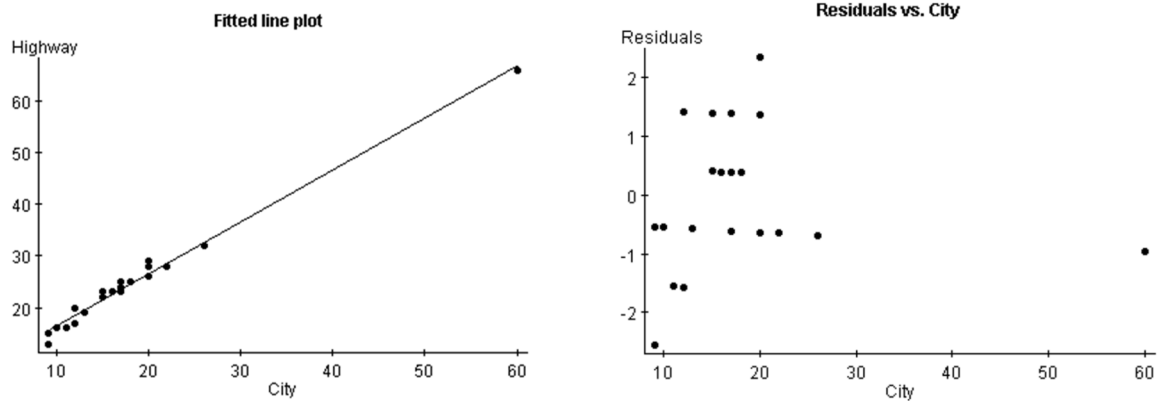
s = 4.505 with 14 - 2 = 12 degrees of freedom

| Source | Sum of Squares | df | Mean Square | F-ratio |
|---------------|----------------|---------------|-------------|---------------|
| Regression | 8330.16 | 1 | 8330.16 | 410 |
| Residual | 243.589 | 12 | 20.2990 | |
| Variable | Coefficient | s.e. of Coeff | t-ratio | Prob |
| Constant | -20.5893 | 3.242 | -6.35 | ≤ 0.0001 |
| No. Teaspoons | 24.3929 | 1.204 | 20.3 | ≤ 0.0001 |



- a) What is the equation of the least squares regression line given by this analysis? Define any variables used in this equation
- b) If someone uses this equation to predict the percentage of weeds killed when 2.6 teaspoons of weed killer are used, which of the following would you expect? i) the prediction will be too large ii) the prediction will be too small iii) a prediction cannot be made based on the information given on the computer output. Explain your reasoning.

6. Data on city and highway gas mileages for 21 two-seater cars, including the Honda Insight gas-electric hybrid car, were collected. The Honda Insight got 60 mpg in the city and 66 mpg on the highway. Least-squares regression was performed on the data. A scatterplot displaying the least-squares line and a residual plot are shown below.



- (a) The Honda Insight is an outlier but does not have the largest residual. Explain why not.
- (b) If the Honda Insight were removed from this set of data, would the correlation increase, decrease, or stay the same? Justify your answer.
- (c) Is the Honda Insight influential on the slope of the regression line? Justify your answer.

7. There is a strong positive association between workers' education and their income. For example, the Census Bureau reports that the median income of young adults (ages 25 to 34) who work full-time increases from \$18,508 for those with less than a ninth-grade education, to \$27,201 for high school graduates, to \$41,628 for holders of a bachelor's degree, and on up for yet more education. In part, this association reflects causation—education helps people qualify for better jobs. Identify a lurking variable that might also contribute to the association. Explain your reasoning.