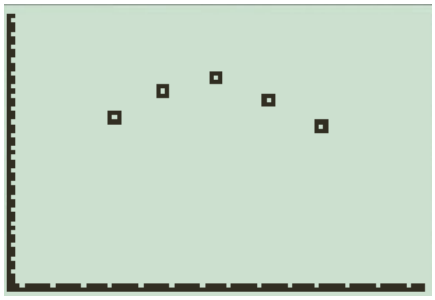
Name:_____ Date:_____

AP Statistics Assignment #3 (with solutions)

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

   a. Make a scatterplot of mileage versus speed



   b. 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?

The relationship between speed and mileage is not linear and shows a parabolic shape. Correlation is equal to zero if there is no linear relationship

Wrong answers: Correlation can not be used to describe curved relationship. <Sure it can, just not that good. We use it for exponential regressions.>

2. Data on the IQ test scores and reading test scores for a group of fifth-grade children give the regression line [*Reading score* = $-33.4 + 0.882$ (*IQ score*)] for predicting reading score from IQ score

   a. Explain what the slope of this line tells you

For every increase in 1unit in the IQ score, our model predicts an average increase of 0.882 units of reading score

   b. Find the predicted reading scores for two children with IQ scores of 90 and 130, respectively

45.98 and 81.26

   c. Draw a graph of the regression line for IQ's between 90 and 130.

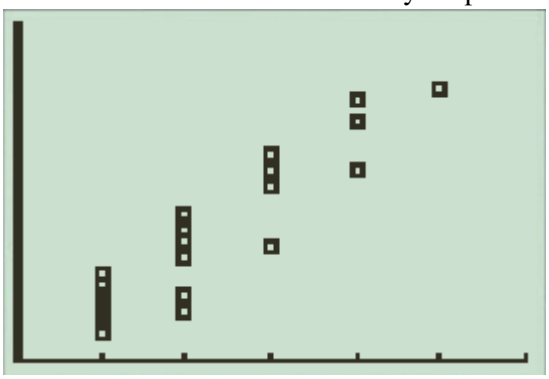
   d. Interpret the y-intercept of this line. Why doesn't this make any sense?

The y-intercept indicates that an IQ score of 0 units, our model predicts a reading score of -33.4 points. This doesn't make sense because of two reasons: a person cannot have an IQ score of 0 and your reading score cannot be negative.

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 diamter, 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.

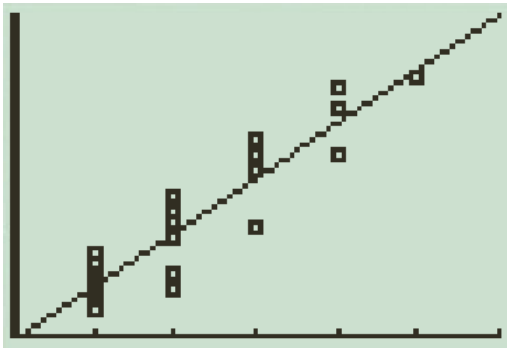
Stumps	2	2	1	3	3	4	3	1	2	5	1	3	2	1	2	2	1	1	4	1	2	1	4
Beetle Larvae	10	30	12	24	36	40	43	11	27	56	18	40	25	8	21	14	16	6	54	9	13	14	50

a. Make a scatterplot that shows how the number of beaver-caused stumps influences the number of beetle larvae clusters. What does your plot show?

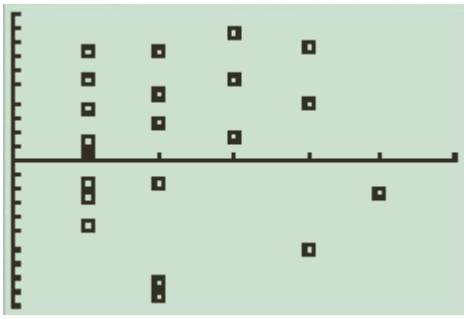


b. Find the least-squares regression line and draw it on your plot

$\hat{y} = -1.2861 + 11.8937x$



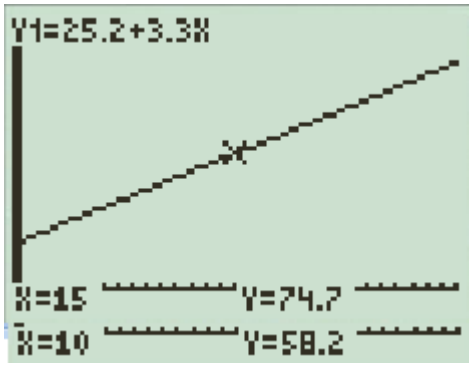
c. Construct a residual plot. How well does the linear model fit the data?



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

$r^2 = 0.839$ 83.9% of the variation in the number of clusters of beetle larvae can be explained by the approximate linear relationship with the the number of stumps from trees cut by beavers

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 correspons to an increase of 5 chirps per minute?

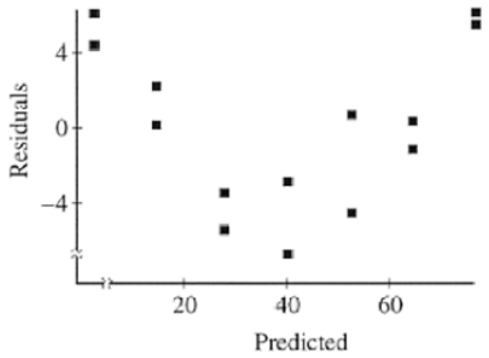


The difference in the resopnse variable is 74.7-58.2 = 16.5 degrees farenheit

5. In a study of the application of a certain type of week killer, 14 fields containing large number of weeks were treated. The week 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 week killer. After a few days, the percentage of weeds killed on each filed 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



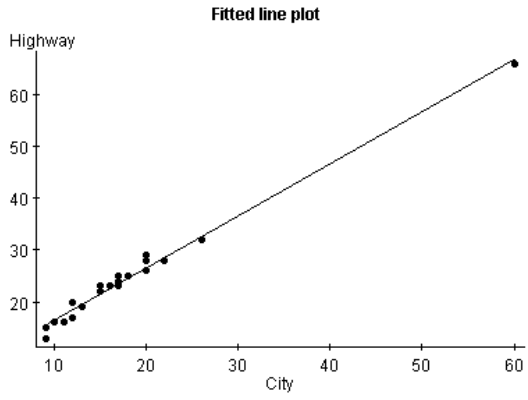
a) What is the equation of the least squares regression line given by this analysis? Define any variables used in this equation

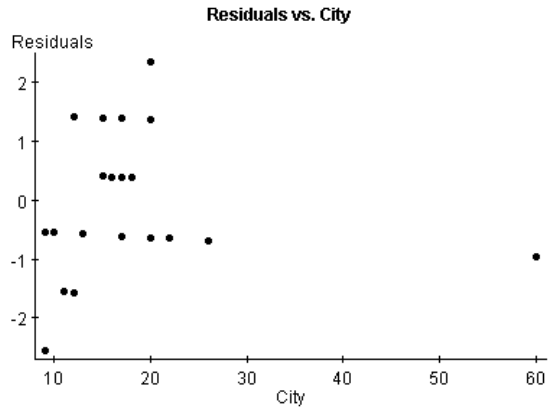
$\hat{y} = -20.5893 + 24.3929x$ $\left(\begin{matrix}Predicted\ \%\\ ofweed\ killed\end{matrix}\right) = -20.5893 + 24.3929\left(\begin{matrix}No.\ of\\ teaspoons\end{matrix}\right)$

b) If someone uses this equation to predict the percentage of weeds killed when 2.6 teaspoons of week 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.

An observation with 2.6 teaspoons would yield a predicted percentage of weed killed by about 42.83%. Based on the residual plot, a predicted percentage of 42.83% would yield a negative residual, meaning that the predicted value will be greater than the observation. Therefore, based on this model, the prediction will be too large.

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.

The honda Insight has exceptional gas mileage and differes greatly from other cars, but the correlation between it's cit mileage and highway mileage is the same as other cars. It is an outlier in the sense that its values are much higher but it falls under the same overall pattern. It does not have the largest residual because its actual value is very close the the prediction from the regression line.

(b) If the Honda Insight were removed from this set of data, would the correlation increase, decrease, or stay the same? Justify your answer.

In this case, removing the observation from the Honday Insight would have very little effect on the regression line because the residual value is very small.

(c) Is the Honda Insight influential on the slope of the regression line? Justify your answer.

No, because the observed value is very close to the regression line.

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.

Examples of Lurking variables:

1. People with more education have a larger network and allow them to get better jobs with higher pay.
2. People with more education tend to be from wealthier backgrounds. They may have a larger network or better access to high paying jobs.
3. People with less education may be from lower income families in the first place. They may settle for lower paying jobs to meet family needs.