Name:_____

AP Statistics Assignment 3.3 Lurking Variables, Residual Plots, and Outliers

1. Define each of the following terms using your own words:

Outliers

Influential observations in regression:

Lurking variable

2. Given each scenario, indicate the possible lurking variables.

i) A study explored the effects of listening to Mozart at a young age and the IQ of the participants when they became young teenagers. It was then donned the "Mozart effect" where listening to Mozart music makes you smarter.

ii) A positive correlation existed in a Miami beach where the number of ice cream sales increased with the number of shark attacks.

iii) A study was done to measure the number of firefighters and the amount of damage done by the fire. Is it enough to claim that the more firefighters you have, the more damage is created?

iv) In WWII, bombers that attacked London tend to be less accurate with the absence of Allied resisting fighter pilots. Whereas, when fighter pilots were defending the city, the bombers were more accurate. What is the lurking variable?

v) Does AP exams cause Global warming? In the past few years the number of AP exams taken have been increasing steadily. Likewise, the average global temperature is also increasing steadily.

3. What are the differences between an “outlier” and an “influential observation” in regression?

4. Given the following table of values,

x	1	2	3	4	5	6
y	1	3	3	4	5	6

a) use your Ti-83 to draw a scatterplot and find the LSRL

b) Calculate the residual for each data point

c) Draw a residual graph

d) A new point (8,3) was collected. Is this new point an “outlier” or an “influencial point” Explain

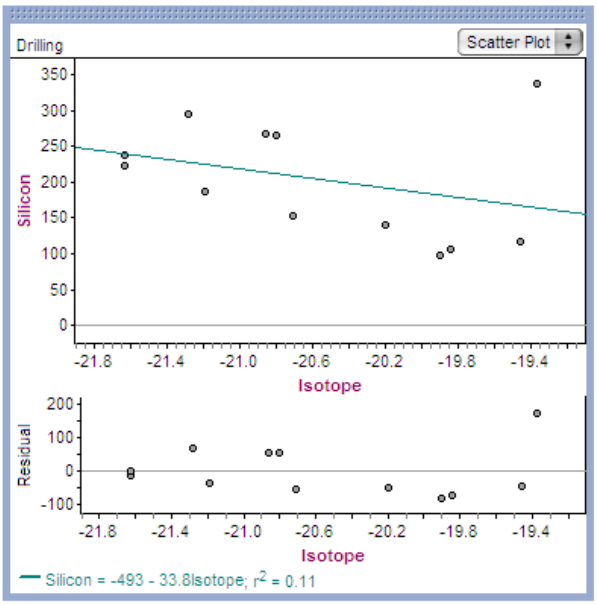
e) How does the new point (8,3) affect the LSRL?

5. What are the two tools that you can use to measure if a LSRL is a good fit?

6. How do you use a residual plot to measure if a LSRL is a good fit? Explain:

7. Suppose r^2 is very close to 1 (ie: 0.96). Under what circumstances will r^2 still be a poor tool for determining if the LSRL is a good fit for the data. Provide an example.

8. Drilling down beneath a lake in Alaska yields chemical evidence of past changes in climate. Biological silicon, left by the skeletons of single-celled creatures called diatoms, is a measure of the abundance of life in the lake. A rather complex variable based on the ratio of certain isotopes relative to ocean water gives an indirect measure of moisture, mostly from snow. As we drill down, we look further into the past. Here is some computer output from a linear regression analysis of data from 2300 to 12,000 years ago:

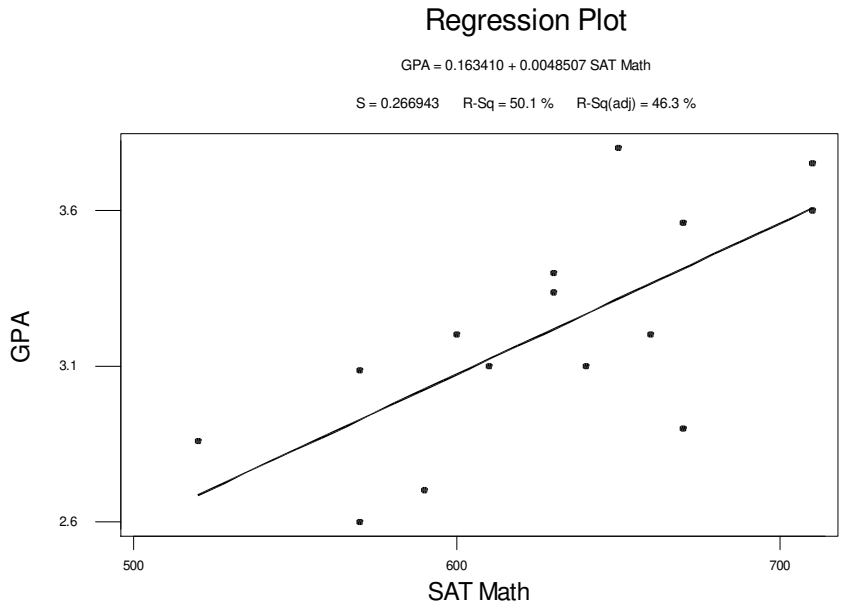


(a) Circle the unusual point in the scatterplot and the residual plot.

(b) If this point was removed, describe the effect on
i) the correlation ii) the least-squares line.

9. What explains grade inflation? Students at almost all colleges and universities get higher grades than was the case 10 or 20 years ago. Is grade inflation caused by lower grading standards? Suggest a lurking variable that might explain higher grades even if standards have remained the same. Justify your answer.

10. Mr. Wright believed that he might be able to use students' SAT Math scores to predict their overall grade point averages. He recorded data on a sample of 15 of his students. The scatterplot below displays the data, along with the least-squares regression line.



One student was absent that day. His SAT Math score is 790 and his grade point average is 3.5. What effect would adding this student's point to the scatterplot have on each of the following? Justify your answers.

(a) the correlation

(b) the slope and y intercept of the regression line

11. The effect of a lurking variable can be surprising when individuals are divided into groups. In recent years, the mean SAT score of all high school seniors has increased. But the mean SAT score has decreased for students at each level of high school grades (A, B, C, and so on). Explain how grade inflation in high school (the lurking variable) can account for this pattern.