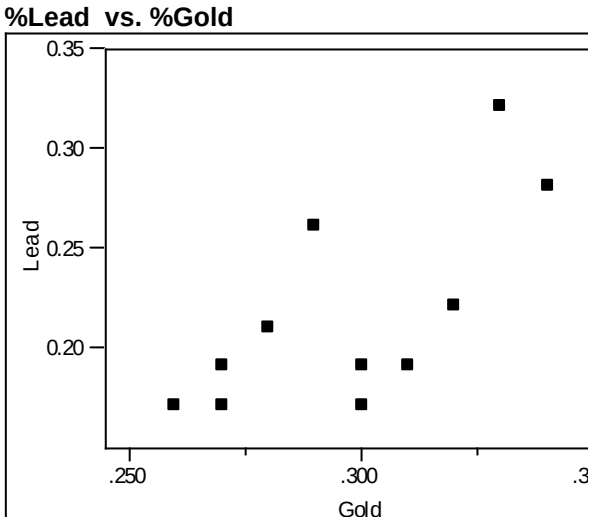
A.P. Statistics
Assignment 2-6

Remember to show your thinking through your work.

1) During the first 3 centuries AD, the Roman Empire produced coins in the Eastern provinces. Some historians argue that not all these coins were produced in local mints, and further that the mint of Rome struck some of them. Because the "style" of coins is difficult to analyze, the historians would like to use metallurgical analysis as one tool to identify the source mints of these coins. Investigators studied 11 coins known to have been produced by local mints in an attempt to identify a trace element profile for these coins, and have identified gold and lead as possible factors in identifying other coins as having been locally minted. The gold and lead content, measured as a % of weight of each coin, is given in the table at right, and a scatter plot of these data is presented below.

Gold % by Weight	Lead % by Weight
0.30	0.17
0.31	0.19
0.28	0.21
0.32	0.22
0.34	0.28
0.29	0.26
0.33	0.32
0.26	0.17
0.30	0.19
0.27	0.19
0.27	0.17



a) What is the equation of the least squares best fit line?

<type answer here>

b) What is the value of the correlation coefficient? Interpret this value.

<type answer here>

c) Suppose that the locally minted coins analyzed in problem 1 are representative of the metallurgical content of mints in the Eastern provinces of the Roman Empire during the first 300 years AD. A locally minted coin is selected at random, and its gold content is 0.30% by weight, calculate the predicted lead content. Be sure to use correct notation and units.

<type answer here>

d) One of the coins used to calculate the regression equations has a gold content of 0.31%. Calculate the residual for this coin. Be sure to use correct notation and units.

<type answer here>

e) The investigators would like to use the regression equation to help assess whether a coin is locally minted or minted in Rome. In a few sentences, discuss the appropriateness of the regression line and how it might be used.

<type answer here>

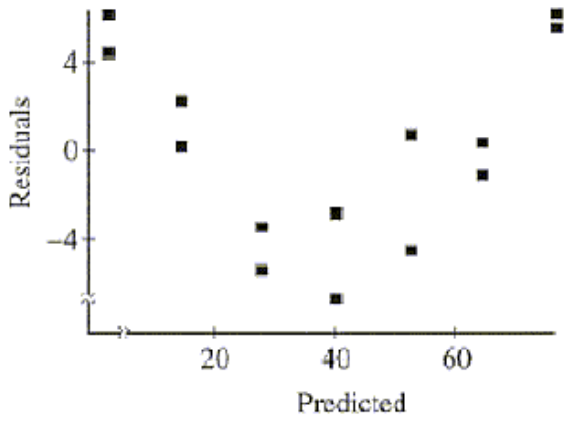
2) The following problem is an actual problem from a past AP exam. Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy of your results and explanation.

4. In a study of the application of a certain type of weed killer, 14 fields containing large numbers 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 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.

(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?

- ○ The prediction will be too large.
- ○ The prediction will be too small.
- ○ A prediction cannot be made based on the information given on the computer output.

Explain your reasoning.

<type answer here>