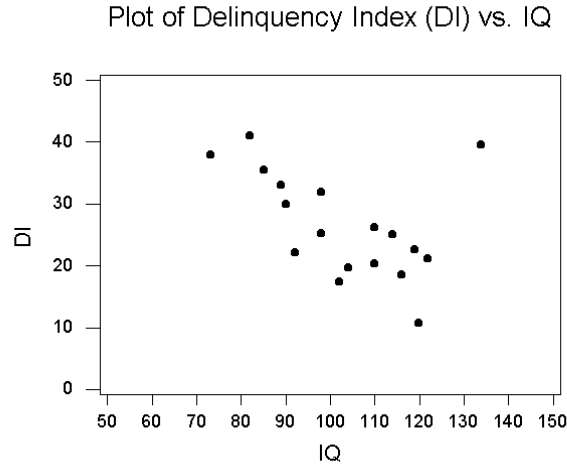


STATISTICS 101 - Homework 4 Answers
Due Friday, February 22, 2008

This assignment is worth 90 points.

- (21 pts) A sociologist at a correctional institution studied the relationship between intelligence and delinquency. A delinquency index (on a scale from 0 to 50 with higher numbers indicating greater delinquency) was constructed using information on the severity and frequency of crimes committed. Intelligence was measured using a standard IQ test. The plot below displays data for 18 inmates at the correctional institution.



- (3 pts) Describe the who and the what for these data.
The who are the 18 inmates at the correctional institution and the whats are the two variables collected for each person, their DI index and their IQ.
- (4 pts; 1 pt for each value) What is the highest value for IQ? What is the associated delinquency index? What is the lowest value for IQ? What is the associated delinquency index?
The highest value for IQ is approximately 132. The associated delinquency index is approximately 40. The lowest value for IQ is approximately 73. The associated delinquency index is approximately 38.
- (4 pts) Describe in words the apparent association between IQ and the delinquency index.
It appears the relationship is a fairly strong negative linear relationship. However, there is one potential outlier at (132,40) that causes the relationship between the two variables to appear weaker than it probably is.
- (2 pts) One of the observations appears to be an outlier. What value is this (give approximate values for IQ and DI)? Why does it appear to be an outlier?
The outlier is at (132,40). It appears to be an outlier because it lies outside the pattern of the other data points.
- (3 pts) The value of the correlation coefficient for the entire set of 18 data points is -0.475 . If this outlying observation were removed, how would the value of the correlation coefficient change. Explain your reasoning.

If the outlier were removed, the correlation coefficient should decrease. Since the outlier has both a greater than average IQ and a greater than average DI index, this point is adding a positive value to the correlation coefficient value. If the point is removed, the correlation coefficient will decrease. This is indicated in the graph by a stronger negative relationship between the two variables.

- (f) (5 pts) Below are summary data for the 17 observations after eliminating the one outlier. Calculate the value of the correlation coefficient, r . Does this agree with your assessment in (e)?

$$n = 17 \quad \sum Y = 438.8 \quad \sum X = 1724$$

$$\sum(X - \bar{X})(Y - \bar{Y}) = -1533.88 \quad \sum(Y - \bar{Y})^2 = 1046.18 \quad \sum(X - \bar{X})^2 = 3454.12$$

First, we need to calculate the standard deviation for both X and Y.

$$s_x = \sqrt{\frac{\sum(X - \bar{X})^2}{n - 1}} = \sqrt{\frac{3454.12}{16}} = 14.69$$

$$s_y = \sqrt{\frac{\sum(Y - \bar{Y})^2}{n - 1}} = \sqrt{\frac{1046.18}{16}} = 8.09$$

Now, we can calculate r .

$$r = \frac{1}{n - 1} \frac{\sum(X - \bar{X})(Y - \bar{Y})}{s_x s_y} = \frac{1}{16} \frac{-1533.88}{14.69(8.09)} = -0.81$$

2. (35 pts) Can the length of a person's forearm be used to predict the length of a person's foot? We will collect data on the two variables during Lab 4. The data below were taken from a sample of 25 female Stat 101 students during a previous semester.

Forearm (cm)	Foot (cm)	Forearm (cm)	Foot (cm)	Forearm (cm)	Foot (cm)
24	24	23	25.5	24.5	25.5
24	27.5	26	27	26	28
25	30	24.5	25	26	30
25.5	26.5	27.5	28	26	28
23.5	27	24.5	27.5	25	24
27	28.5	26	25.5	24	27
29	31	24	26	27	26.5
27	29	25	25	28	27
28	28				

Follow the instructions from the JMP Guide to enter the data above into JMP (you should enter two variables, Forearm and Foot) and obtain a scatterplot with regression line, regression coefficients and a residual plot. Then print out the results and include them with your assignment. Use your printout to answer the following questions.

5 pts for the JMP output

- (a) (3 pts) Describe the who and the what for these data.
The who are the sample of 25 women and the whats are the variable length of foot and length of forearm.
- (b) (3 pts; 1 pt for variables; 2 pts for explanation) What is the explanatory variable? What is the response variable? Briefly explain your choice.
The explanatory variable is the length of the forearm and the response variable is the length of the foot. The variables were selected for these roles since we wish to predict the length of the foot using the length of the forearm.
- (c) (4 pts) Describe the general relationship between the two variables.
The forearm length and foot length of the 25 women have a fairly weak positive linear association. The general relationship indicates that women with longer forearm lengths will have longer foot lengths, but there is a lot of variation around the line.
- (d) (4 pts; 1 pt for value; 3 pts for explanation) Give the value for the slope of the least squares regression line. Give the interpretation of this value within the context of the problem.
The slope is 0.6311 cm. This means that a 1 cm increase in forearm length is associated with a 0.6311 cm increase in foot length, on average.
- (e) (3 pts) Give the value for the intercept of the least squares regression line. Explain why the interpretation of this intercept value does not make sense in this context.
The intercept for the regression line is 10.9240 cm. The intercept is a predicted value of the foot length when the forearm length is 0 cm. This of course doesn't make sense given the definition of the variables.
- (f) (4 pts; 2 pts for equation; 2 pts for predicted value) Give the equation of the least squares regression line for this problem. Use this equation to predict the length of a person's foot given the length of their forearm is 29 cm.
The predicted foot length = $10.9240 + 0.6311 * \text{forearm length}$. For a woman with a forearm length of 29 cm, their predicted foot length is = $10.9240 + 0.6311 * 29 = 29.2259$ cm.
- (g) (3 pts) Would you use this regression line to predict the foot length of a women with a forearm length of 32cm? Explain your answer.
No, the largest forearm length of the 25 women is 29cm. We do not have any information on the relationship between forearm length and foot length for women with forearm length greater than 29cm. Making a prediction of this manner would be extrapolation.
- (h) (2 pts) One woman had a forearm length of 29 cm. What is the residual for this woman?
The residual is the actual value - predicted value. Going to the table above, the foot length for this woman was 31 cm. So the residual is $31 - 29.2259 = 1.7741$.
- (i) (4 pts; 1 pt for value; 3 pts for explanation) Give the value of R^2 for this regression. Give the interpretation of this value in the context of the problem.
The value of R^2 is 30.06%. This means that 30.06% of the variation in the foot length of the 25 women can be explained by the linear regression with their forearm lengths.

3. (34 pts) Many physical characteristics of people and animals are related. For example, taller people tend to weigh more. In this problem, we will be looking at the relationship between two physical characteristics of butter clams, length and width. In particular, we will be

looking at predicting the length of a butter clams from the width of the butter clam. Follow the instructions in the JMP Guide to download the data from the main course webpage into JMP and then to obtain a scatterplot with regression line, regression coefficients and a residual plot. The data are the length and width in centimeters of a sample of 88 butter clams.

5 pts for the JMP output.

- (a) (3 pts) Describe the who and the what for these data.
The who are the 88 butter clams and the whats are the two variables, length and width for each clam.
- (b) (3 pts; 1 pt for variable; 2 pts for explanation) Which variable is the response variable in this regression? Explain why you chose this variable as your response.
The response variable is the length. Length was chosen for this role since we want to use the width to predict the length.
- (c) (2 pts) What is the least squares regression equation relating the length of butter clams to their width?
 $\text{Predicted length} = 0.2569 + 1.2201 * \text{width}$
- (d) (3 pts) Does the interpretation of the intercept make sense in the context of these data? Explain your answer.
No, the intercept is a prediction value when the width of a butter clam is 0cm. This doesn't make sense in the context of the problem since a butter clam must have at least a small width.
- (e) (4 pts) Give the interpretation of the slope within the context of the problem.
The slope of the regression line is 1.22 cm. This means that on average, a 1 cm increase in width will mean an increase of 1.22 cm of length.
- (f) (2 pts) Use the least squares regression to predict the length of a butter clam who has a width of 7 cm.
 $\text{Predicted length} = 0.2569 + 1.2201 * 7 = 8.7976 \text{ cm.}$
- (g) (2 pts) The data includes a clam whose width is 7 cm and length is 9.5 cm. Find the residual for this clam.
The residual is actual value - predicted value = $9.5 - 8.7976 = 0.7024$.
- (h) (4 pts; 1 pt for value; 3 pts for explanation) What is the value of R^2 for these data? Give the interpretation of this value in the context of the problem.
The value of R^2 is 97.87%. This means that 97.87% of the variation in the length of the 88 butter clams can be explained by the linear regression with width.
- (i) (2 pts) What is the value of r, the correlation between the width and length of a butter clam?
Since the slope is positive, r will be the positive square root of R^2 . This is $r = 0.9893$.
- (j) (4 pts) Describe what you see in the residual plot. What does this tell you about using a regression line to summarize the relationship between the width and length of a butter clam?
There is a megaphone pattern in the residual plot. This means that for large widths, there is more variability in the length of butter clams than for smaller widths. The regression line is an appropriate way to summarize this relationship but caution must be used during further analysis.