9.07
Introduction to Statistical Methods
Homework 7

Name: _____

1. A researcher measures the following scores for a group of people. X indicates the number of errors on a math test, and Y indicates the person's level of satisfaction with their performance.

X	9	8	4	6	7	10	5
Y	3	2	8	5	4	2	7

(a) Plot the scatter plot for this data.
(b) Find the best fit line, Y=a+bX
(c) Plot the residual plot
(d) What is the Pearson's Product-Moment Correlation coefficient, r?

2. You want to know if a nurse's absences from work in one month can be predicted by knowing her score on a test of psychological "burnout". From the data shown here, (a) what is the regression equation? (b) Plot the regression line with the data. (c) What is r? Is this a strong relationship between the two variables?

Absences	4	7	6	9	6	8	7	10	11
Burnout	2	1	2	3	4	4	7	7	8

3. You measure how much people are initially attracted to a person, and how anxious they become during their first meeting with him or her. Here's the data:

Initial attraction	2	6	1	3	6	9	6	6	4	2
Anxiety level	8	14	5	8	10	15	8	8	7	6

(a) Compute r.
(b) Compute the linear regression equation predicting anxiety level from attraction.
(c) What anxiety score do you predict for a person who has an attraction score of 9?
(d) Is the slope of the regression equation, b, significantly different from 0?

4. (a) What proportion of the variance in anxiety level is accounted for by the relationship in problem 3? (b) What is proportion of the variance not accounted for? (c) When using this linear relationship, what is the "average" amount of error you should expect in your predictions, $s_{y'}$?

5. Using two questionnaires, a researcher measures how positive a person's mood is (X) and how creative they are (Y), obtaining the following scores:

X	10	8	9	6	5	3	7	2	4	1
Y	7	6	11	4	5	7	4	5	6	4

9.07
Introduction to Statistical Methods
Homework 7

Name: _____

(a) What is the predicted creativity score for anyone scoring 3 on the mood questionnaire? (b) If your prediction is in error, how much error do you expect to have (compute $s_{y'}$)? (c) How much smaller will your error be if you use the regression equation than if you merely use the overall mean creativity score as the predicted score? (d) Give confidence intervals for your prediction in (a).

6. You are given the following height and weight data for a set of students:

Height	61	56	73	65	70	69	68	63	63	66	62	62	76	68	61	68	62	61	67	73
Weight	108	96	139	115	144	118	111	136	115	116	128	115	158	170	111	173	135	90	106	163

(a) Find the best-fit line for predicting the student's weight from their height.
(b) Find the best-fit line for predicting the student's height from their weight. Show that this is or is not the line found in (a). (c) If the lines are not the same, where do they intersect?

7. For the data in problem 6, what is the proportion of the variance accounted for by the regression in (a)? What is the proportion of the variance accounted for by the regression in (b)? What is the correlation coefficient?

8. We can think of the regression model for the whole population as a linear relationship of the form $y = \alpha + \beta x + \varepsilon$, where α is the population parameter for the intercept, β is the population parameter for the slope, and ε is normally distributed random noise with mean 0 and standard deviaton, $\sigma(\varepsilon)$. (a) For the data in problem 6, consider the regression equation for estimating weight from height. Estimate $\sigma(\varepsilon)$. (b) Find the standard error for the intercept, a, and find 95% confidence intervals for α. (c) Find the standard error for the slope, b, and find 95% confidence intervals for β. Is the slope significantly different from 0, at the 0.05 level?

9. For the data in problem 6, find the standard error in the estimate of the height, when the weight is 160 lbs. Give 95% confidence intervals for the height, for a weight of 160 lbs. Give the standard error and 95% confidence intervals for the height, for a weight of 120 lbs. What do you notice about the size of the confidence interval for the 160 lbs case vs. the 120 lbs case?

10. For the data in problem 6, test the hypothesis that the population correlation, ρ, is significantly different from 0 at the 0.05 level.