Homework 1, Stat 3

1. Consider a linear regression model:

$$y_i = \alpha + \beta x_i + e_i, \qquad i = 1, 2, \cdots, n$$

where x_i's are fixed and e_i's are independent random errors with mean 0 and variance σ^2. Define two estimators of β as follows

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \quad \text{and} \quad \hat{\beta}_2 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

   (a) Obtain an unbiased estimator of β as a linear combination of $\hat{\beta}_1$ and $\hat{\beta}_2$.

   (b) Find the mean squared errors (square of bias plus variance) of $\hat{\beta}_1$ and $\hat{\beta}_2$.

2. Consider two random variables (X, Y) distributed as bivariate normal. Based on a random sample from this bivariate distribution, the fitted least squares regression line of Y on X and that of X on Y were as follows:

$$Y = 22 - 3X \qquad X = 5.84 - 0.12Y$$

Find the sample means of X and Y, the sample correlation between X and Y and the ratio of the sample variances of X and Y.

[Normality is not required here except to indicate that both X and Y are random and continuous.]

3. Use the anscombe dataset is available in R. It consists of 4 set of data, each with 11 bivariate observations. Compute the following statistics for each dataset: means of the two variables, variances of the two variables, correlation coefficient between the two variables, regression line and R^2 of the second variable on the first. Plot the four datasets and comment on the suitability of a linear regression model in each case.

[This exercise demonstrates the need for plotting before modeling the data.]

4. Suppose you have fitted a regression line $Y = b_0 + b_1 X$ on n pairs of observations where b_0 and b_1 are the least squares estimators. For a new value x_0 of X, what is the predicted value of Y? What is the variance of this prediction? How does the variance depend on x_0? Interpret this result.