Statistics 108, Homework 1

Due : October 5th, 2017, In Class (turn in paper form)

* You need to show the steps to get full credits.

This homework is to practice on the basics of the simple linear regression. Total: 90 points.

1. (30 points) *Understanding the simple linear regression model.*

   (a) (10 pt) When asked to state the simple linear regression model, a student wrote it as follows: $E(Y_i) = \beta_0 + \beta_1 X_i + \varepsilon_i$. Do you agree? Why?

   (b) (20 pt) Consider the normal error simple linear regression model. Suppose that the parameter values are $\beta_0 = 49$, $\beta_1 = 4$, and $\sigma = 3$. State the distributions of Y at $X = 10, 24$, and 38, and briefly explain why.

2. (35 points) *Data analysis.* We are interested in establishing the relationship between weight and height of men. We set weight as the response variable and use height as the predictor variable. We are provided by researchers 150 pairs of observations collected by them and the data is in "weight_full.txt" on canvas.

   (a) (10 pt) Draw a scatter plot using R with height on x-axis and weight on the y-axis. Use proper labels on both axes. Report what you feel on the data.

   (b) (15 pt) Calculate the least squares estimates of the intercept and slope using formulas given in the lecture. Add the estimated line onto the plot in (a). Does the line look like to fit the data well?

   (c) (5 pt) Report the values of $\bar{X}$ and $\bar{Y}$. Does the line you obtain in (b) pass through the point $(\bar{X}, \bar{Y})$?

   (d) (5 pt) Report the following four quantities:

   $$\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}), \quad \sum_{i=1}^{n}(X_i - \bar{X})Y_i, \quad \sum_{i=1}^{n}X_i(Y_i - \bar{Y}), \quad \sum_{i=1}^{n}X_iY_i.$$

   Are they all the same? If not, are some of them the same?

3. (25 points) *Rigorous deviations.*

   (a) (10 pt) Based on part (c) in Problem 2, do you think the statement "The least squares line in simple linear regression always passes the center of the data $(\bar{X}, \bar{Y})$" is correct? If so, show it rigorously.

   (b) (15 pt) If you found the some quantities in part (d) of Problem 2 to be the same, are these quantities always the same no matter what X_i's and Y_i's are? If so, show it rigorously.