Homework 1
Due Tuesday, September 17

Show all your work. For Problems 4, 5 and 6, please try to use R, to gain some practice using the software. Data files available on class website (for Problem 6).

1. Recall that $SXX = \sum(x_i - \bar{x})^2$ and $SXY = \sum(x_i - \bar{x})(y_i - \bar{y})$. Verify that

   (a) $SXX = \sum(x_i - \bar{x})x_i = \sum x_i^2 - n\bar{x}^2$

   (b) $SXY = \sum(x_i - \bar{x})y_i = \sum(y_i - \bar{y})x_i = \sum x_i y_i - n\bar{x}\bar{y}$

2. Consider the model
$$y_i = \alpha + \beta_1(x_i - \bar{x}) + e_i, \quad i = 1, \ldots, n$$
   Starting from $RSS(\alpha, \beta_1) = \sum(y_i - \alpha - \beta_1(x_i - \bar{x}))^2$, show that the least squares estimates are given by
$$\hat{\alpha} = \bar{y}, \quad \hat{\beta}_1 = \frac{SXY}{SXX}$$

3. Show that
$$RSS = \sum \hat{e}_i^2 = SYY - \frac{(SXY)^2}{SXX} = SYY - \hat{\beta}_1^2 SXX$$

4. Consider the following data

X	1	2	3	4	5
Y	1.9	2.2	2.9	3.2	5.2

   Please find $\bar{x}$, $\bar{y}$, SXX, SYY, SXY, along with $\hat{\beta}_0$ and $\hat{\beta}_1$, se($\hat{\beta}_0$) and se($\hat{\beta}_1$). Draw the data points and the regression line.

5. Consider the following dataset

   ```
         ht    wt
   1   169.6 71.2
   2   166.8 58.2
   3   157.1 56.0
   4   181.1 64.5
   5   158.4 53.0
   6   165.6 52.4
   7   166.7 56.8
   8   156.5 49.2
   9   168.1 55.6
   10  165.3 77.8
   ```

This gives `ht` = height in centimeters and `wt` = weight in kilograms for 10 children. Interest is in predicting weight from height.

Please find $\bar{x}$, $\bar{y}$, SXX, SYY, SXY, along with $\hat{\beta}_0$ and $\hat{\beta}_1$, se$(\hat{\beta}_0)$ and se$(\hat{\beta}_1)$. Draw the data points and the regression line.

6. Please find the dataset `simple1.csv`. Here the response is `wage` and the predictor is `educ`.

Please find $\bar{x}$, $\bar{y}$, SXX, SYY, SXY, along with $\hat{\beta}_0$ and $\hat{\beta}_1$, se$(\hat{\beta}_0)$ and se$(\hat{\beta}_1)$.