COURSE: MATH 50 DARTMOUTH COLLEGE (MWF 11:15 AM-12:20 PM), FALL 2015
INSTRUCTOR: NISHANT MALIK
HOMEWORK SHEET NUMBER: 2
POSTED ON: 09/25/2015
DUE ON: 10/02/2015

Directions: Any problem marked with asterisk (*) should be completed using IPython Notebook (Jupyter) and can be uploaded at https://dropitto.me/m50f15 .

1*. **Muscle mass**. A person's muscle mass is expected to decrease with age. To explore this relationship in women, a nutritionist randomly selected 15 women from each 10-year age group, beginning with age 40 and ending with age 79. The results follow; X is age, and Y is a measure of muscle mass. Assume that first-order regression model is appropriate.

Data source: `https://netfiles.umn.edu/users/nacht001/www/nachtsheim/Kutner/Chapter%20%201%20Data%20Sets/CH01PR27.txt`

a. Obtain the estimated regression function. Plot the estimated regression function and the data. Does a linear regression function appear to give a good fit here? Does your plot support the anticipation that muscle mass decreases with age?

b. Obtain the following: (1) a point estimate of the difference in the mean muscle mass for women differing in age by one year, (2) a point estimate of the mean muscle mass for women aged X =60years, (3) the value of the residual for the eighth case, (4) a point estimate of σ^2 .

Reference: *This problem is copied from the textbook, Kutner et. al. "Applied Linear Regression Models", Ed. 5 (problem no 1.27 on page 36). The accompanying data can accessed either using the above link or from the book cd.*

2*. **US Temperatures** The data referenced below gives the normal average January minimum temperature in degrees Fahrenheit with the latitude and longitude of 56 U.S. cities. (For each year from 1931 to 1960, the daily minimum temperatures in January were added together and divided by 31. Then, the averages for each year were averaged over the 30 years.)

Data source and description: `http://lib.stat.cmu.edu/DASL/Datafiles/USTemperatures.html`
Data abstract and reference: `http://lib.stat.cmu.edu/DASL/Stories/USTemperatures.html`

Consider latitude as predictor variable and temperature as response variable and assume the first order regression function.

a. Estimate regression function and plot it on scatter plot of X,Y. The color of each marker should represent the longitude.

b. Plot a distribution of residuals.

c. Show that your fitted line goes through the point $\overline{X}$,$\overline{Y}$.

3. Derivations, proofs and statements:

a. Let $\hat{Y}_i = b_0 + b_1 X_i$ is a least squares based estimate for the first order regression function $E\{Y_i\} = \beta_0 + \beta_1 X_i$. Derive the expressions for b_0 and b_1. Defining residuals as $e_i = Y_i - \hat{Y}_i$, state and prove all the properties for the fitted line.

b. State and explain the assumptions in normal error regression model. What is the difference between the estimator of σ^2 by maximum likelihood estimator for the normal error regression model and the least squares based estimator for the simple regression model.