Assignment #1

Data Mining Due: September 10, 2018

1. Write programs to implement the multiple linear regression analysis. Do not use an R command such as 'lm' or 'glm'.
    a. Make the program to accept file names for data and output from the user. Give the user a prompt to type-in the data file and output file names. (Hint: use 'readline' command in R)

R example

```
data=readline("Enter the data file name: " )
cat("Select the data coding format(1 = 'a b c' or 2 = 'a,b,c'):  ")
fm = scan(n=1, quiet=TRUE)
if(fm==1) {form = ""} else {form = ","}
data=read.table(data, sep=form)
```

    b. The program must print out the coefficient for each X variable and save it into an output file.
    c. The program must calculate the fitted values and save it into an output file.
    d. The program can be a naïve one, thus you don't have to worry about many issues such as missing values, collinearity, ANOVA table, etc.
    e. Turn in the program file at the course website. It must be executable without any modification on the program. After the run, it must generate one output file.
    f. Use a data file named "harris.dat" for this assignment. Assume that the first one is the response variable.
    g. The output file generated by the program must look like the below (the sample output is not the true one).

```
Coefficients
-------------
Constant:  5.312
Beta1:  1.345
Beta2:  .236
Beta3:  -.439
Beta4:  .457

ID, Actual values, Fitted values
--------------------------------
1, 9.5, 9.8
2, 4.6, 4.8
3, -2.3, -3.2
(continue)

Model Summary
-------------
R-square = .5689
MSE = .234
```