

APM 153 Assignment Six – Calculating a Least Squares Regression Using Matrix Math

Define the Problem

(1) This week we will be writing a function called **least_squares.m** that calculates and plots the least squares regression line for a set of x and y data points.

Define the Inputs and Outputs

(2) The inputs for the function will be a set of x and y data points stored as vectors. The output will be a plot of the data points (as points) and a plot of the regression line.

Get the Data

(3) We will download three data files, SITE1.DAT, SITE2.DAT and SITE3.DAT from the APM 153 course website <http://www.esf.edu/pse/scott/class/apm153/apm153.html>

(4) Copy the data files to your USB drive. Double clicking on the name of one of the data files starts the “import wizard”. Import only the data itself.

(5) Once you have imported the data into Matlab, it is stored in a variable called “data”. If you are importing the data from SITE1.DAT, **rename** the variable **site1**.

(6) Repeat the process to import the data from SITE2.DAT into a variable named **site2**, and the data from SITE3.DAT into a variable named **site3**.

Extra Twist

(7) For full credit (a 90% or higher), your program must also ask the user if they would like to transform the data using the **log10** command before plotting the data points and calculating the regression line. For today however, just write the basic function.

One Line of Code

(10) Because we are writing a function, the first line of code in the function will be,...

```
function least_squares(x, y)
```

(11) The input arguments are the x and y data points. There is no output argument.

Review of Matrix Math for Calculating a Least Squares Regression

(12) From Wednesday's class you will recall that we need to put the x values into an **augmented matrix**, which we could call **Aug_x**.

(13) Given the augmented matrix Aug_x with the x values in the first column and all **ones** in the second column, the equation for the slope and intercept of the regression line is,...

$$\text{slope_intercept} = \text{inv}(\text{Aug_x}' * \text{Aug_x}) * (\text{Aug_x}' * y)$$

(14) If Aug_x is the matrix with the x values, what is **Aug_x'** ?

(15) Once you have calculated the slope_intercept, the function that represents the regression line is,...

$$f = \text{slope_intercept}(2) * \text{slope_intercept}(1) * a$$

where **a** is a **symbol** (How do we create a set of symbols in Matlab?)

What to Turn In

(16) Assignment Six is due in lab on March 10th in lab. (Do not blow off class on March 10th!) You will turn in **one** MS Word document (with page numbers) that contains,...

Part One – Your Algorithm written as pseudocode and a flowchart

Part Two – A copy of the mfile of your function with all proper documentation

Part Three – Figures showing the data points and regression lines for SITE1.DAT, SITE2.DAT and SITE3.DAT with titles, axis labels, and legends.

Part Four - (for a 90% or higher) Figures showing the regression lines for the data points in SITE2.DAT and SITE3.DAT transformed using the **log10** command.