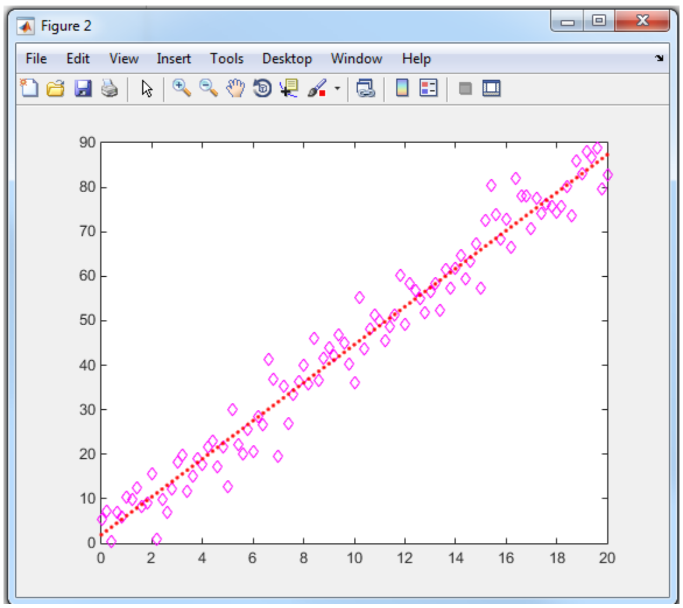
In Class Activity Class 10 – Regression in Matlab

OBJECTIVE: Explore the impact of sample size on the distribution of the sample mean.

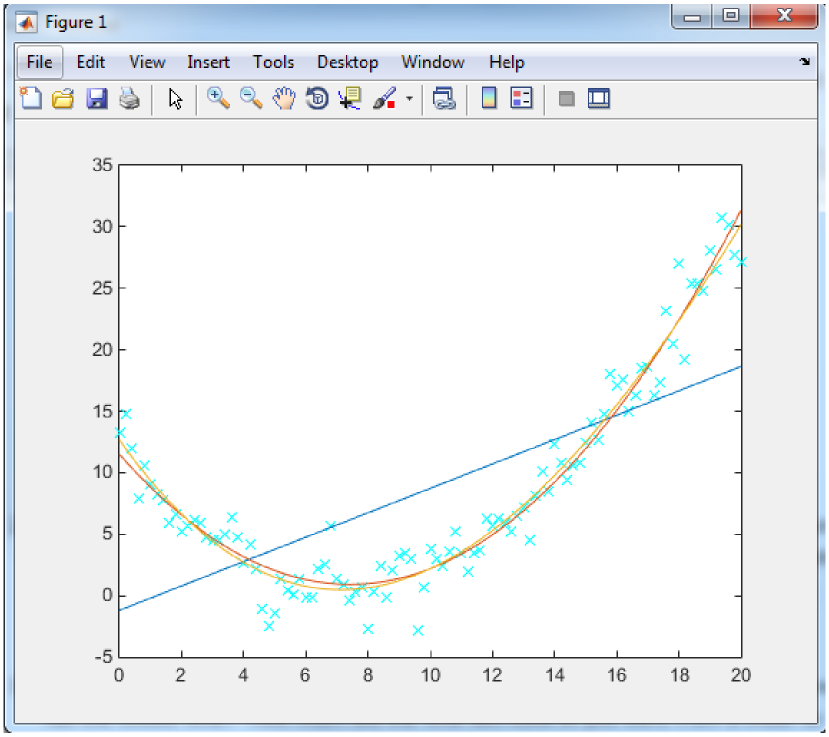
You should work on this task **individually**, but you may discuss methods and results with your group.

1. Download the data file Class10Data1.mat from eCampus. Write a function in which you a) plot the data using a scatter plot of y vs. x to check that a linear model makes sense, b) use polyfit to fit a linear model to the data and output the slope and intercept, and c) plot the best-fit line on the same graph as the scatter plot. Your output should look like this:




```
slope is 4.269476, and intercept is 1.921243
```

2. Download the data file Class10Data2.mat from eCampus. Again, make a scatter plot of y vs. x. Things shouldn't look quite so linear. Use polyfit to fit a a) linear, b) quadratic, and c) cubic polynomial to the data. Plot the three fitted lines on your scatter plot. Your plot should look like this:



3. Now calculate the mean squared error (MSE) for each of the 3 fitted lines. Which model has the lowest MSE? How much did the MSE decrease as you went from a linear to quadratic? How much did the MSE decrease from quadratic to cubic? In your opinion, was it worthwhile fitting a higher order model in this example?

4. Now let's go back to the data set from problem one, and make confidence intervals for the linear model. As presented in the reading you did before class, if n = the number of pairs used in the regression is large , an approximate confidence interval for the slope is given by

$$b_1 \pm z_{\alpha/2} \frac{s_y}{\sqrt{n} s_x}$$

where $s_y$is the standard deviation of y, and s_x is the standard deviation of x. A confidence interval for the intercept is given by

$$b_0 \pm z_{\alpha/2} [\sum_{i=1}^{n} (x_i)^2]^{1/2} \frac{s_y}{n s_x}$$

Add code to your function that will calculate and output these confidence intervals for a degree of confidence of 95%.