

**A simple linear regression problem (20 Points)**  
with some perhaps unintuitive results....

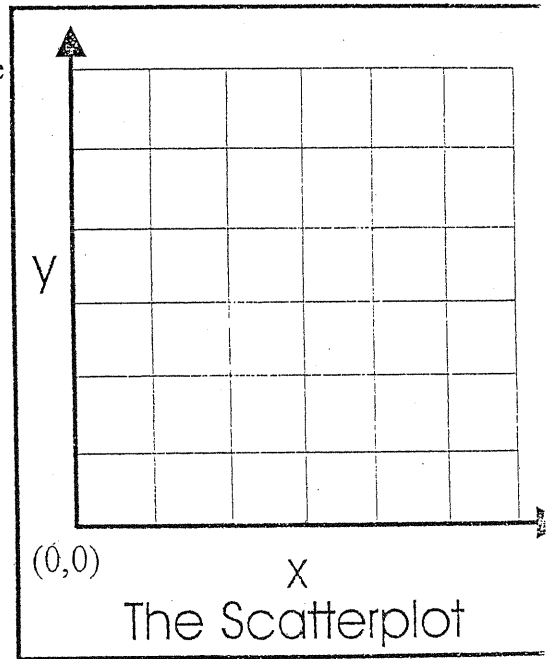
This problem will be completely abstract in the sense that the numbers will be just numbers. Suppose you have been provided with the following X,Y coordinate pairs for six observations of some bivariate phenomenon.

**The Data**

X	Y
1	1
3	2
4	3
2	3
5	5
3	4

Perform the following tasks:

- 1) Plot the observed points as a scatterplot. (draw these 'actual' points as solid dots on the graph provided)
- 2) Draw a dashed line through the points using your eye as judge as to the 'best fit' line. (The line should pass through the grid intersections - the data is set up so that should be 'intuitive')
- 3) Run a 'Simple Least Squares' (SLR) Regression on the points to calculate the slope 'b', and intercept 'a' of the SLR 'best fit' for these points. (show your work by filling out the provided table and fill in the blanks provided)
- 4) Calculate the estimated 'Yhat' values for all the actual X values in the provided data. Use these values to calculate the following: SSR(sum squared regression), SSE (sum squared error), SST(sum squared total), 'Yhat' values, and  $R^2$
- 5) Plot the estimated 'Yhat' values for every actual X value (draw these points as an 'X' with a circle around it, connect these points with a solid line representing the SLR 'best fit' line)
- 6) Answer the question(s) at the bottom of the page.



'b' Slope: \_\_\_\_\_

'a' Intercept: \_\_\_\_\_

'SSR' : \_\_\_\_\_

'SSE' : \_\_\_\_\_

'SST' : \_\_\_\_\_

'R<sup>2</sup>' Correlation: \_\_\_\_\_

**The Table:** (label your columns to show your work)


Question: compare your 'intuitive' line with the SLR line. Use two comparison methods A) Total squared error; and, B) Total error. Discuss the implications. (use back of this page)