

Anupam Vinay Singh

24B2120

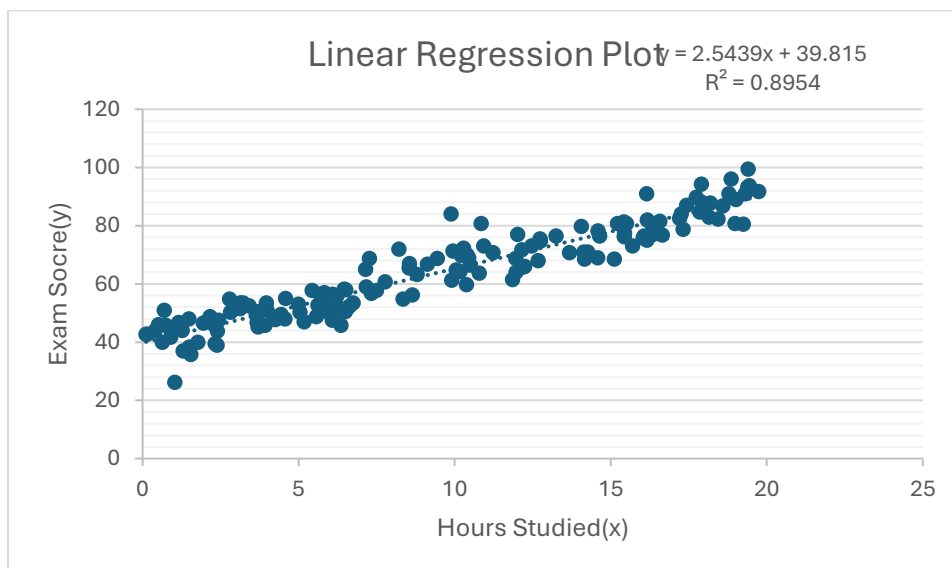
1.

The given document is numeric and has data of hours of study vs marks in the exam.

- ➔ 2 columns containing hours of study and exam score.
- ➔ It has 150 columns meaning it is data of 150 students.
- ➔ Data type is numerical between 1 to 100.
- ➔ Level of measurement is Ratio.

2.

A Linear trend is visible in the data, there are no curvatures visible hence the data can be said a simple linear regression model.



3.

I calculated the value of β_0 and β_1 with the help of excel. I calculated it and also got its value on the plot and both were almost same.

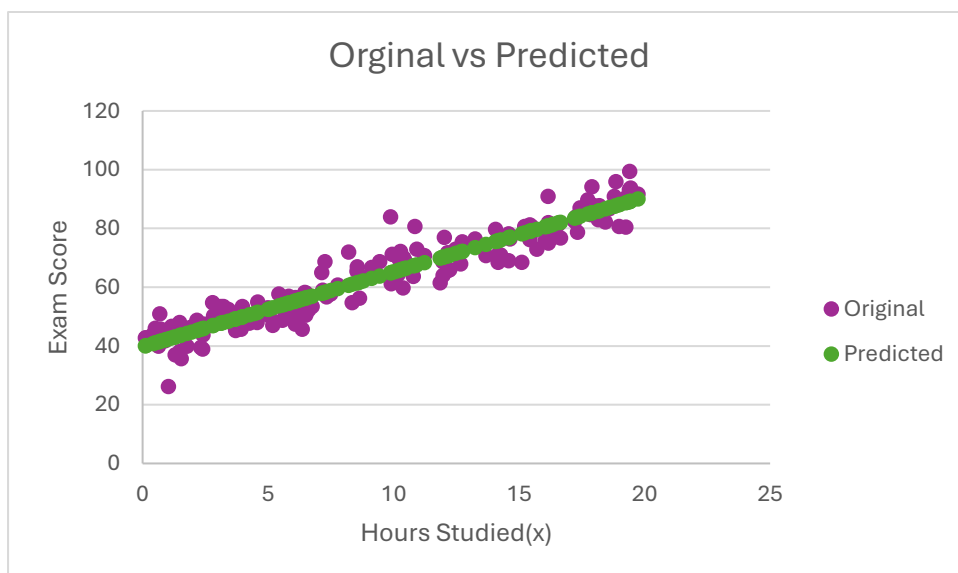
$\beta_0 = 39.8147$ (intercept)

$\beta_1 = 2.5439$ (slope)

4.

Using the regression model, predicted scores were generated for each data point. These predictions were later used to compute residuals and assess the model's performance. Done on spreadsheet.

5. Plot by superimposing the predicted points (x_i, y^{\wedge}_i) over the scatter plot of (x_i, y_i) created earlier.



6.

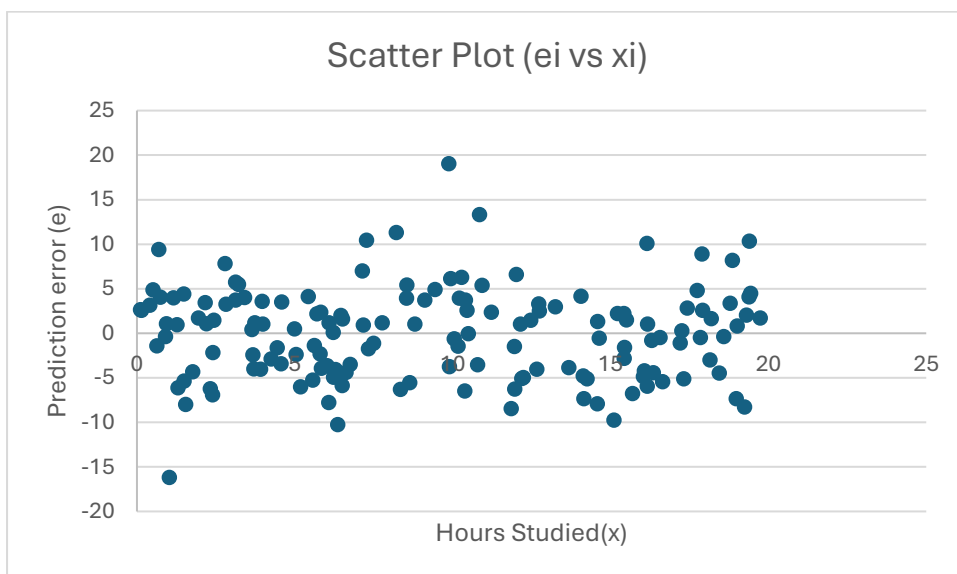
Calculated SSE (Sum of Squared Errors), MSE (Mean Squared Error), RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) using required functions in the spreadsheet, on the lower side of the sheet.

- $SSE = 3960.05$
- $MSE = 26.4003$
- $RMSE = 62.9289$
- $MAE = 4.11214$

Contexts & Applications of Error Metrics:-

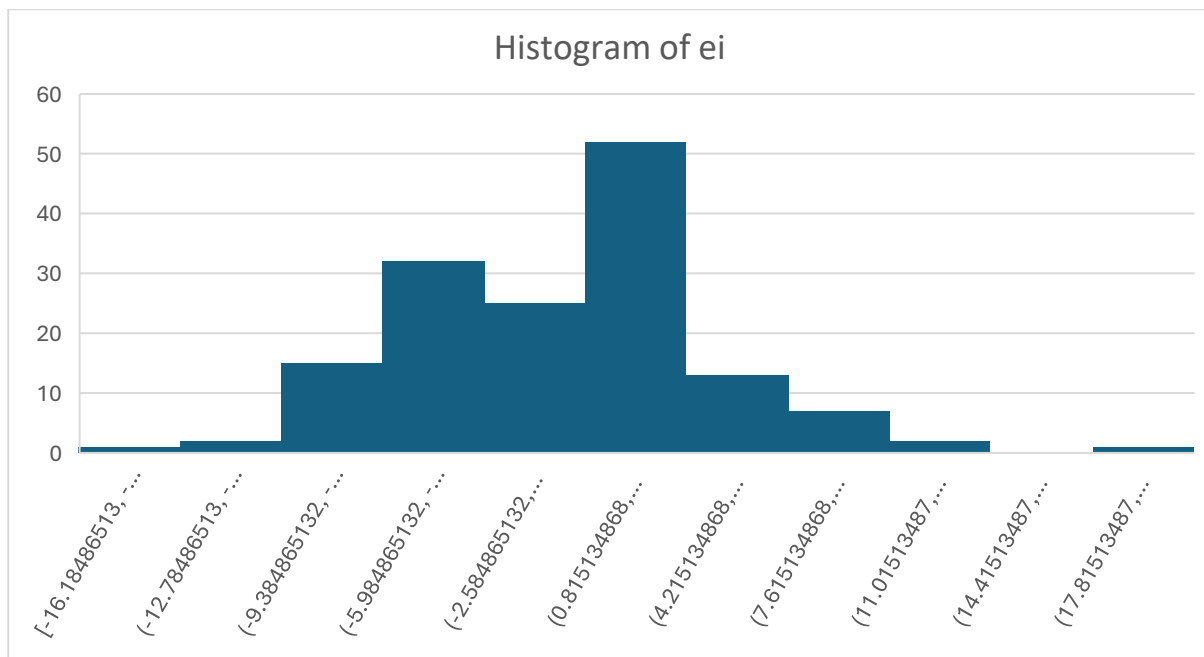
- **SSE:-** Used in model fitting (e.g., to minimize error)
- **MSE:-** Penalizes larger errors more; used in machine learning & regression diagnostics
- **RMSE:-** More interpretable than MSE; standard in forecasting and regression
- **MAE:-** Less sensitive to outliers; useful in robust models or when all errors matter equally

5. Scatter plot of e_i v/s x_i



We observe that there is no pattern in our e_i vs x_i plot and the points are randomly scattered that means our model is working fine.

6. Histogram and error related



My histogram is roughly bell shaped that means the errors are close to zero and there are not much outliers. This shows that my regression model is good.

7. Skewness and Kurtosis

Using inbuilt functions of excel, I calculated the skewness and kurtosis which came as follows:-

Skewness = 0.26146

Kurtosis = 0.87142

- Skewness value is close to zero which means that our plot does not have much outliers which is a property of normal distribution. It means our plot is symmetric.
- In excel kurtosis function gives excess kurtosis after subtracting 3 from it. So our actual kurtosis is 3.87. For normal distribution, kurtosis should be close to 3. Our Kurtosis is also close to 3 so it's roughly a normal distribution. Kurtosis tells us how often outliers occur.

8. R^2

The value of R^2 came to be 0.89543. R^2 is basically the proportion of the variance in the dependent variable (y) that is explained by the independent variable (x). Our R^2 is very high that means our model is very good. It's meaning is that 89.543 % values of y can be explained by the value of x which pretty high.

9.

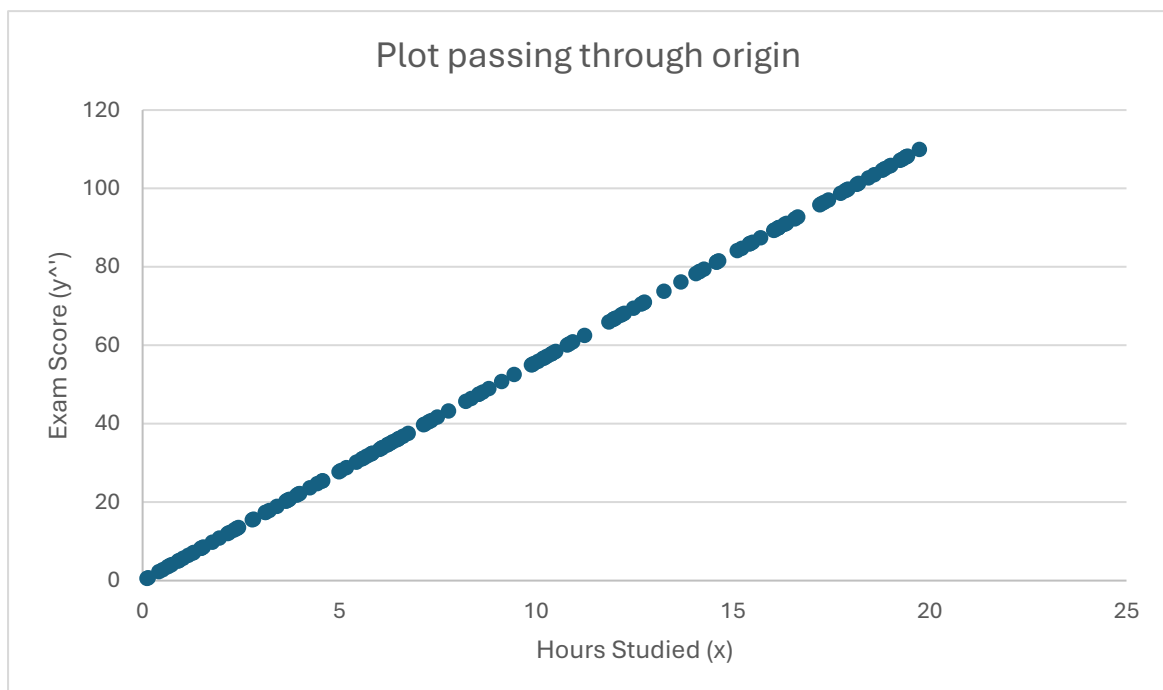
(a)

For this, since there is no intercept, we have to calculate β_1 by:-

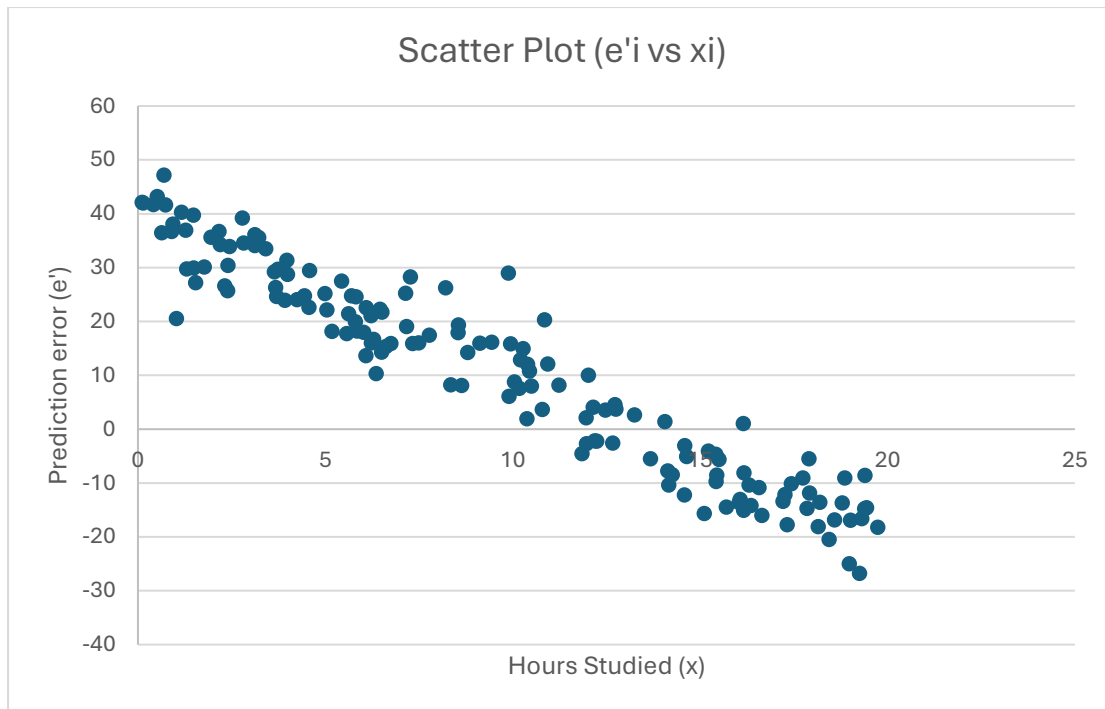
$$\beta_1 = \text{summation}(x_i * y_i) / \text{summation}(x_i^2)$$

(b) Plots

(c) \hat{y}_i vs x_i



(d) new e_i vs x_i



(e) SSE, MSE, RMSE, MAE for this model

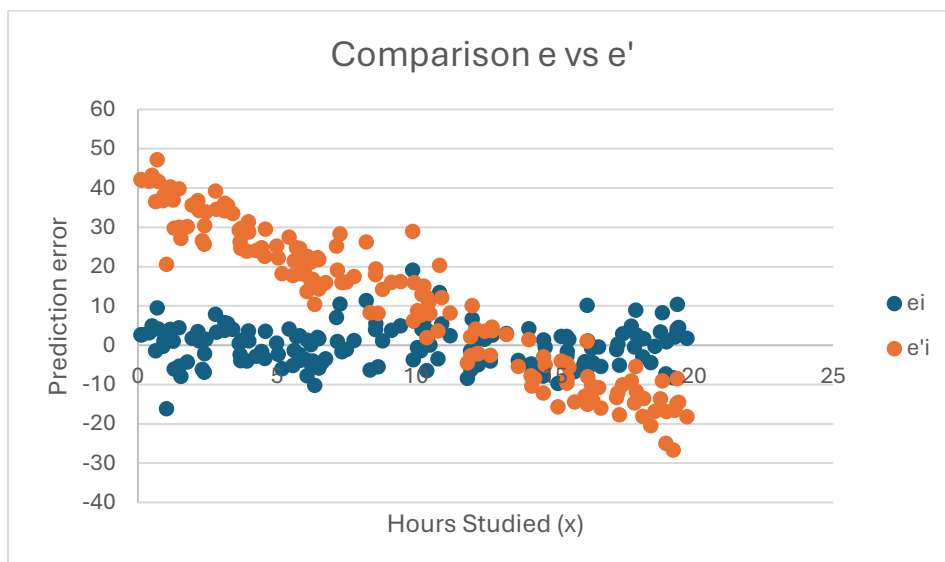
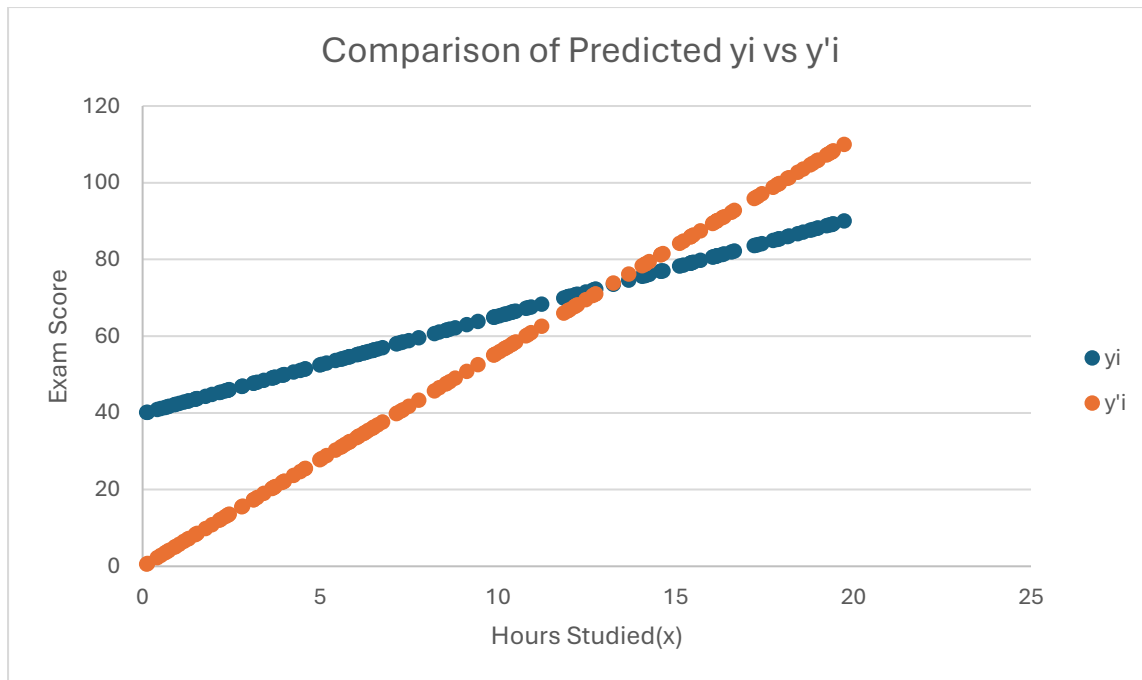
SSE = 70747

MSE = 471.647

RMSE = 265.983

MAE = 18.6231

(f) Comparison of these plots and metrics with the corresponding plots and metrics of the earlier model



Comparison of SSE, MSE, RMSE and MAE:-

All these values increase when our line passes through origin. This is because at first we did the best fit. So for best fit, the error is minimised which in turn reduces these quantities. But when we force our line to go through the origin so that intercept is zero, this results in greater error both at the start of the plot and the end of the plot. Since SSE, MSE, RMSE and MAE are all related to error, these values increase.

Note:- In the excel sheet I have referred to the values which we found out for making intercept zero marked by ' sign (like y^{\wedge} ' and e_i ', etc.) I have also written (new) below these so that there is no confusion.