R-Square ():

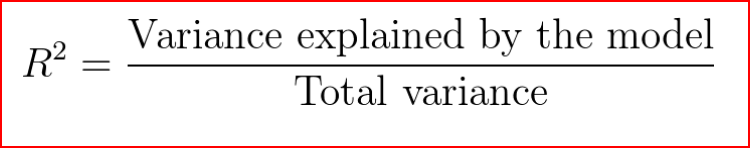
***How do I measure the performance of my regression model?***

*A few statistical tools like coefficient of determination also called as R², Adjusted R² and Root mean square Error -RMSE are commonly used to evaluate the performance of the regression model.*

***What is***R²***?***

It is also called as coefficient of determination.

R²**gives us a measure of how well the actual outcomes are replicated by the model or the regression line**. This is based on the total variation of prediction explained by the model.



R² is always between 0 and 1 or between 0% to 100%.

value of 1 means that the model explains all the variation in predicted variable around its mean.

***Sum square of errors(SSE) or Residuals, how far did we predict a value when compared to the actual value***

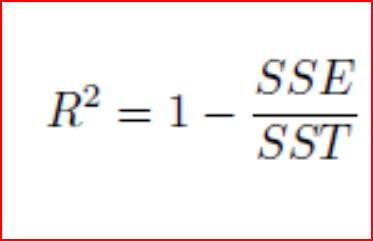
**SSE = Actual value -Predicted value**

***Sum square of total (SST), how far is the actual value when compared to the mean value***

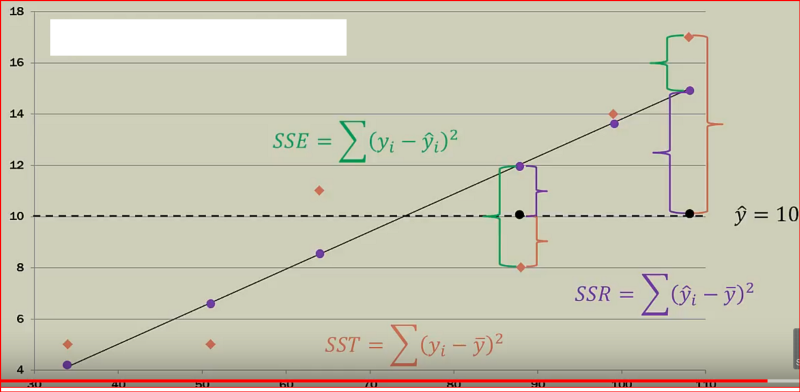
**SST = Actual value -Mean value**

***Sum square of Regression(SSR), how far is the actual value when compared to the mean value***

**SSR = Predicted value -mean value**



If the error in prediction is low then SSE will be low and R² will be close to 1.



A caution of note here, when we add more independent variables, R² gets higher value. R² value keeps on increasing with addition of more independent variables even though they may not really have a significant impact on the predictions. This does not help us to build a good model.

To overcome this issue, we use **Adjusted R²**. Adjusted R² penalizes the model for every addition of an insignificant independent variable.

A value close to 1 for R² means a good fit.we can also calculate **root mean square error**also referred as **RMSE**.

**Root Mean Square Error**

RMSE shows the variation between the predicted and the actual value. Since the difference between predicted and actual values can be positive or negative . To offset that difference we take the square of the difference between predicted and actual value.

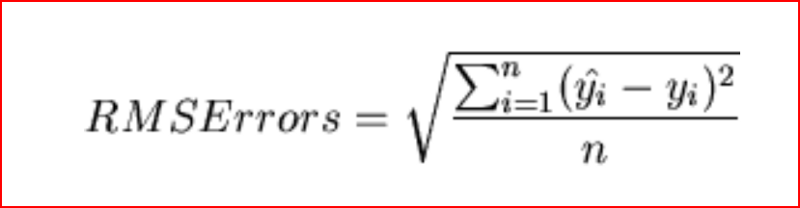
**Step 1:** Find the difference between predicted and actual value for every observation and square the value and add them

Sum of all observation (predicted value — actual value)²

**Step2:** divide the sum by number of observations

Sum of all observation (predicted value — actual value)²/number of observations

**Step 3:** Take the square root of the value from step 2



***How do I measure the performance of my classification model?***

For Binary classification we expect an output of 0 or 1. The output is a predictive score which conveys the probability of the output to be either a 0 or a 1. Typically if the score is the above a certain threshold value then we set the output to 1 else the output will be 0. This threshold value is usually selected as 0.5 but can vary.