R-squared & Mean Squared Error

Presented By

Name: Apoorv Yadav

GitHub Profile: https://github.com/Apoorv-17

Batch No: 15

Serial No: 537

R-squared

- **R-squared** is a statistical measure of how close the data are to the fitted regression line.
- It is the percentage of the response variable variation that is explained by a linear model.

$$\text{R-squared} = \frac{Explained\ Variation}{Total\ Variation}$$

- It is also called a **coefficient of determination**, or **coefficient of multiple determination** for multiple regression.

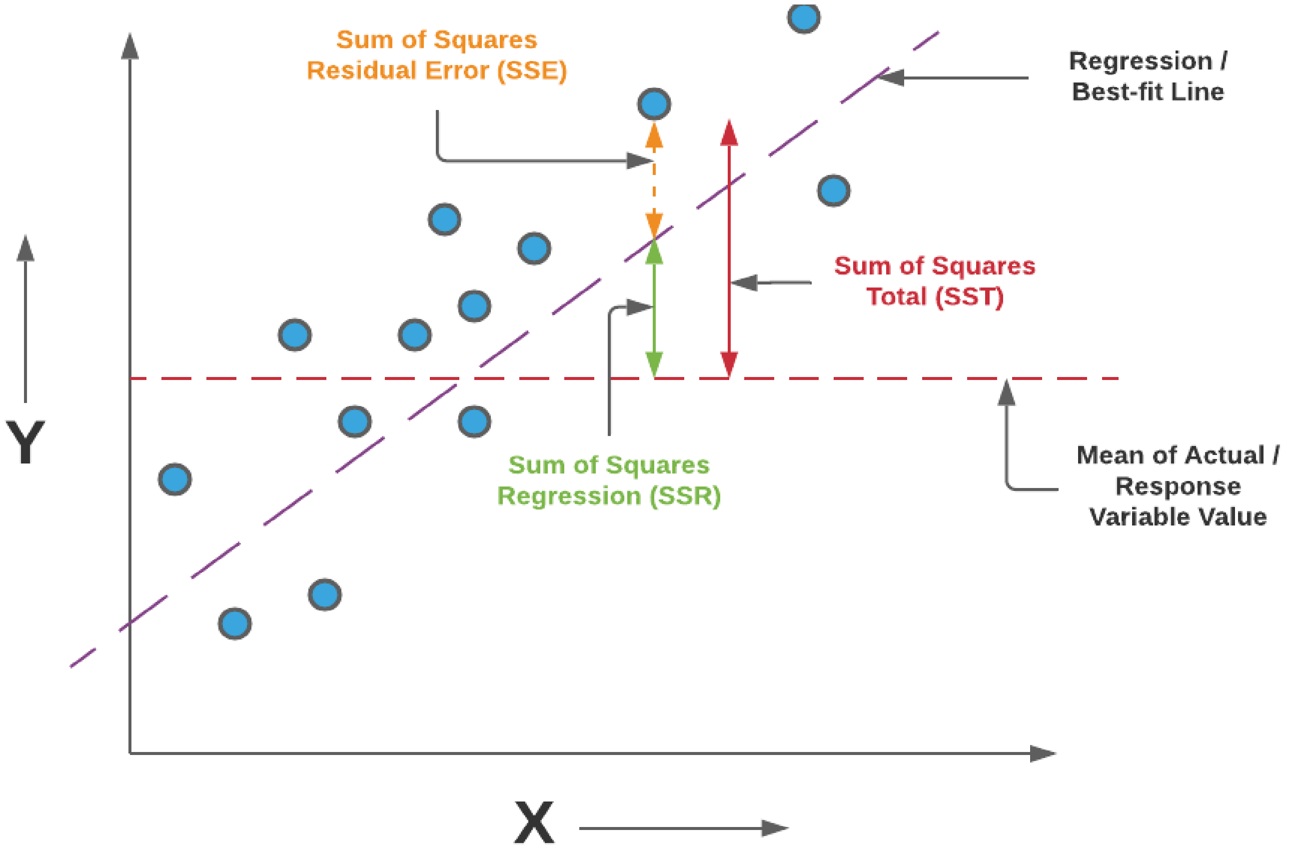
Parameters

The Parameter in deciding performance the model:

- R-squared is always between 0 and 100%:
  - 0% indicates that the model explains none of the variability of the response data around its mean.
  - 100% indicates that the model explains all the variability of the response data around its mean.
- The high value of R-square determines the less difference between the predicted values and actual values and hence represents a good model.

Graphical Representation of R-squared

Plotting fitted values by observed values graphically illustrates different R-squared values for regression models.



Key Limitations of R-squared

- R-squared cannot determine whether the coefficient estimates, and predictions are biased, which is why you must assess the residual plots.
- R-squared does not indicate whether a regression model is adequate. You can have a low R-squared value for a good model, or a high R-squared value for a model that does not fit the data!

Mean Squared Error

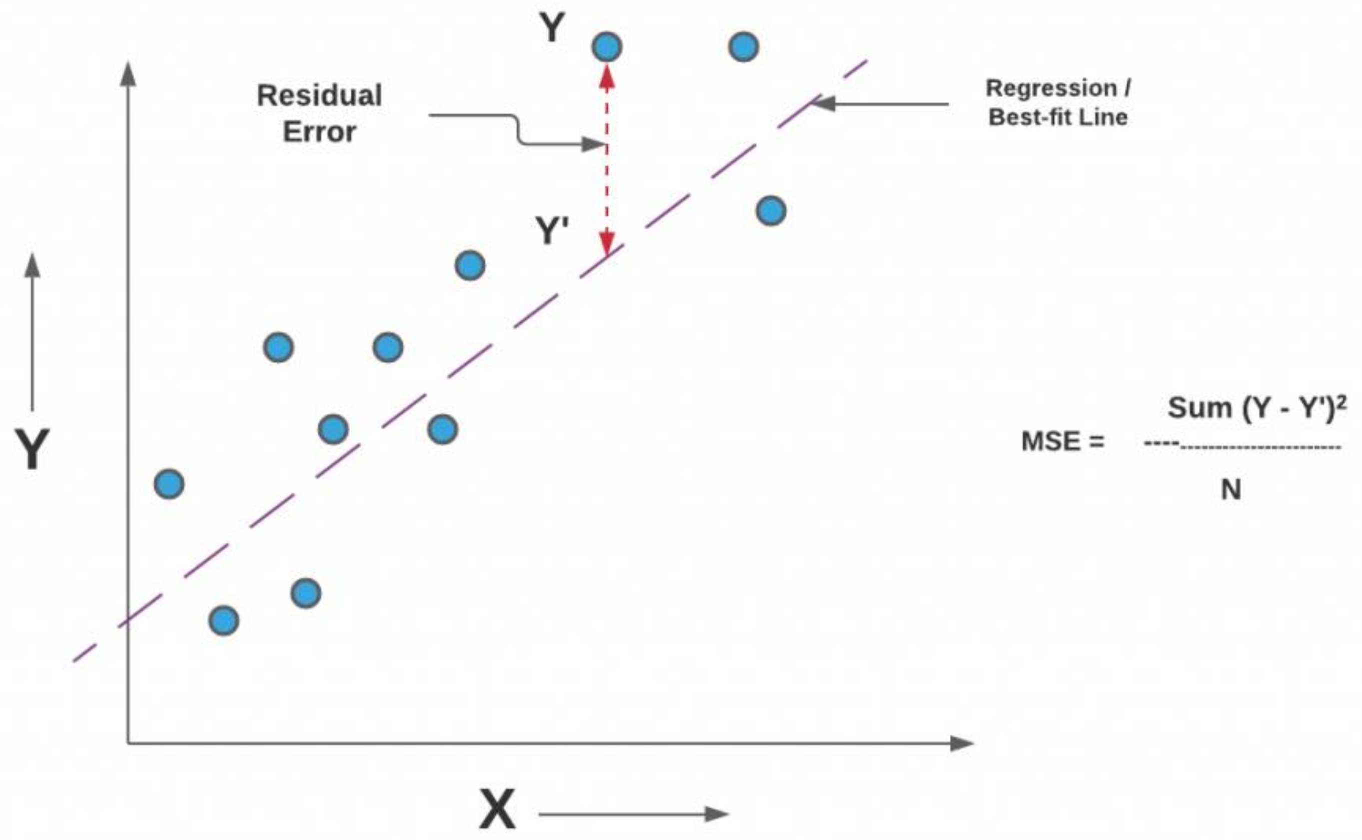
- The Mean Squared Error (MSE) or Mean Squared Deviation (MSD) of an estimator measures the average of error squares i.e., the average squared difference between the estimated values and true value.
- It is a risk function, corresponding to the expected value of the squared error loss.
- It is always non – negative and values close to zero are better.
- The MSE is the second moment of the error (about the origin) and thus incorporates both the variance of the estimator and its bias.

Formula

- MSE formula $= \left(\frac{1}{n}\right) \sum_{i=0}^{n} [actual - forecast]^2$
  - n = number of items,
  - Σ = summation notation,
  - Actual = original or observed y-value,
  - Forecast = y-value from regression

Graphical Representation of MSE

Suppose we have a set of values, and we start by drawing some regression line parameter sized by a random set of weight and bias value as before.



Conclusion

So, in conclusion we come out with what to use either mean square error (MSE) or R-Squared in the given condition:

- MSE represents the residual error which is nothing but sum of squared difference between actual values and the predicted / estimated values. R-Squared represents the fraction of response variance captured by the regression model.
- R-squared does not indicate whether a regression model is adequate. You can have a low R-squared value for a good model, or a high R-squared value for a model that does not fit the data!
- The disadvantage of using MSE is that the value of MSE varies based on whether the values of response variable is scaled or not. If scaled, MSE will be lower than the unscaled values.

THANK YOU