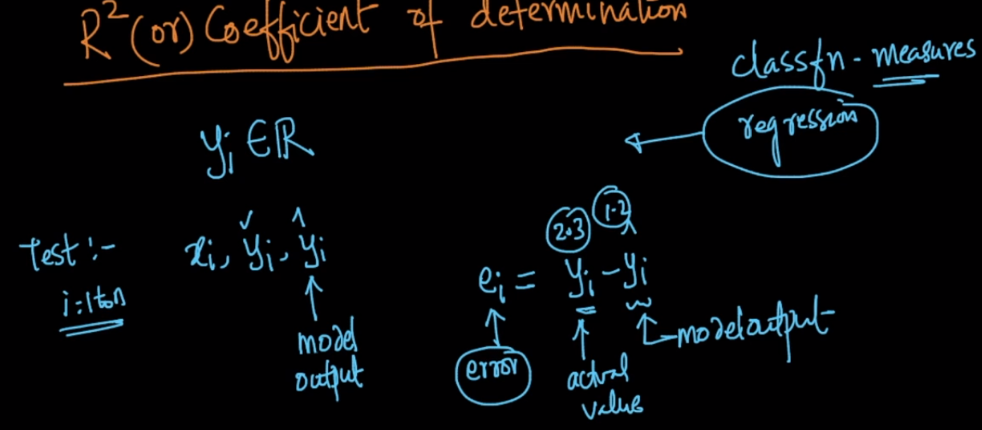
R-squared is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regression.

Here we are talking about Regression and not classification. So for regression, what would be the error value.

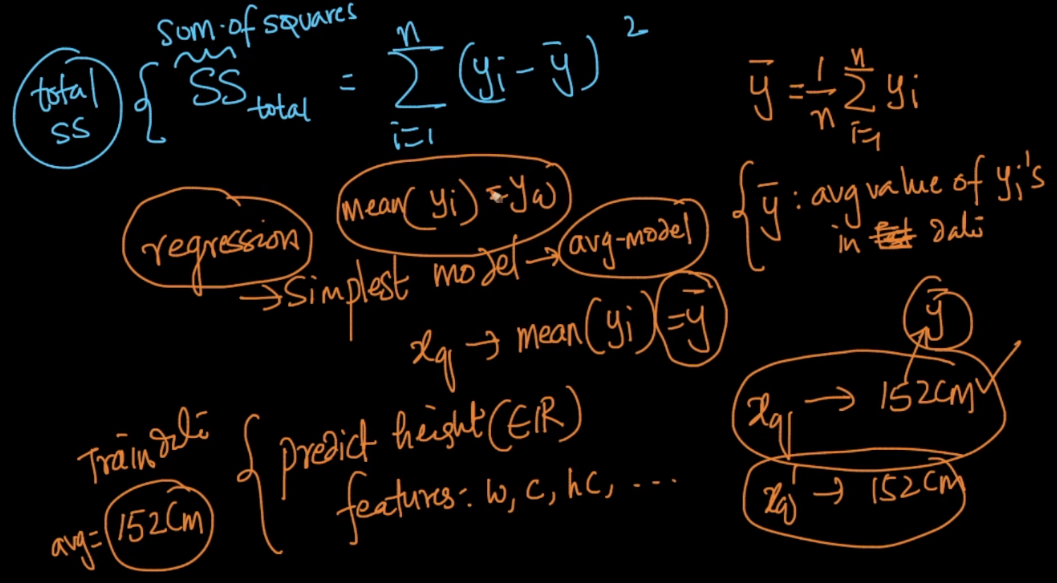
Error = actual value – predicted value.

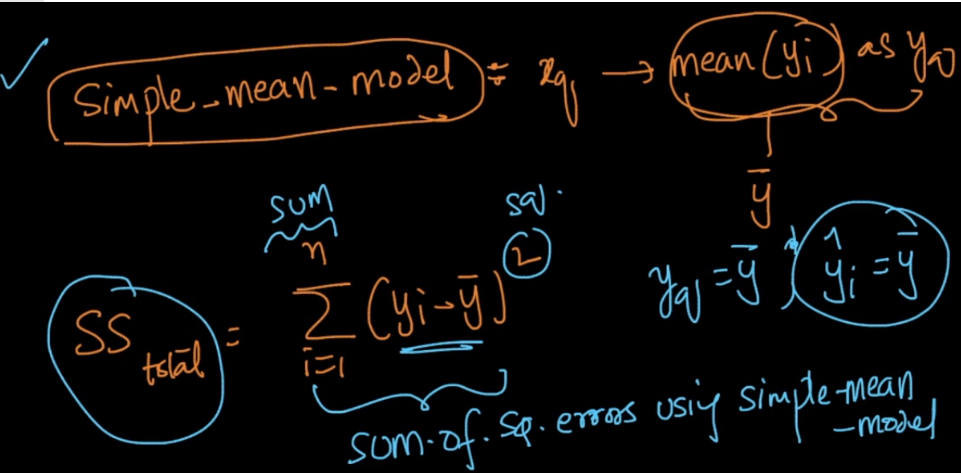


Before defining R2, let’s define two terms SStotal and SSresidue.

**SStotal:** Let’s we build a model, which for any new data point it always return the mean of the output we have in train data. We can say it as simple mean model.

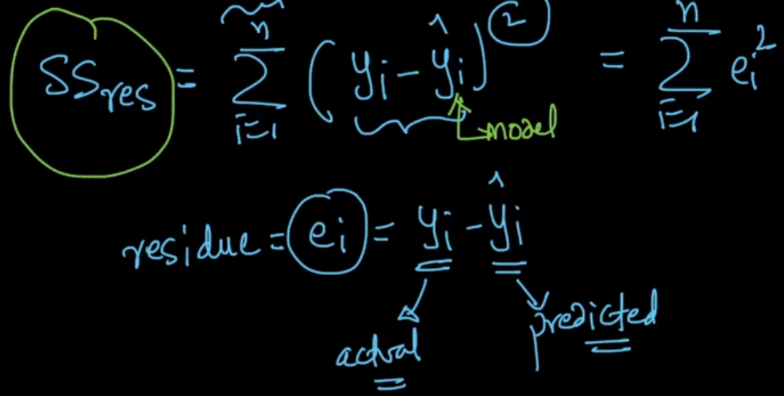
Let’s say the average value the model return is y bar, therefore SStotal ­is defined as given in below image where yi  is the actual output.





**SSres:** It’s defined as **average of the sum of square of error** where error = actual – predicted value.

It’s formula is given in below image.



Therefore R2 is defined as given in below image, and by this we can draw following conclusions:

**Case 1:**

If **SSres = 0,** that means error that our model generate is 0, in such case R2 = 1, which is the best case.

**Case 2:**

If **SSres < SStotal,** that means error that our model generate is less than the error generated by simple mean model, in such case R2 = 0 to 1

**Case 3:**

If **SSres = SStota,** that means error that our model generate is equal to the error generated by simple mean, in such case R2 = 0, which means that model is same as simple mean model.

**Case 4:**

If **SSres > SStotal,** that means error that our model generate is more than the error generated by simple mean model, in such case R2 = -ve, that means our model is worst than mean average model.

**Conclusion: The more the R2 value, the best our model is.**

<https://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit>

