**ADJUSTED R2**

Concept of R2 can also be applied to multiple linear regression.

y = b0 + b1x1 + b2x2

R2 – goodness of fit

(greater is better)

Problem When we have more than 2 variables

When you add a variable it will affect how the model looks like. We are trying to minimize ssres so that somehow the coefficient(b3) may help to reduce it, then R2 will increase.

But if this happens:

That whatever coefficient you give to your new variable, you cannot decrease ssres.

In such worst case scenarios, b3 is making the model completely worse thus we just put zero instead of this coefficient. Thus ssres and R2 won’t change.

b3 is rarelyzero because there is always some slight random co-relation between independent and dependent variable. Then R2 will slightly increase or decrease and you will not know if the new variable added is helping your model. R2 is biased, it is always increasing regardless of the actual improvement or non-improvement.

Thus adjusted R2 comes in.

p - number of regressors(number of independent variables)

n – sample size

p increases when number of independent variable increases

(n-p-1) decreases

((n-1) / (n-p-1)) increases

(1-R2)((n-1) / (n-p-1)) increases

1 - (1-R2)((n-1) / (n-p-1)) decreases

=> AdjR2 decreases(i.e going away from 1)