# Linear Regression

RSS = residual sum of squares

Deviations predicted from actual observed data; amount of variability left unexplained

TSS = total variance in the response, amount of variability in the response *before* regression

**Assessing model accuracy**

**R2 statistic:**

The amount of variability that is explained by the model (0,1)

# Multiple Linear Regression

A single variable regression may imply a relationship between the predictor and response, but multiple variable regression may imply the opposite

* **Correlation matrix** 🡪 may show one variable is a “surrogate” for another and gets “credit” for the relationship of another variable on the response
* **Example:** Ice cream sales and shark attacks
  + Hot temperatures bring more people to the beach (thus increasing ice cream sales)
  + Ice cream sales aren’t the reason for more shark attacks, hot temperatures are

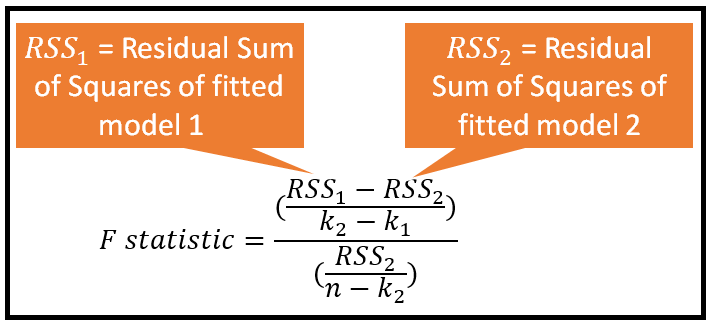
**F-statistic**

Assess multiple coefficients at the same time, will you be able to improve your regression model by adding more predictors?

H0 = all coefficients are equal to 0

H1 = there is at least one coefficient > 0

<https://towardsdatascience.com/fisher-test-for-regression-analysis-1e1687867259>

  
Difference in unexplained variance between two models/unexplained variance in more complex model

What fraction of the more complex model’s unexplained variance does the simpler model not explain?

F-statistic 🡪