Youtube Link: https://www.youtube.com/watch?v=2AQKmw14mHM

**R-squared tells us how well our regression line predicts the actual values**

R-squared is normally between 0 to 1 but it can be negative

**R2 is the percentage of variance explained by the relationship between two variables**

R2 = [Var(mean) – Var(line)] / Var(mean)

Var(mean) = difference of data points from average line

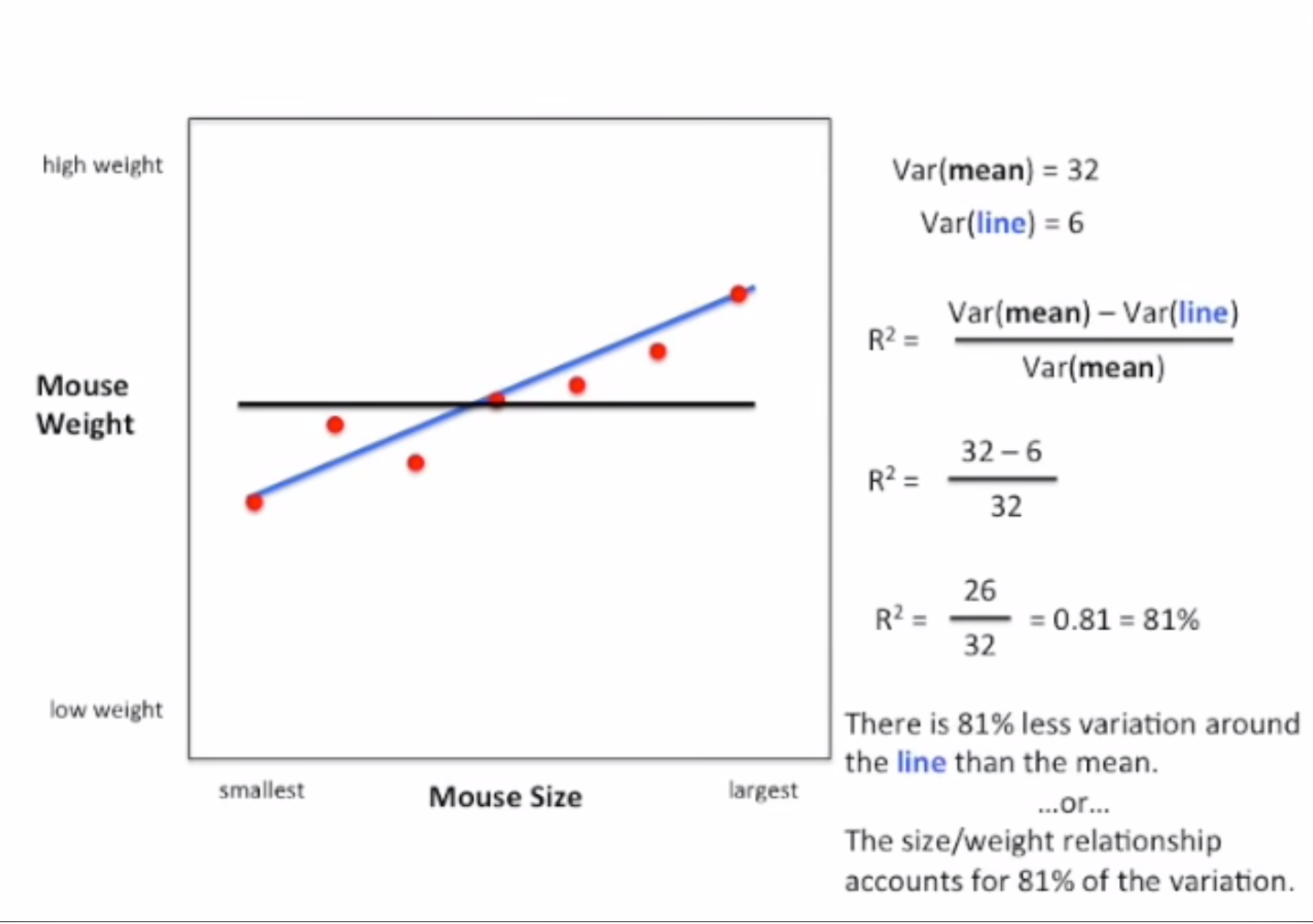
Var(line) = difference of data points from regression line

Interpretation: This formula makes R2 a percentage. Let’s assume R2 is 0.81.

This means there is **81%** less variation around the line than the mean

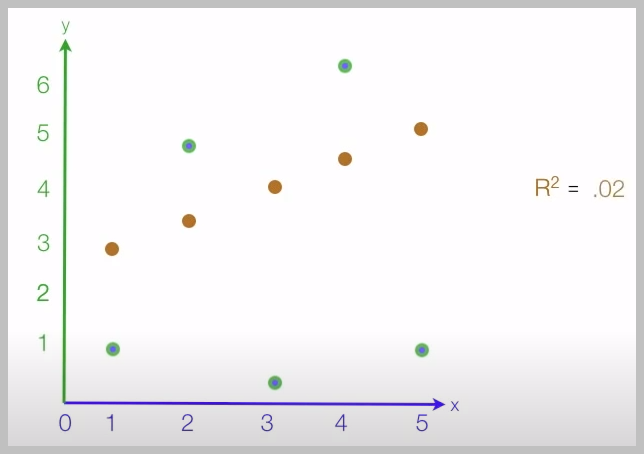
OR

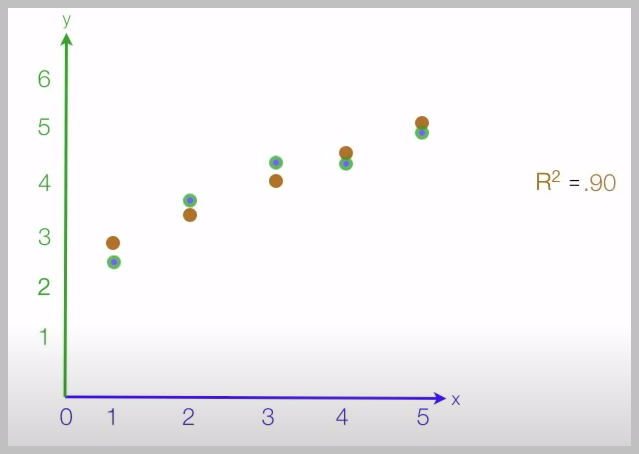
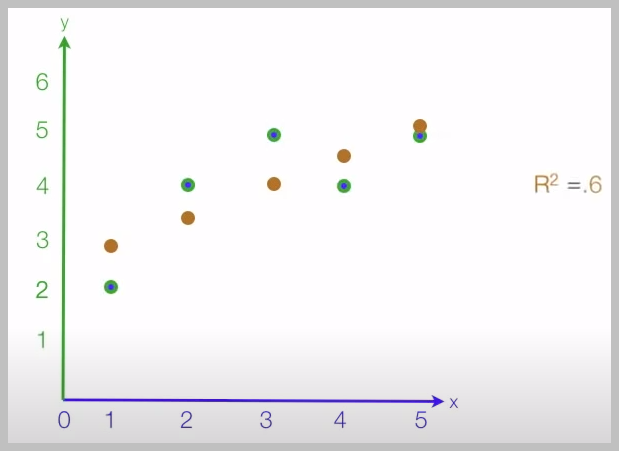
The 2 variables’ relationship accounts for **81%** of the variation.

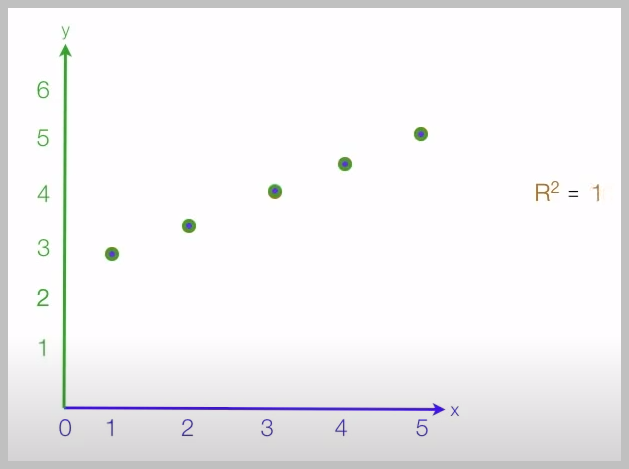


Why do we square the points at each step: Because otherwise the sum of (point – average) will always be 0. That’s what the whole concept of average is. Middle value of all points

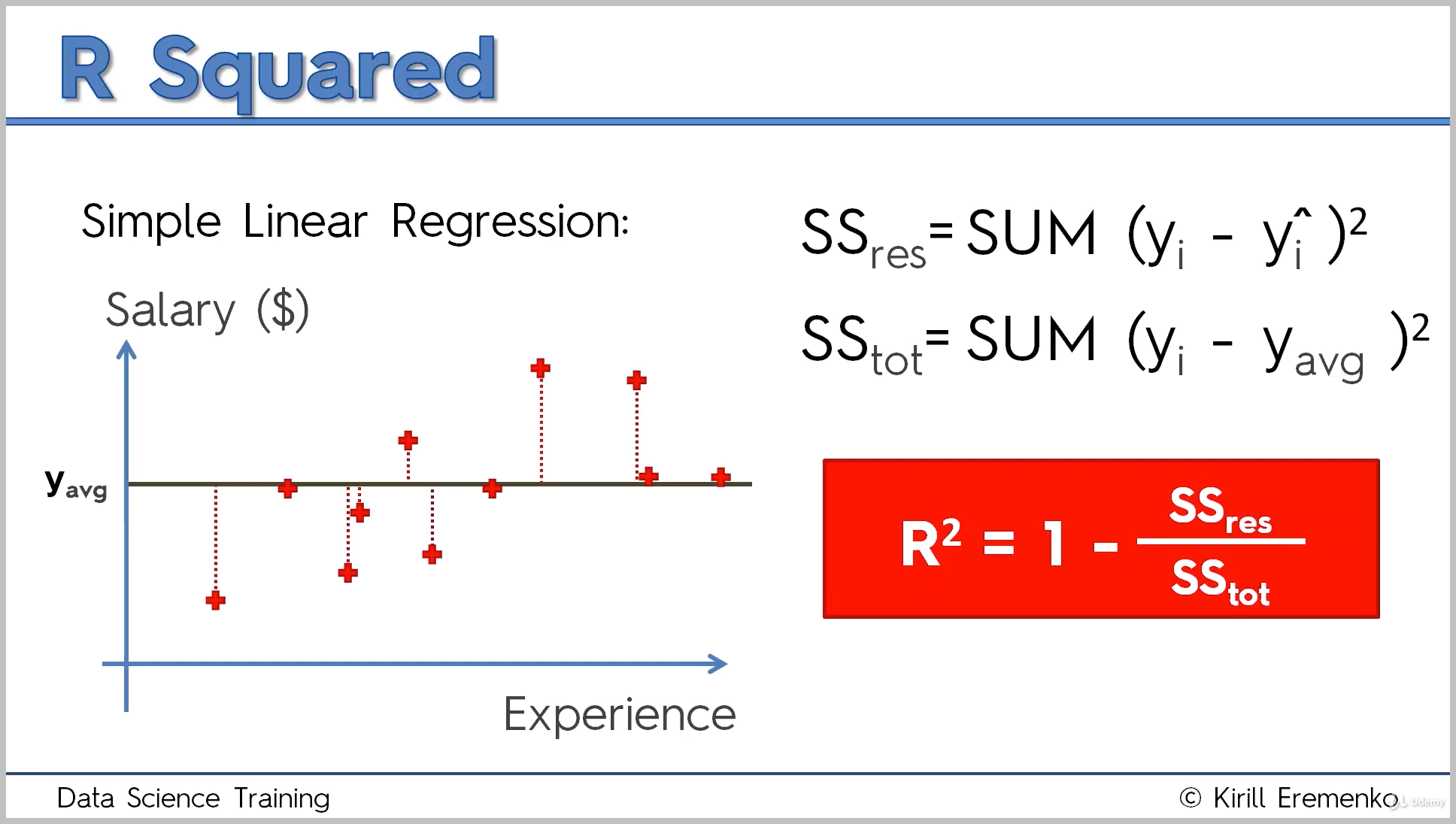
Brown: Predicted Line. Green: Actual Points

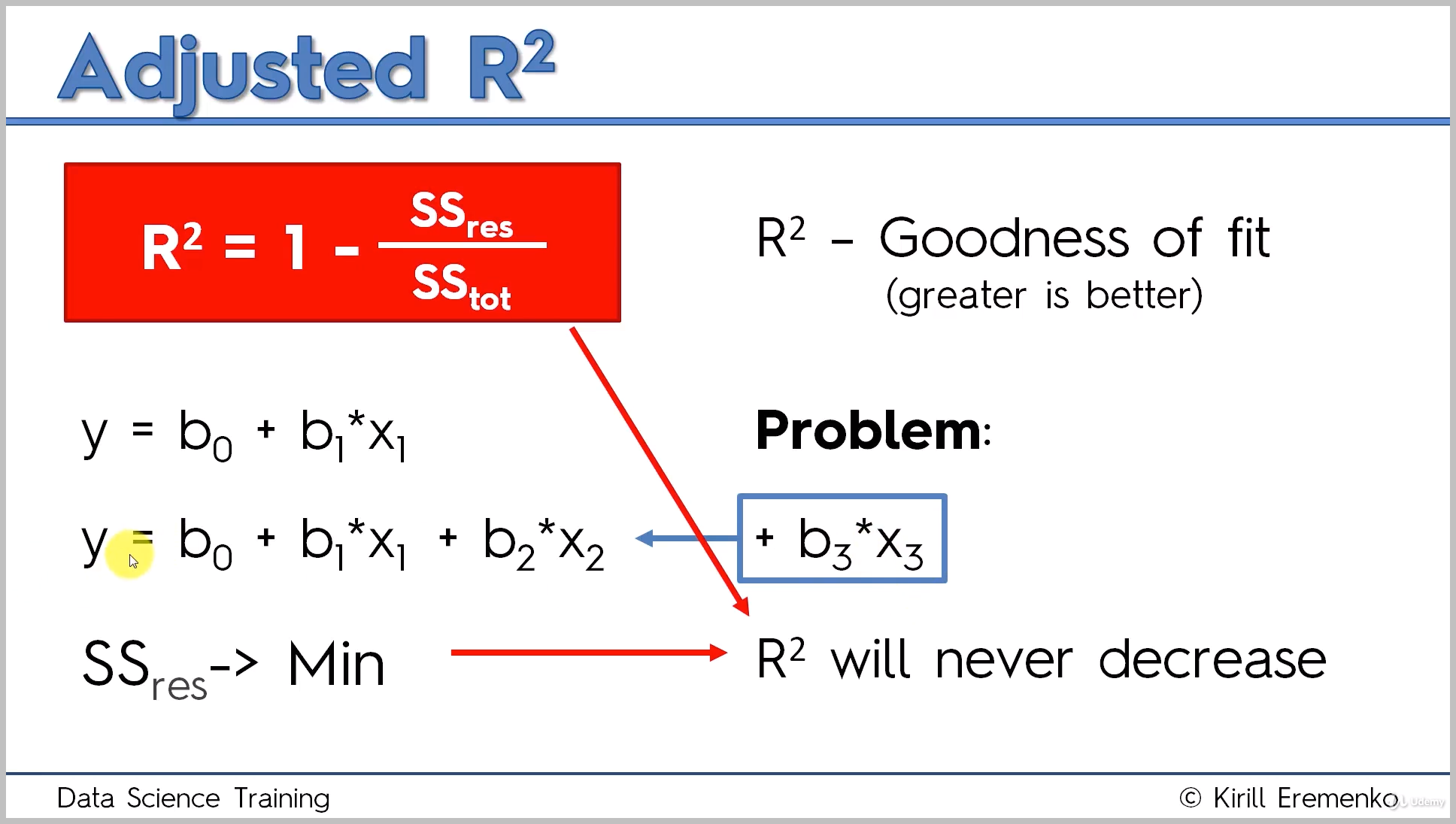






Note: To get value of R2, we only need predicted points and actual points. Nothing else.





P: number of independent variables

