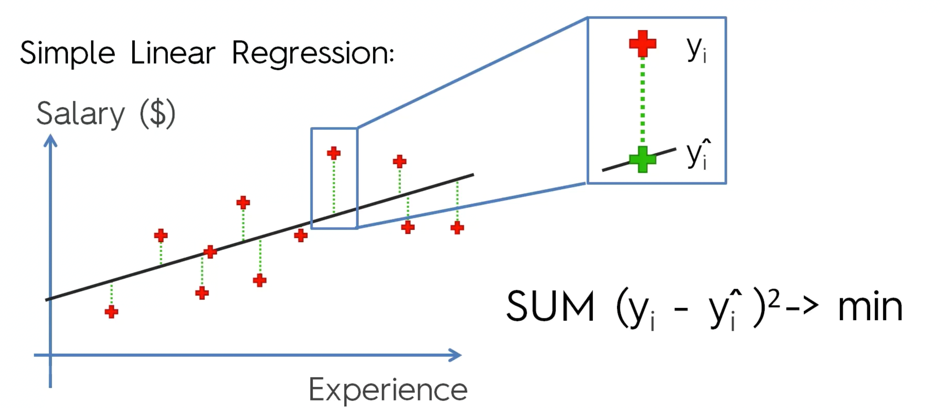
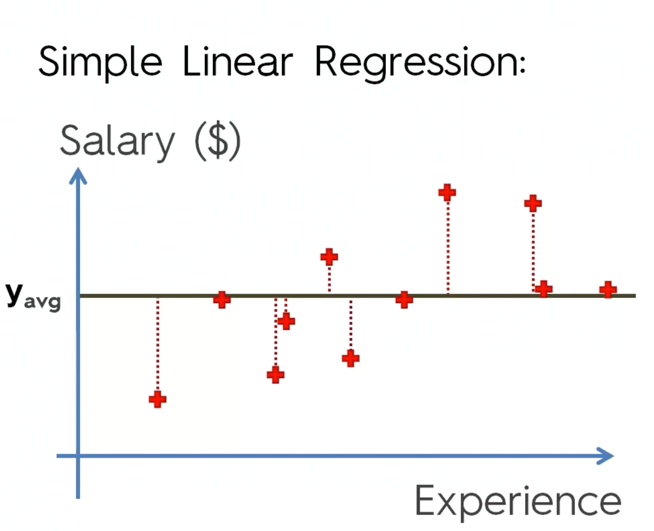
**R-SQUARED**

* SUM (Squares of Residuals) [SSres] = SUM (yi – yi^)2 (sum of squared difference between the actual value and predicted value)



* SStot (Total Sum of Squares) = SUM (yi – yavg)2 (Sum of squared difference between the actual value and the average value)



* R2 =
* So, what it is saying is that there is always going to be a Sum of Total Squares (Not all your values will equal to the average of the data set), and what we are trying to do with our regression is fit a line for the regression to minimize the Sum of Residual Values to make it as small as possible.
* To think of it that way, the average line on the chart is also a type of a trend line, it is just horizontal. You can think of it as model that is fit to the model. Not the best model, but it’s just y avg.
* So, what you are trying to do by fitting a slopped line and minimizing the sum of squared residuals is you are trying to fit the best line.
* And what R-squared is telling us is - how good is your line, compared to the average line, which anybody can think of.
* You always try to minimize SS Residuals, and as you try to minimize it, it always grows smaller.
* And that way SSRes provided by SSTot becomes smaller, and 1 – (SSRes/SSTot) becomes greater.
* So ideally, if your SSRes is 0, so basically your trend line that your modelling goes through all your records then in that case R2 = 1. And that’s the ideal scenario. But that normally never happens.
* Therefore, the closer R2 is to 1, the better.
* R2 can also be negative, if your SSRes fits your data worse than your average line.
* Normally R2 is between 0 and 1. The closer it is to 1 the better.