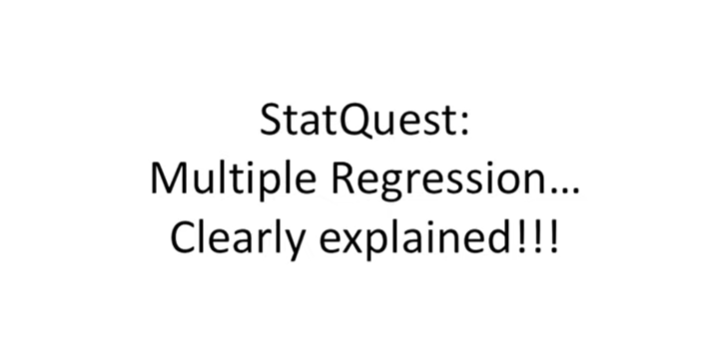
<https://www.youtube.com/watch?v=zITIFTsivN8&list=PLblh5JKOoLUIzaEkCLIUxQFjPIlapw8nU&index=4>



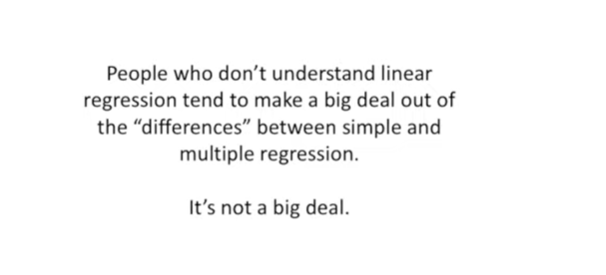
Hello I'm Josh stommer and welcome to stack quest.

Stack quest is brought to you by the friendly folks in the genetics department at the University of North Carolina at Chapel Hill.

Today we're gonna be talking about multiple regression and it's gonna be really strange.

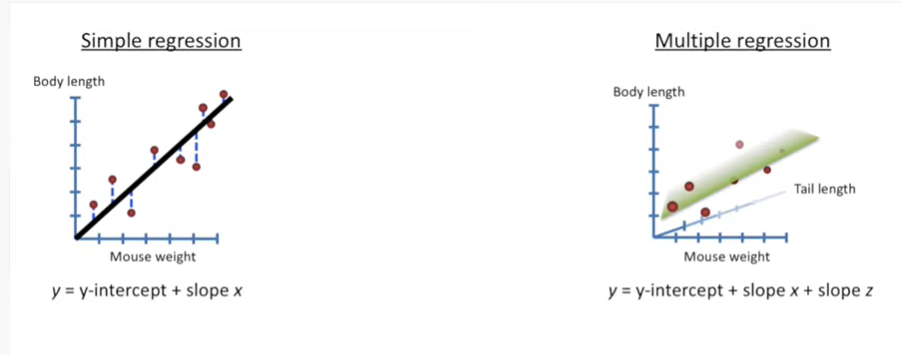
This stack was two builds on the one for linear regression.

So if you haven't already seen that one yet check it out.



Alright now let's get to it people who don't understand linear regression tend to make a big deal out of the differences between simple and multiple regression.

It's not a big deal in a static quest on simple linear regression already covered most of the concepts we're going to cover here.



You might recall from the stat quest on the linear regression it's simple regression is just fitting a line to data.

We're interested in the r-squared and the p-value to evaluate how well that line fits the data.

In that same stat quest I also showed you how to fit a plane to data well that's what multiple regression is you fit a plane or some higher dimensional object to your data.

A term like higher dimensional object sounds really fancy and complicated but it's not.

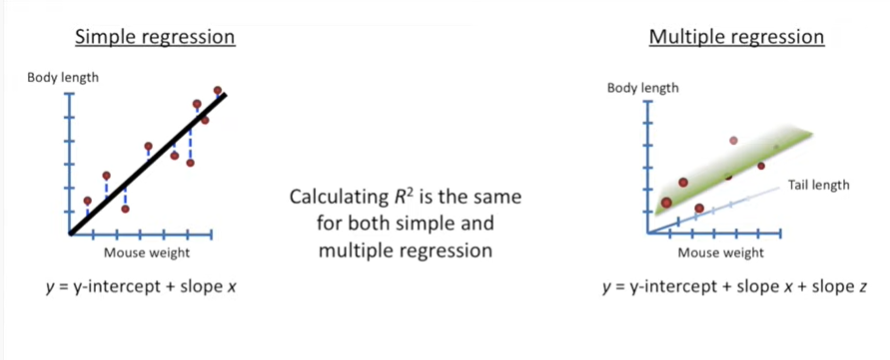
All it means is that we're adding additional data to the model.

In the previous example all that meant was that instead of just modeling body lengths by mouse weight.

We modeled body lengths using mouse weight and tail length.

If we added additional factors like the amount of food eaten or the amount of time spent running on a wheel, while those would be considered additional dimensions.

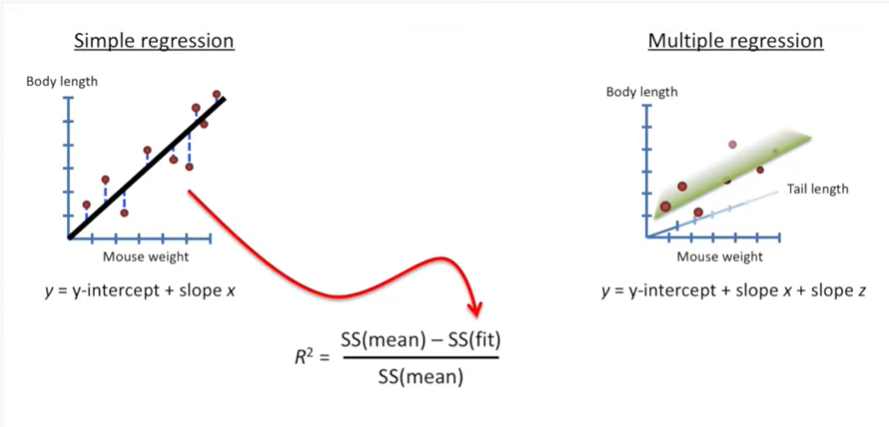
But they're really just additional pieces of data that we can add to our fancy equation.



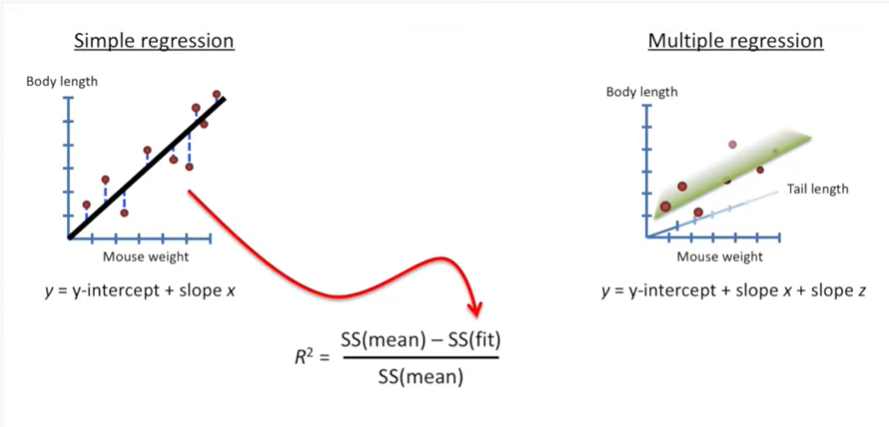
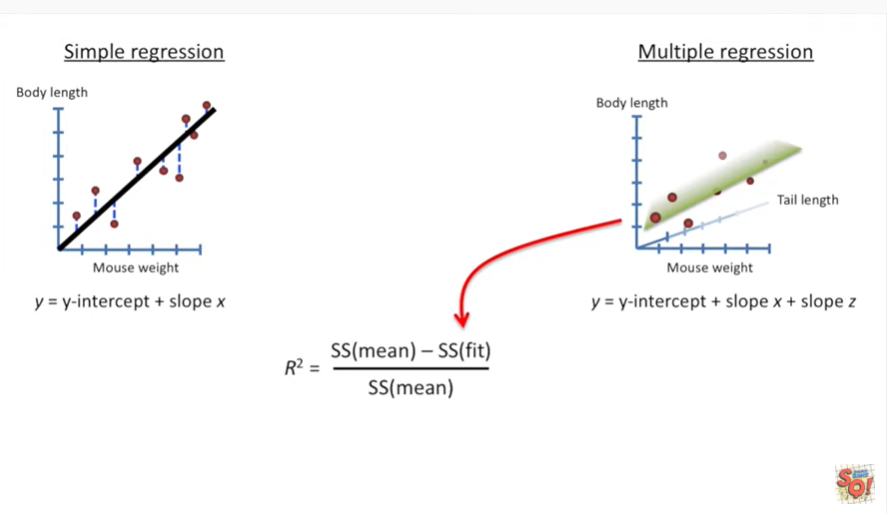
So from the stat quest on linear regression you may remember the first thing we did was calculate R squared.

Well the good news is calculating r squared is the exact same for both simple regression and multiple regression.

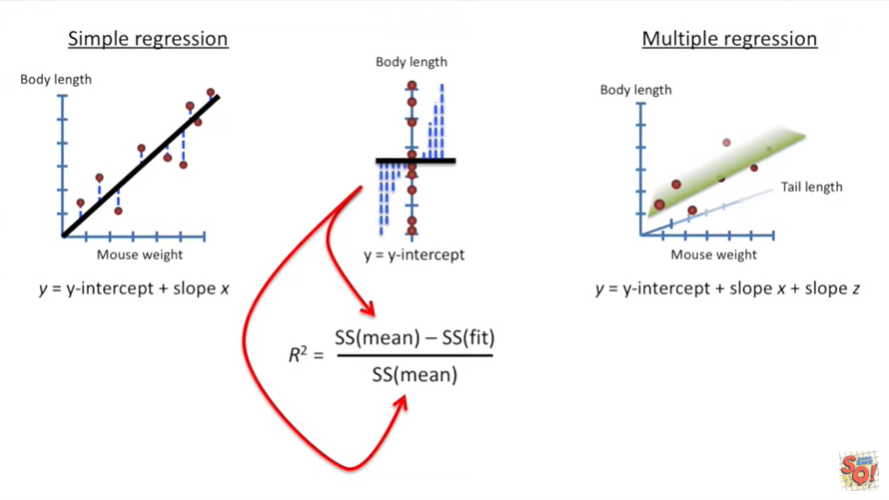
There's absolutely no difference.



Here's the equation for R squared.

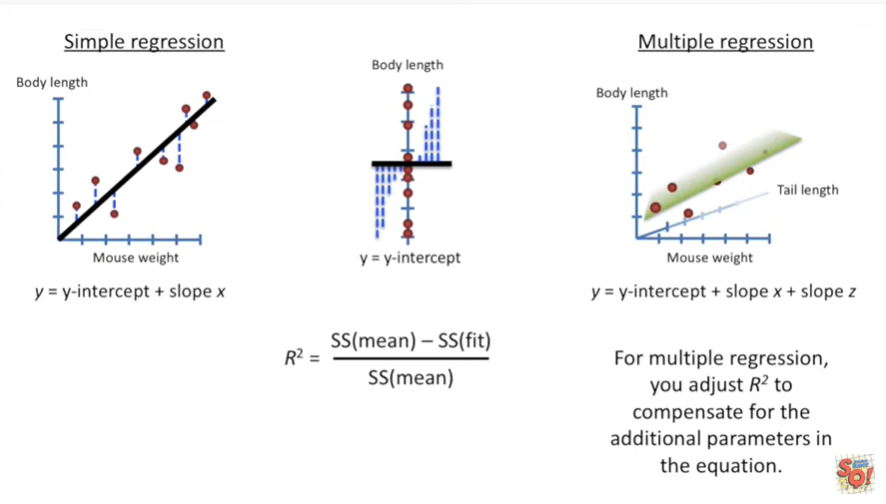


And we plug in the values for the sums of squares around the fit.



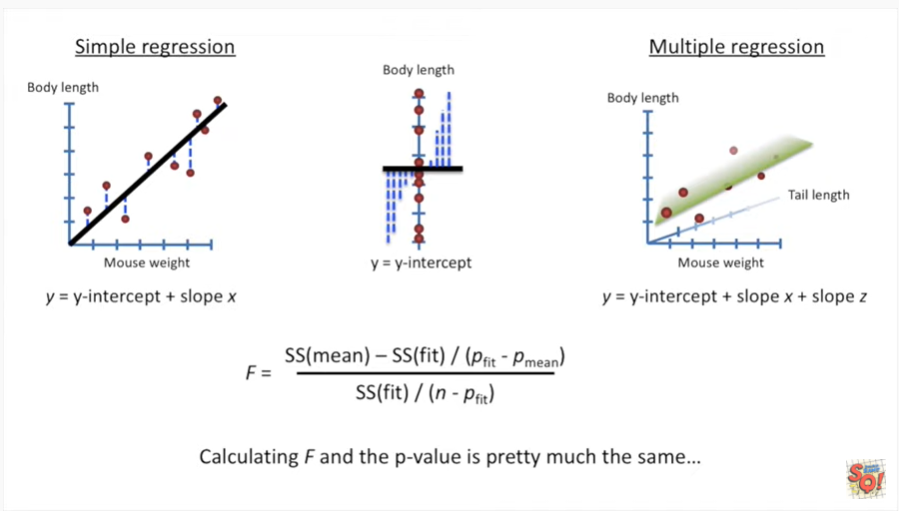
And then we plug in the sums of squares around the mean value for the body length.

Regardless of how much additional data we add to our fancy equation if we're using it to predict body lengths then we use the sums of squares around the body length.



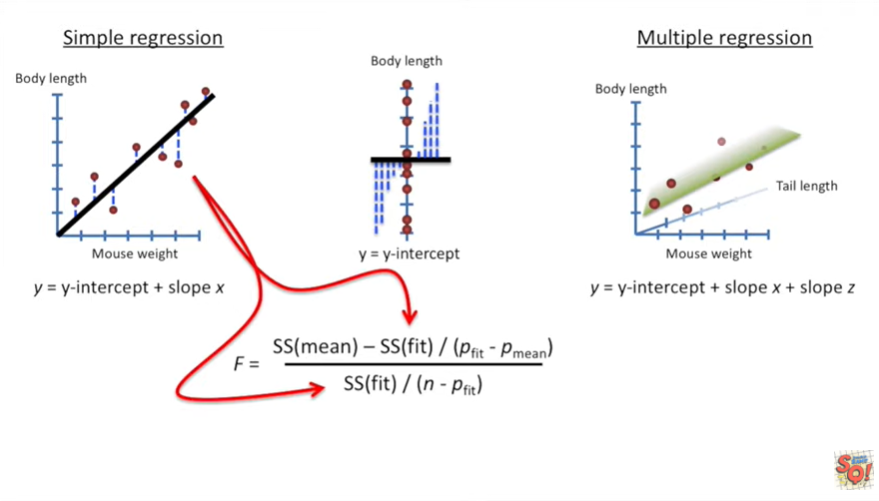
One caveat is for multiple regression you adjust r-squared to compensate for the additional parameters in the equation.

We covered this in the stat quest for linear regression so it's no big deal.

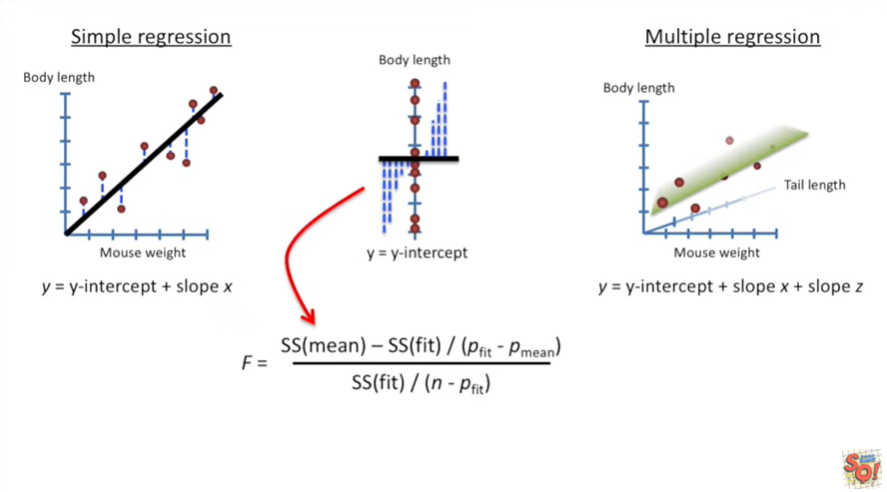


Now we want to calculate a p-value for our r-squared.

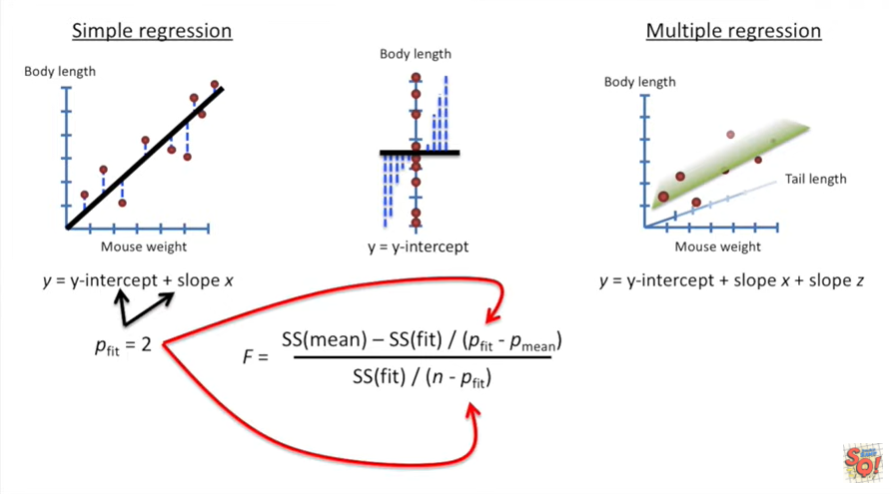
Calculating F and the p-value is pretty much the same.



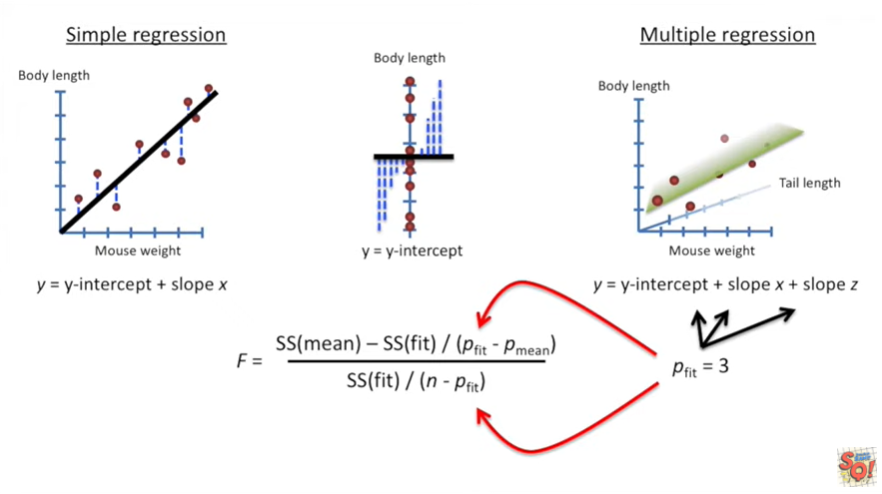
You plug in the sums of squares around the fit



and then you plug in the sums of squares around the mean.

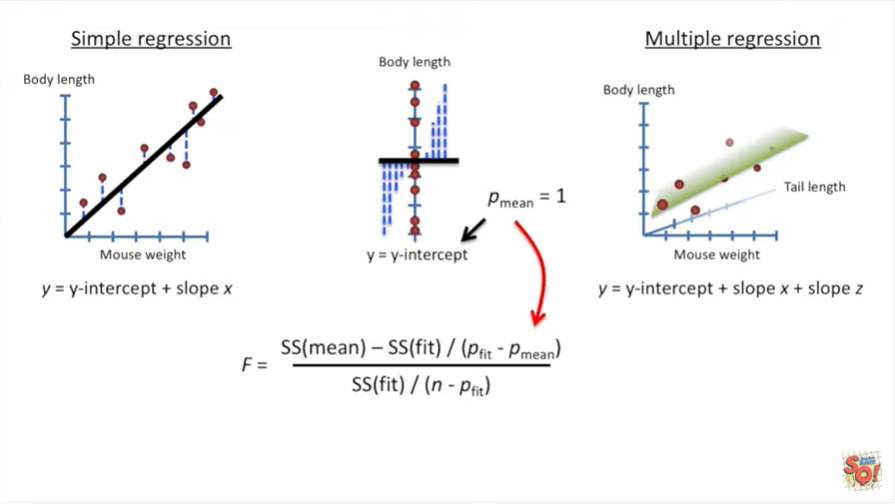


For simple regression P fit equals 2 because we have two parameters in the equation that least squares has to estimate .

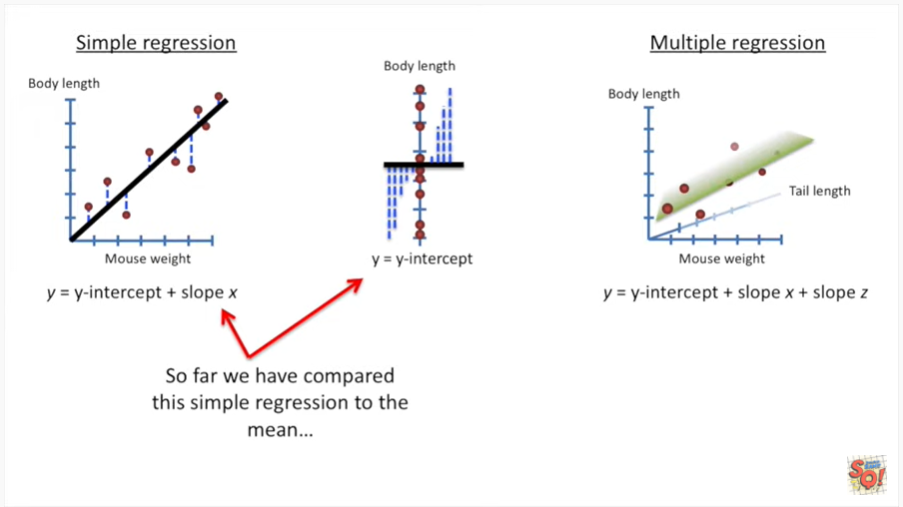


And for this specific example the multiple regression version of P fit equals 3 because least squares had to estimate three different parameters.

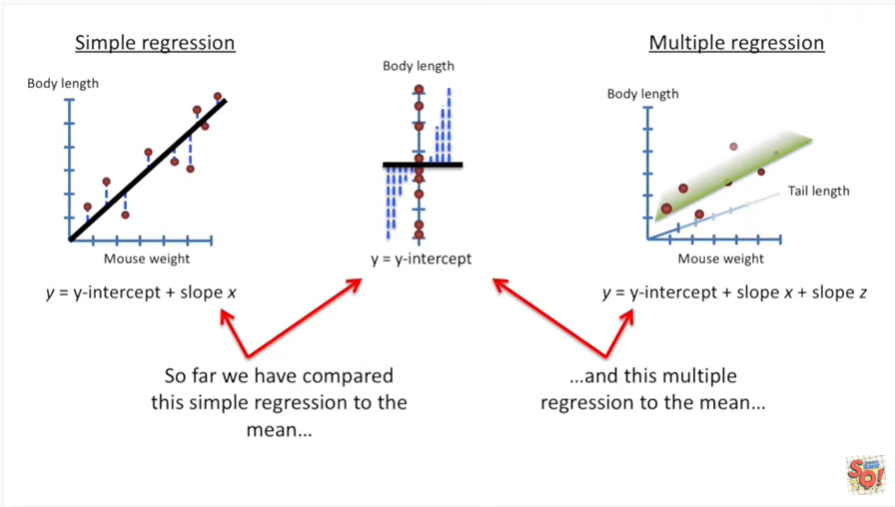
If we added additional data to the model for example the amount of time a mouse spins running on a wheel then we have to change P fit to equal the number of parameters in our new equation.



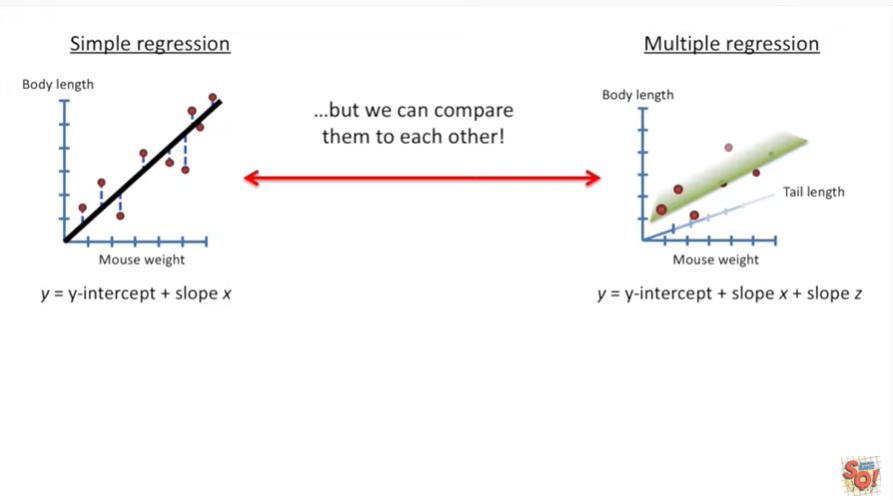
And for both simple regression and multiple regression P mean equals 1 because we only have to estimate the mean value of the body length.



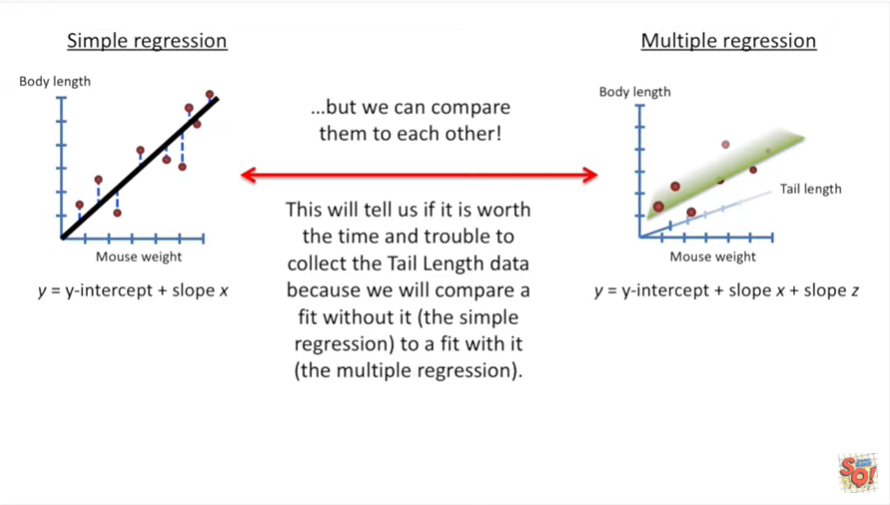
So far we have compared this simple regression to the mean.



And this multiple regression to the mean.

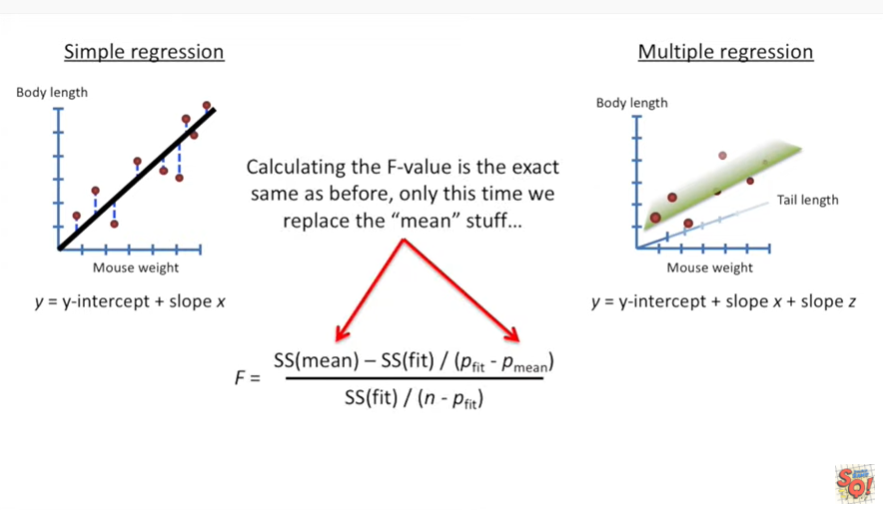


But we can compare them to each other.

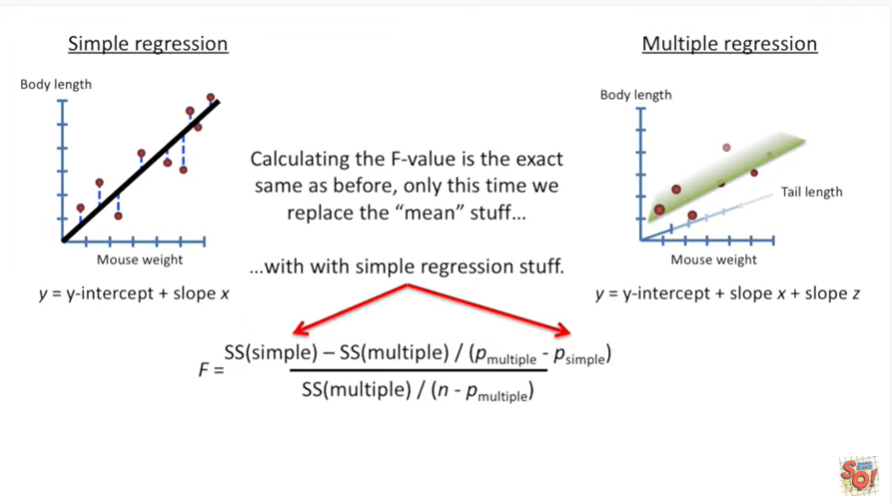


And this is where multiple regression really starts to shine.

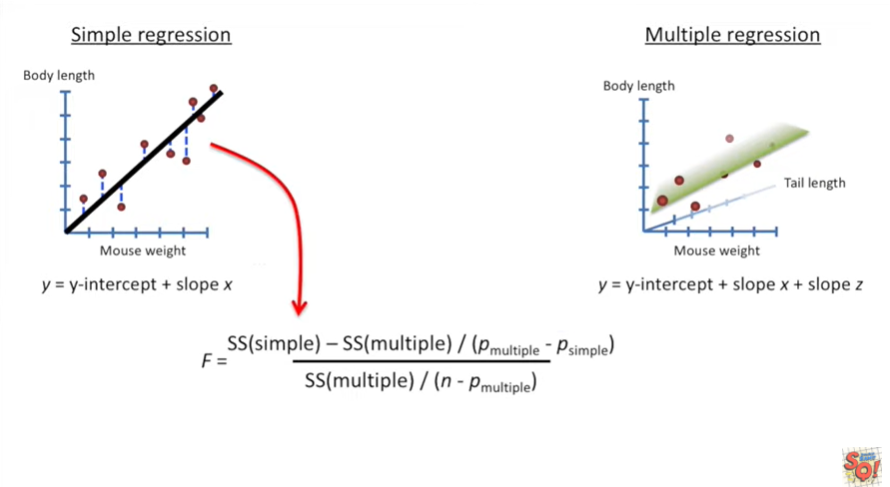
This will tell us if it's worth the time and trouble to collect the tail length data because we will compare a fit without it the simple regression to a fit with it the multiple regression.



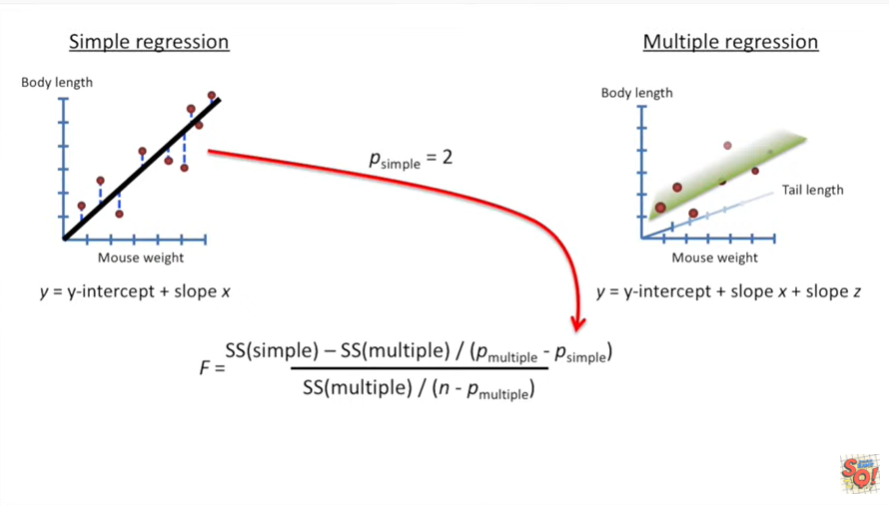
Calculating the F value is the exact same as before.



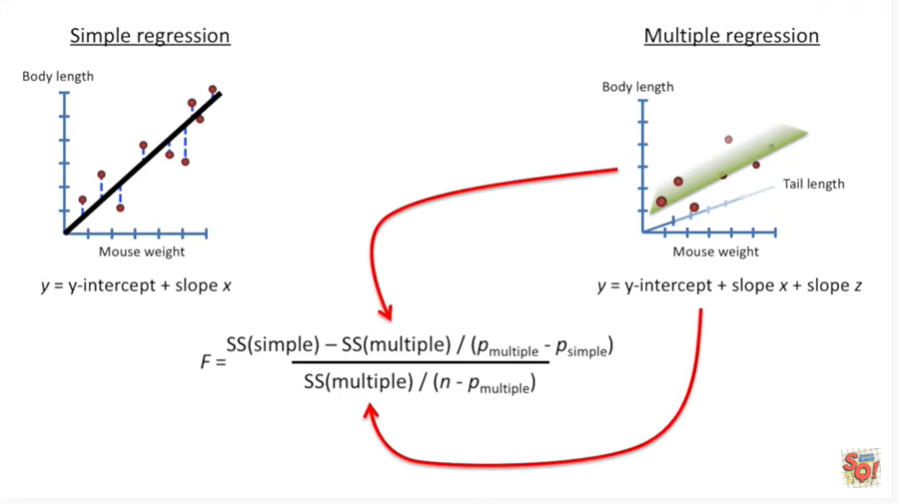
Only this time we replace the mean stuff with the simple regression stuff.



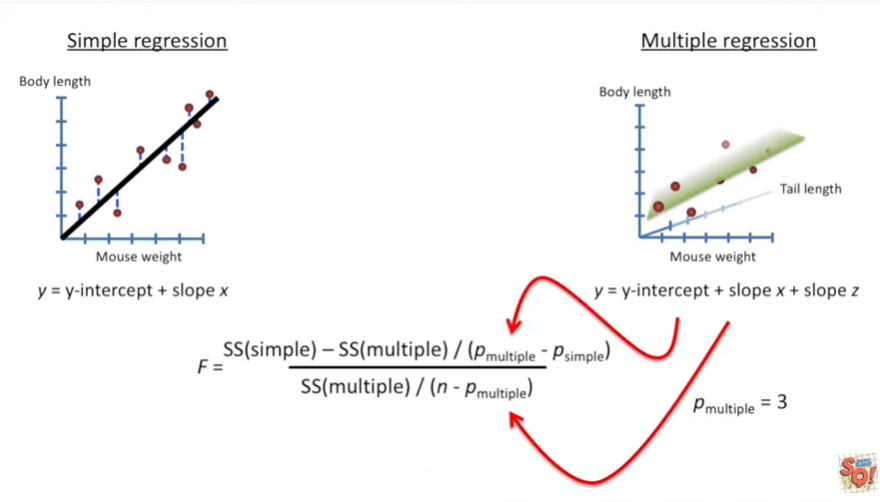
So instead of plugging in the sums of squares around the mean, we plug in the sums of squares around the simple regression.



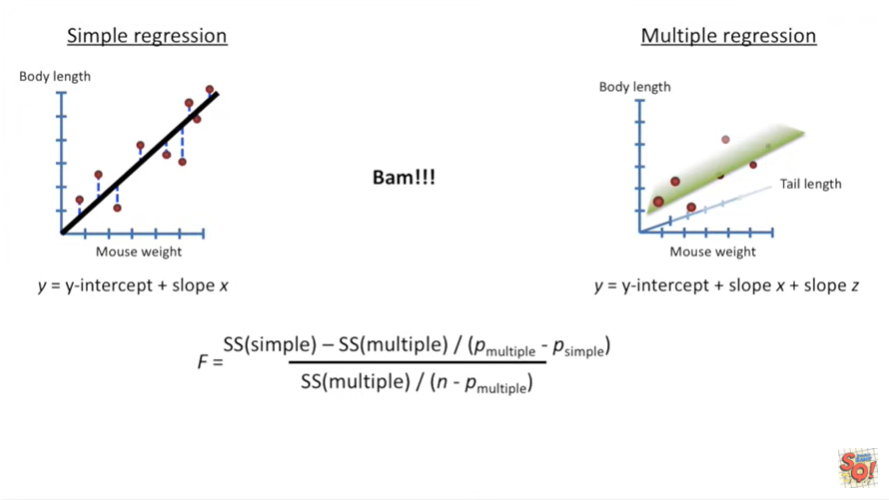
And instead of plugging in P mean we plug in P simple which equals the number of parameters in the simple regression that's 2.



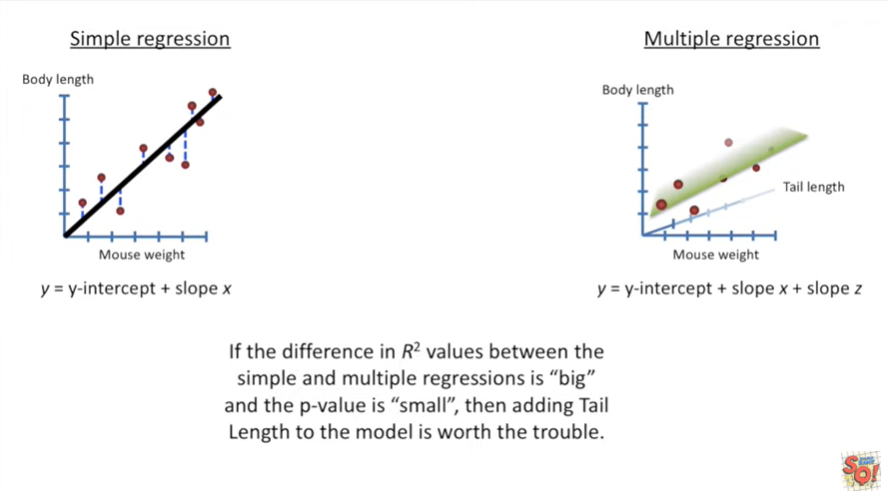
And then we plug in the sums of squares for the multiple regression.



And we plug in the number of parameters in our multiple regression equation.



BAM.



If the difference in our squared values between the simple and multiple regression is big and the p value is small then adjusting tail length to the model is the trouble.