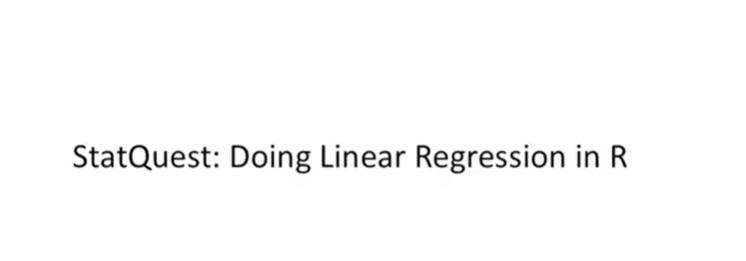
<https://www.youtube.com/watch?v=u1cc1r_Y7M0&list=PLblh5JKOoLUIzaEkCLIUxQFjPIlapw8nU&index=3>



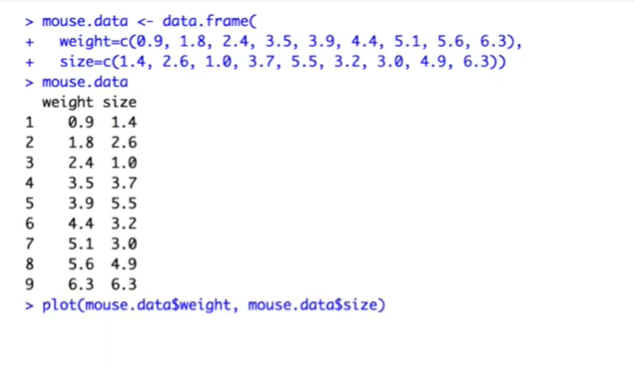
Hello and welcome to stack West.

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

Today we're going to be talking about doing linear regression in R.

This particular stack West is intended to be a companion video for the stat quest on linear regression.

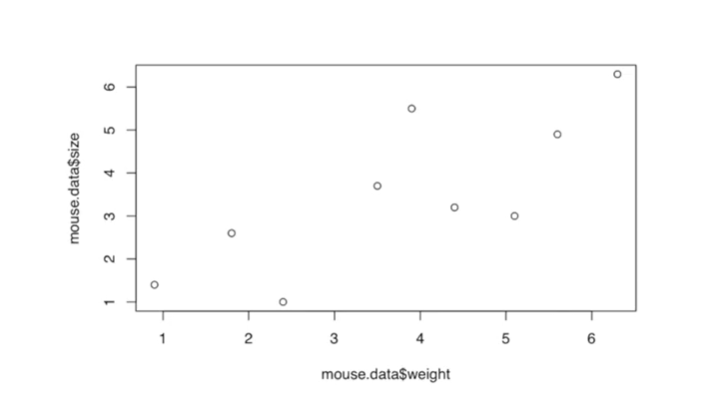
For this tutorial I'm going to assume that you already know how to get your data into R and instead I'm going to focus on how to get that data into a linear regression model and how to interpret the results.



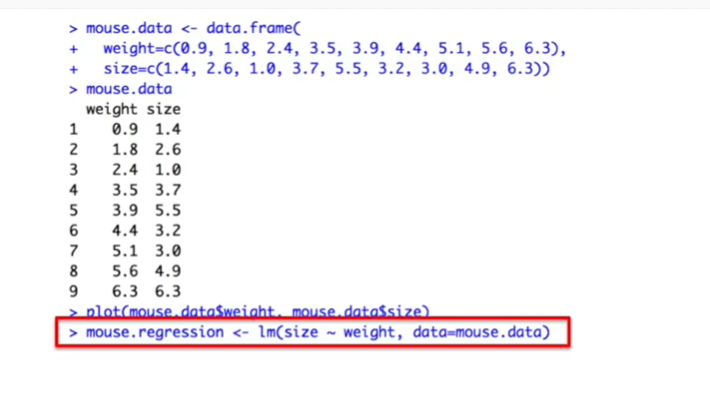
This is how I created the data for the stat quest on linear regression.

I created a data frame with two columns weight and size.

If I just type Mouse data and then press return, R will print out the data frame and a nice column format.

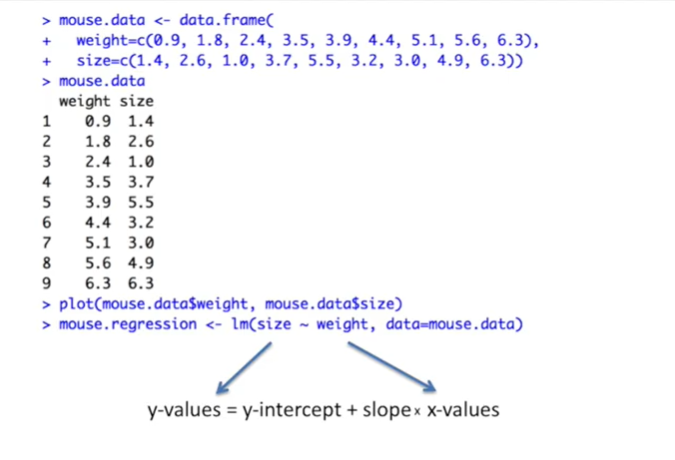


I then use the plot function to plot the data on an XY graph.

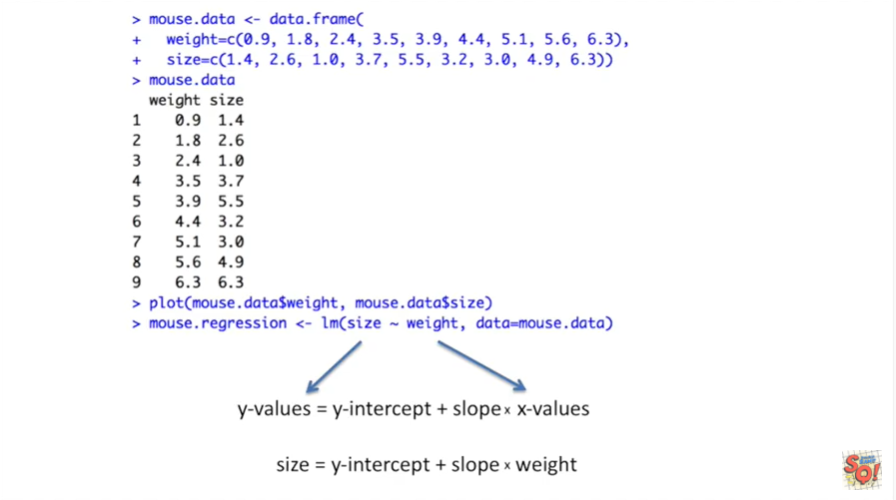


This is where I set up the actual linear regression.

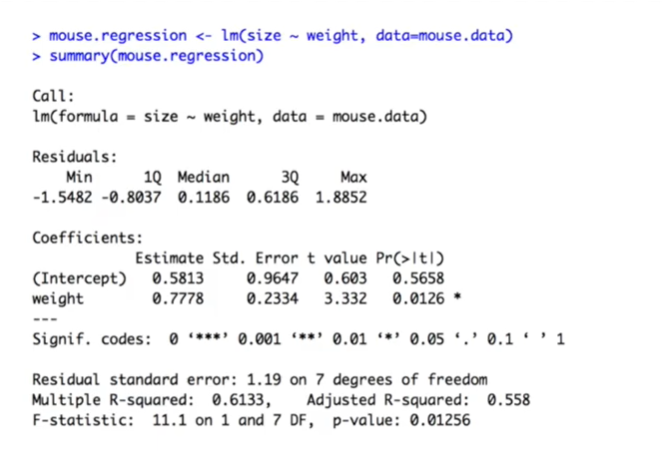
I call the function L M which stands for linear models and I pass it a formula and I pass it the mouse data.



The way I've specified the formula means that size are considered to be the Y values and weight are considered to be the X values.

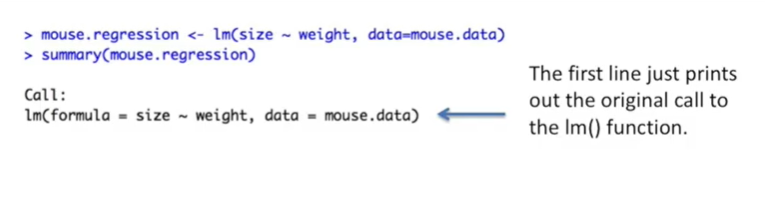


The linear models function then calculates the least squares estimates for the y-intercept and the slope.

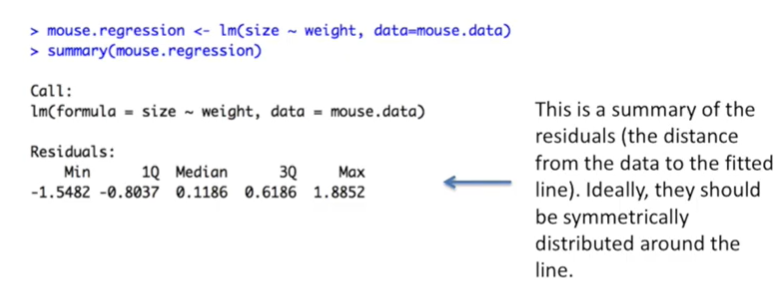


In R the meat of doing or regression is in the summary function.

This function generates all kinds of output and I'm going to walk through it one step at a time.



The first line just prints out the original call to the LM or linear models function.



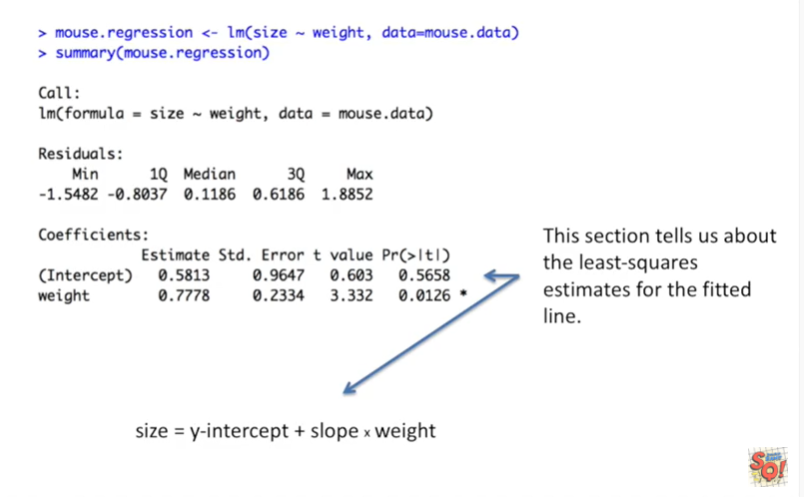
After that you get a summary of the residuals.

Those are the distance from the data to the fitted line.

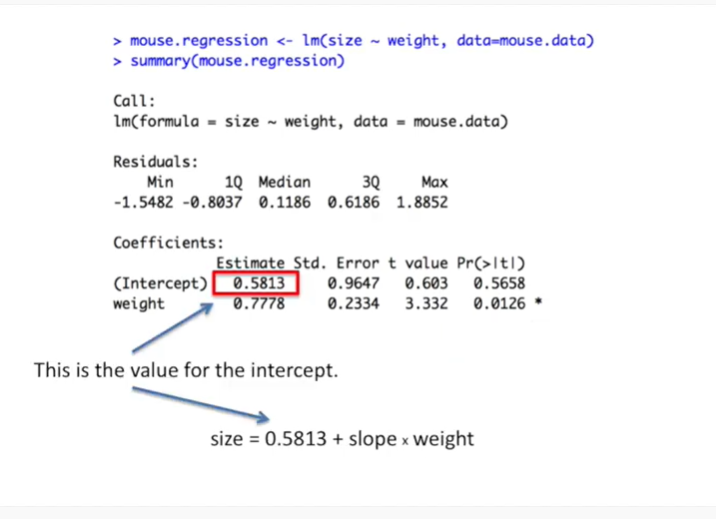
Ideally they should be symmetrically distributed about the line that means you want the min value and the max value to be approximately the same distance from 0.

Likewise you'd like the first quantile or 1q and the third quanta or 3q to be equidistant from 0.

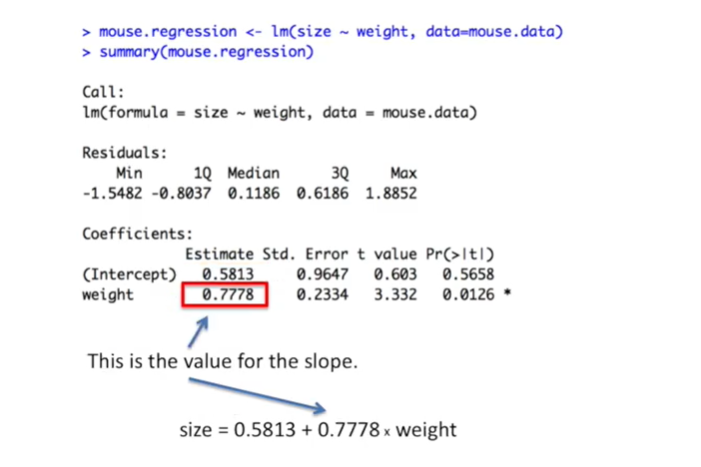
Also it's nice to have the median close to 0 as well.



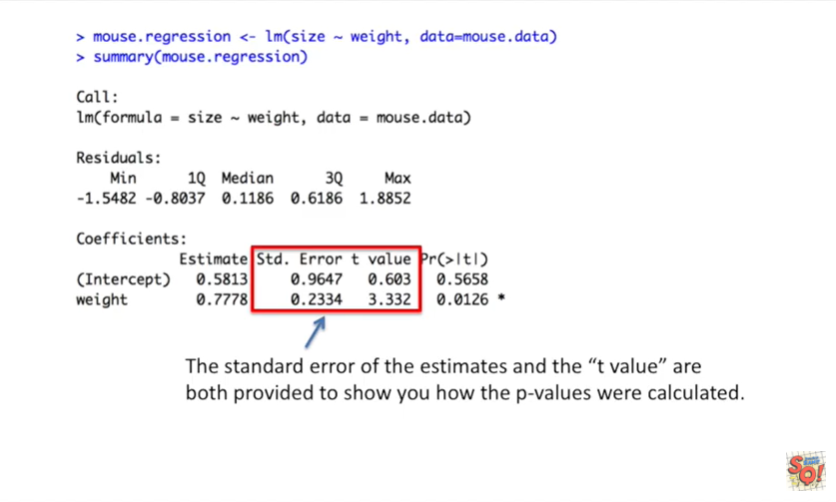
This next section tells us about the least squares estimates for the fitted line.



This value is for the intercept.



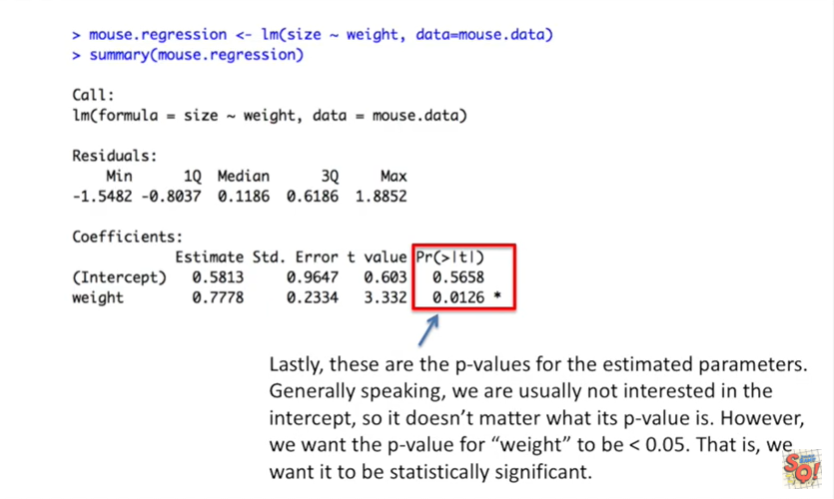
And this value is for the slope.



The standard error of the estimates and the T value are both provided to show you how the P values were calculated.

These P values test whether the estimates for the intercept and the slope are equal to 0 or not.

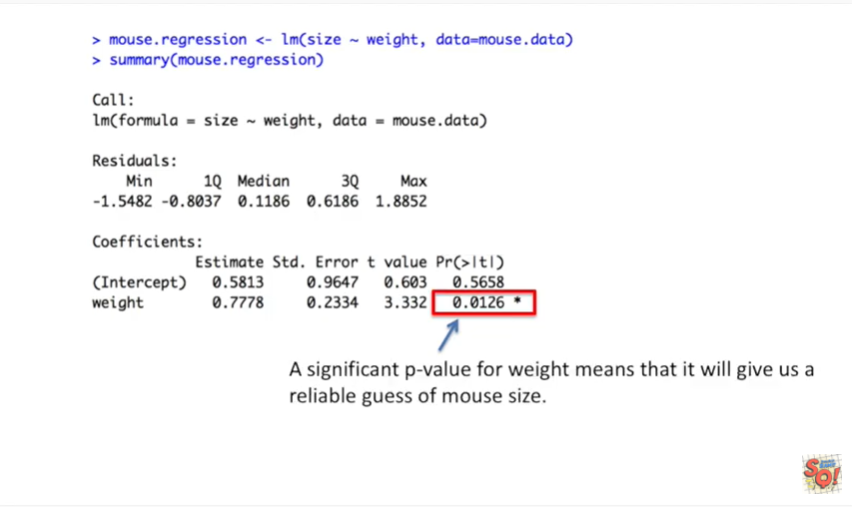
If they're equal to 0 that means they don't have much use in the model.



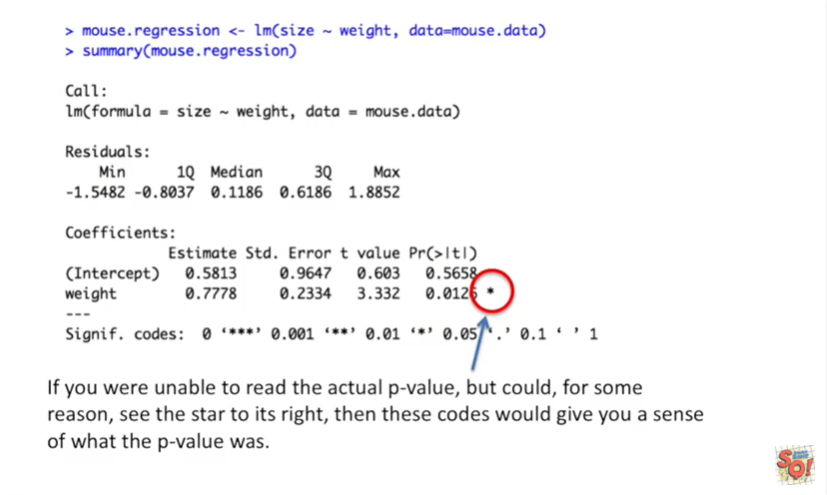
Lastly these are the P values for the estimated parameters.

Generally speaking we are usually not interested in the intercept so it doesn't matter what it's p value is.

However we want to p value for weight to be less than 0.05 that is we want it to be statistically significant.

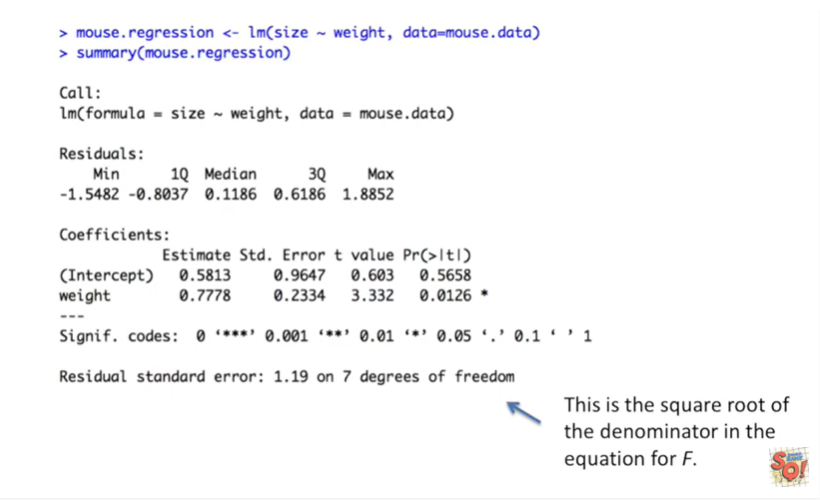


A significant p value for weight means that it will give us a reliable guess of mouse-sized.

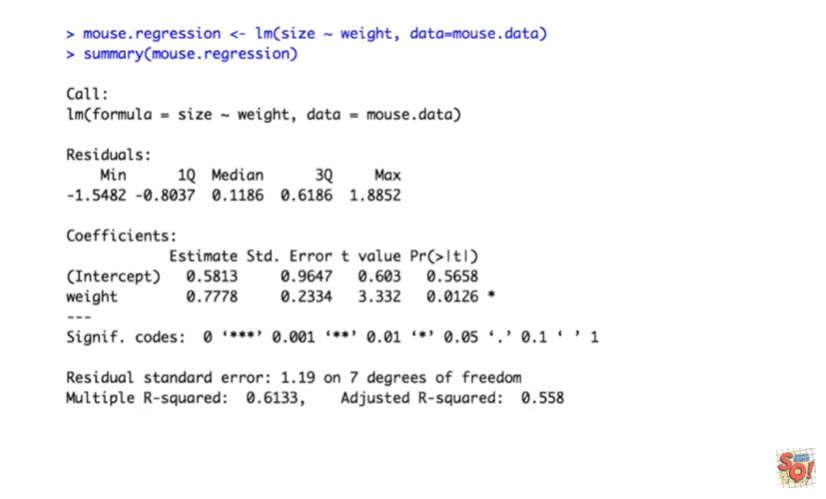


If you were unable to read the actual p value but could for some reason see the star to its right.

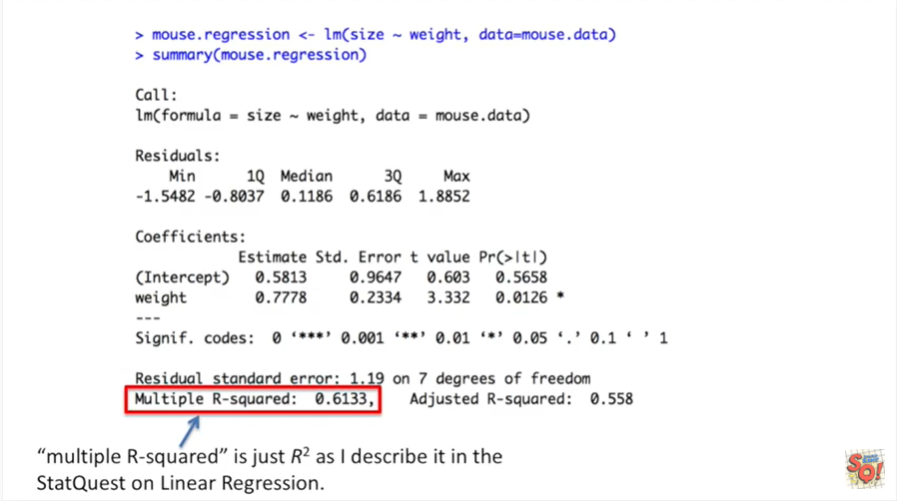
Then these codes would give you a sense of what the p value was.



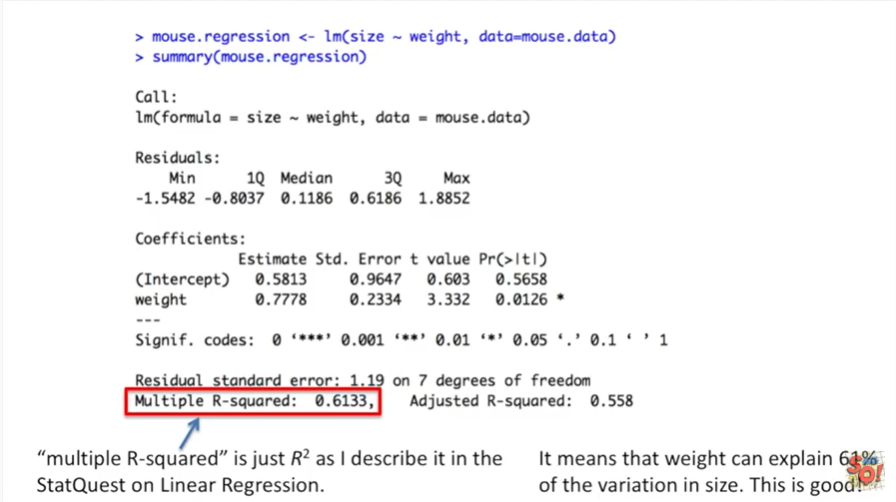
The next line the residual standard error is the square root of the denominator in the equation for f.



The next line tells us the multiple r-squared and adjusted r-squared values.

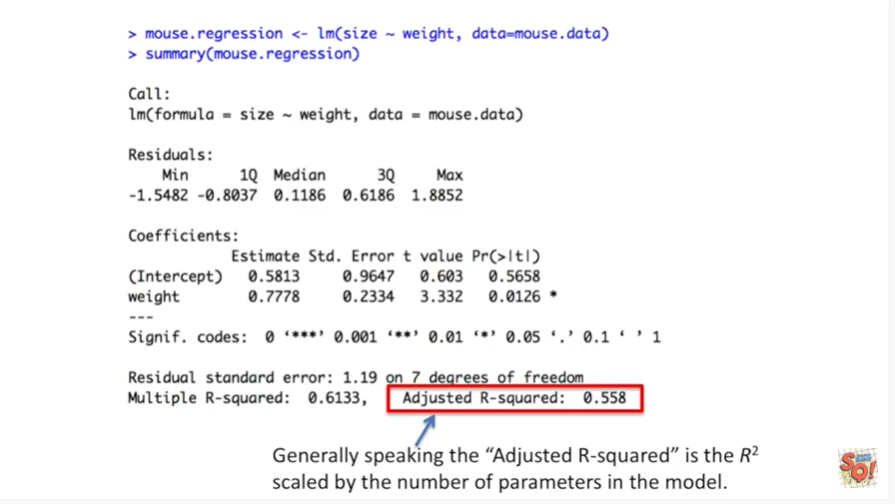


Multiple r-squared is just R squared as I described it in the stat quest on linear regression.



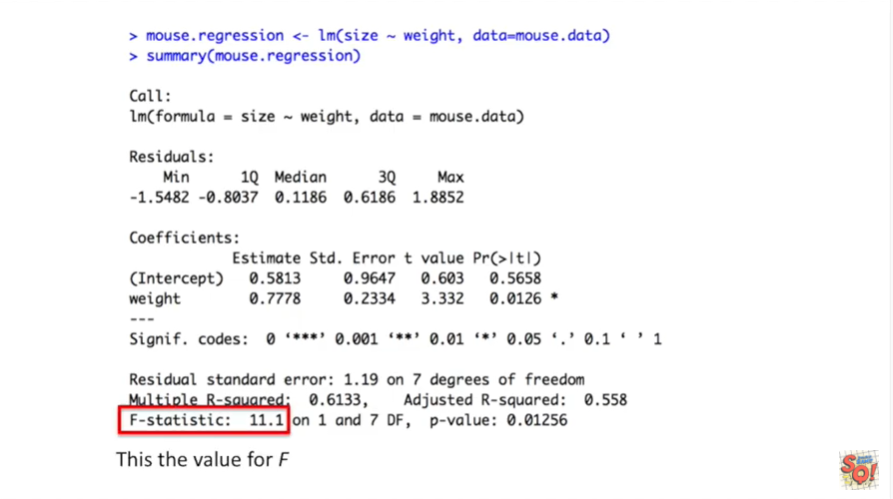
It means that weight can explain 61% of the variation in size.

This is good.

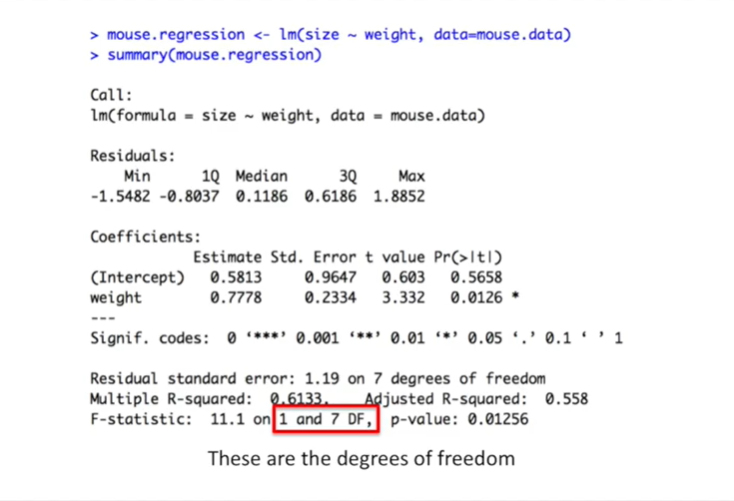


Generally speaking the adjusted R squared is the r-squared scaled by the number of parameters in the model.

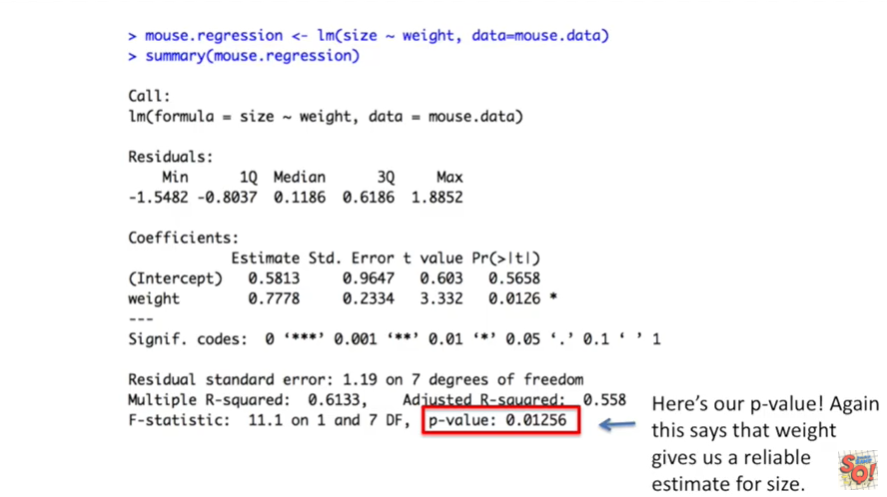
The next line tells us if the r-squared is significant or not.



This is the value for F.

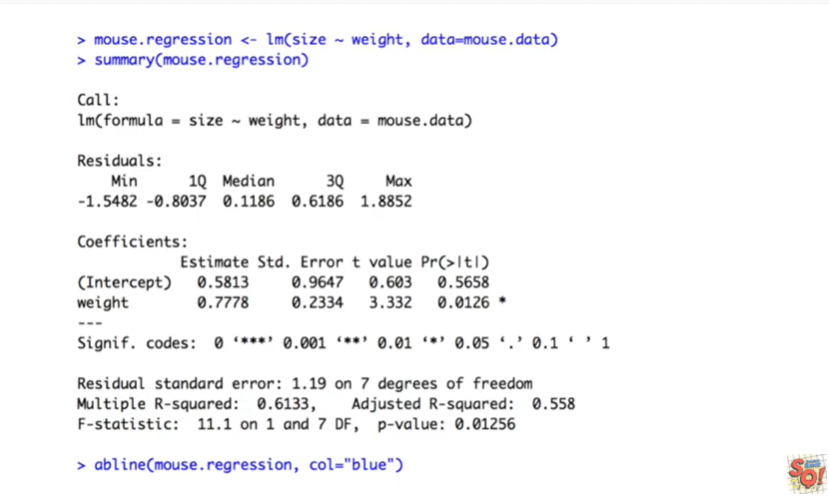
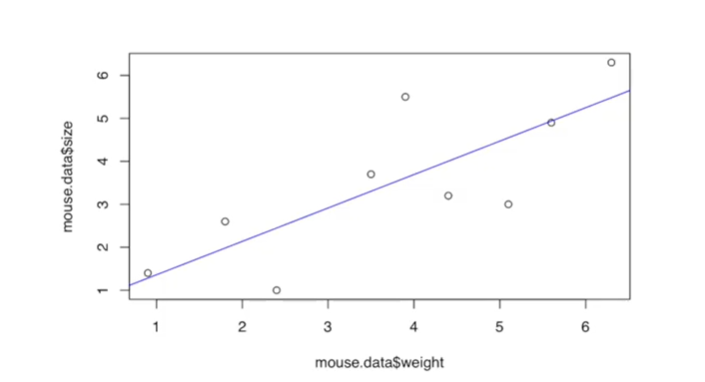


These are the degrees of freedom.



And here's our p-value.

Again this says that weight gives us a reliable estimate for size.

Lastly we can add the regression line to the XY graph we started your on earlier.