

Rsquared
by Statquest

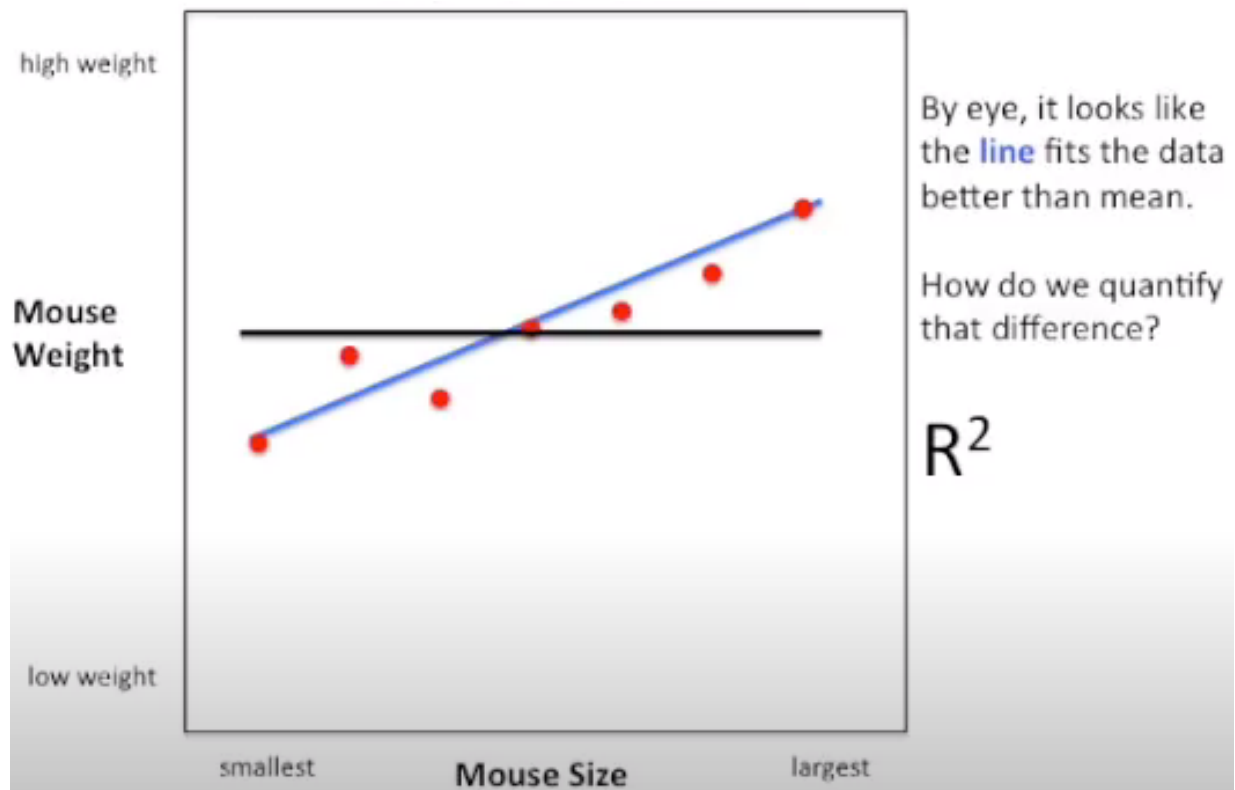
why?

$R = 0.7$ is twice as good as $R = 0.5$

Rsquared = 0.7 is what it looks like 1.4 times as good as Rsquared = 0.5

variation of the data = $\text{Sum}(\text{weight for mouse } i - \text{mean})^2$

points are squared so that points below the mean don't cancel points above the mean



**key thing here is reorganizing the data

instead of using Mouse ID we switched x axis to Mouse Size

this doesn't change any of the data so the mean stays the same

but now we are better able to fit the line of best fit

$R^2 = (\text{Var}(\text{mean}) - \text{Var}(\text{line})) / \text{Var}(\text{mean})$

$\text{Var}(\text{mean}) = \text{sum of differences between the points and the mean}$

$\text{Var}(\text{line}) = \text{sum of differences between the points and the line}$

R^2 range is from 0 to 1 because the variation around the line will never be greater than the variation around the mean

and it will never be less than 0

division within the equation makes R^2 a percentage

example

$$\text{Var}(\text{mean}) = 32$$

$$\text{Var}(\text{line}) = 6$$

$$R^2 = \frac{\text{Var}(\text{mean}) - \text{Var}(\text{line})}{\text{Var}(\text{mean})}$$

$$R^2 = \frac{32 - 6}{32}$$

$$R^2 = \frac{26}{32} = 0.81 = 81\%$$

There is 81% less variation around the **line** than the mean.

...Or...

The size/weight relationship accounts for 81% of the variation.

Means that 81% of the variation in the data is explained by the size/weight relationship

Now, when someone says...

"The statistically significant R^2 was 0.9..."

You can think to yourself..

"Very good! The relationship between the two variables explains 90% of the variation in the data!"

And when someone else says...

"The statistically significant R^2 was 0.01..."

You can think to yourself...

"Dag! Who cares if that relationship is significant, it only accounts for 1% of the variation in the data.

Something else must explain the remaining 99%."

Conceptualizing

Statistically significant R is 0.9

that means R^2 is 0.9×0.9 which is .81 or 81%

that means the relationship between those two variables accounts for 81% of the variation in the data

or

R is 0.5 > R^2 is only 25% accountability of the variation

*remember R^2 does not indicate direction

