This is a Khan Academy document. I selected it carefully from the gazillions of web hits on r-squared. It is a high-quality writeup for developing intuition.

R-squared intuition

AP.STATS: DAT-1 (EU), DAT-1.G (LO), DAT-1.G.4 (EK)

When we first learned about the correlation coefficient, r, we focused on what it meant rather than how to calculate it, since the computations are lengthy and computers usually take care of them for us.

We'll do the same with r^2 and concentrate on how to interpret what it means.

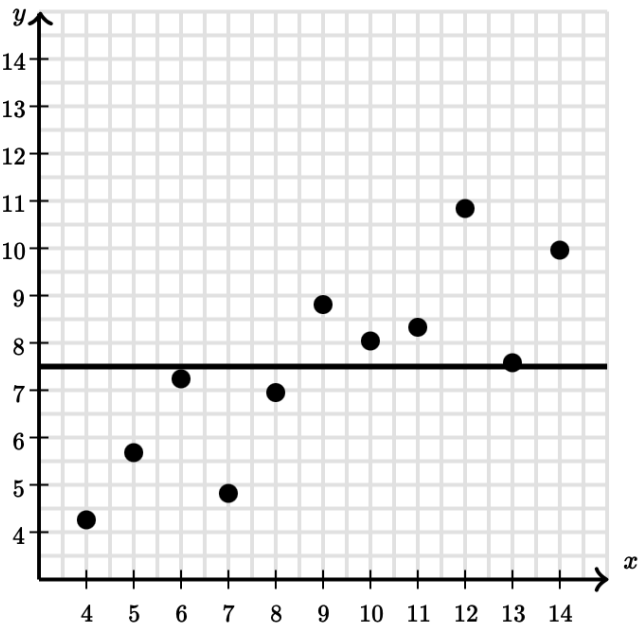
In a way, r^2 measures how much prediction error is eliminated when we use least-squares regression.

Predicting without regression

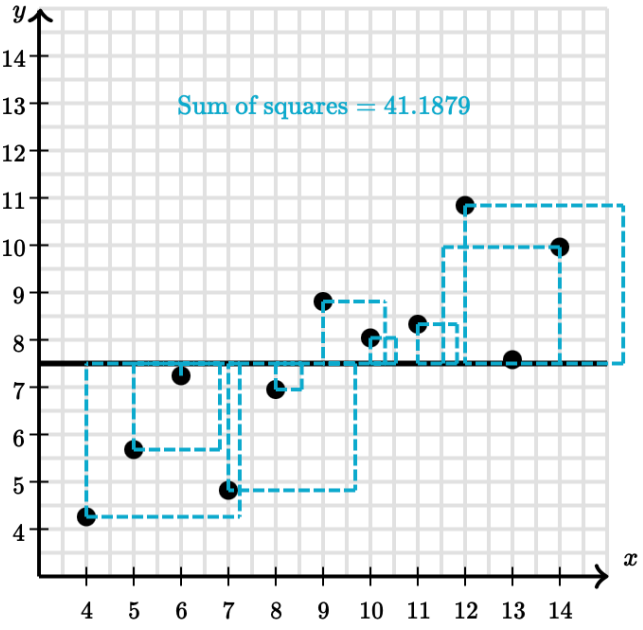
We use linear regression to predict y given some value of x. But suppose that we had to predict a y value without a corresponding x value.

Without using regression on the x variable, our most reasonable estimate would be to simply predict the average of the y values.

Here's an example, where the prediction line is simply the mean of the y data:



Notice that this line doesn't seem to fit the data very well. One way to measure the fit of the line is to calculate the sum of the squared residuals—this gives us an overall sense of how much prediction error a given model has.

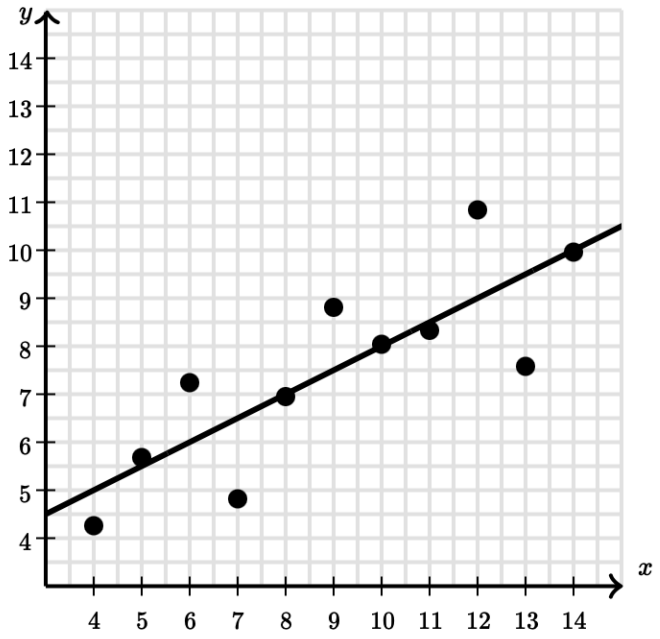


So without least-squares regression, our sum of squares is 41.1879

Would using least-squares regression reduce the amount of prediction error? If so, by how much? Let's see!

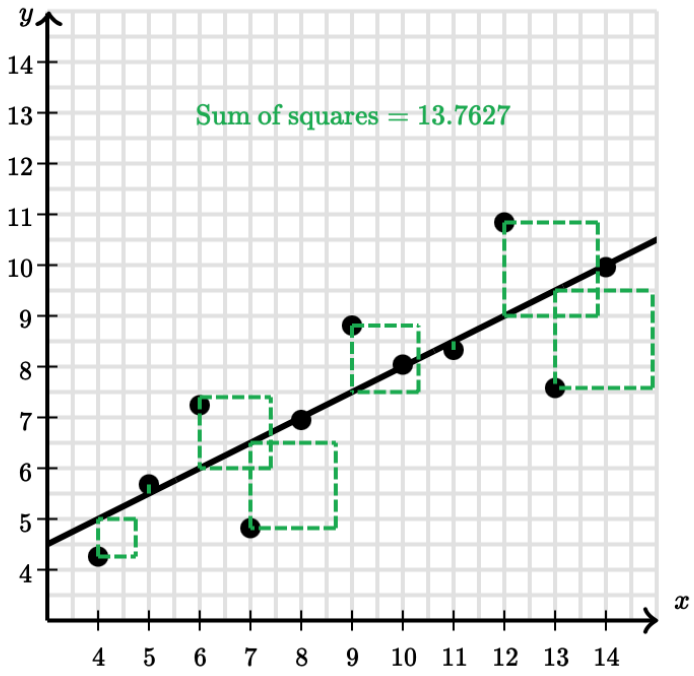
Predicting with regression

Here's the same data with the corresponding least-squares regression line and summary statistics:



Equation	r	r^2
$\hat{y} = 0.5x + 1.5$	0.816	0.6659

This line seems to fit the data pretty well, but to measure how much better it fits, we can look again at the sum of the squared residuals:



Using least-squares regression reduced the sum of the squared residuals from 41.1879 to 13.7627.

So using least-squares regression eliminated a considerable amount of prediction error. How much though?

R-squared measures how much prediction error we eliminated

Without using regression, our model had an overall sum of squares of 41.1879. Using least-squares regression reduced that down to 13.7627.

So the total reduction there is $41.1879 - 13.7627 = 27.4252$.

We can represent this reduction as a percentage of the original amount of prediction error:

$$\frac{41.1879 - 13.7627}{41.1879} = \frac{27.4252}{41.1879} \approx 66.59\%$$

If you look back up above, you'll see that $r^2 = 0.6659$.

R-squared tells us what percent of the prediction error in the y variable is eliminated when we use least-squares regression on the x variable.

As a result, r^2 is also called the **coefficient of determination.**

Many formal definitions say that r^2 tells us what percent of the variability in the y variable is accounted for by the regression on the x variable.

It seems pretty remarkable that simply squaring r gives us this measurement. Proving this relationship between r and r^2 is pretty complex, and is beyond the scope of an introductory statistics course.

Hah! Not at Deep Springs it isn't.
We will be deriving this relationship.