

# What is the difference between linear regression on y with x and x with y?

Asked 7 years, 8 months ago   Active 8 months ago   Viewed 173k times

Make your own McKinsey

McKinsey  
& Company

[View all 6 job openings!](#)

▲  
97

The Pearson correlation coefficient of x and y is the same, whether you compute  $\text{pearson}(x, y)$  or  $\text{pearson}(y, x)$ . This suggests that doing a linear regression of y given x or x given y should be the same, but I don't think that's the case.

▼  
★  
111

Can someone shed light on when the relationship is not symmetric, and how that relates to the Pearson correlation coefficient (which I always think of as summarizing the best fit line)?

regression

correlation

linear-model

pearson-r

edited Jun 22 '13 at 8:18



Nick Cox

40.9k

6

91

136

asked Feb 13 '12 at 5:15



user9097

1,158

5

13

10

- 1 Every correlation matrix will be symmetric because  $\text{cov}(x, y) = \text{cov}(y, x)$   $\text{cov}(x,y)=\text{cov}(y,x)$ . I encourage you to work out the math to see that this is indeed true. If you know the relationship between xx and yy (or whatever the variables of interest are) is not symmetric *a priori*, it might benefit you to look in to other methods of analysis. – [Phillip Cloud](#) Feb 13 '12 at 6:02
- 14 Interesting points were made on a related question, [Effect of switching response and explanatory variable in simple linear regression](#). – [chl](#) Feb 13 '12 at 10:20

## 9 Answers

▲  
159  
▼  
✓

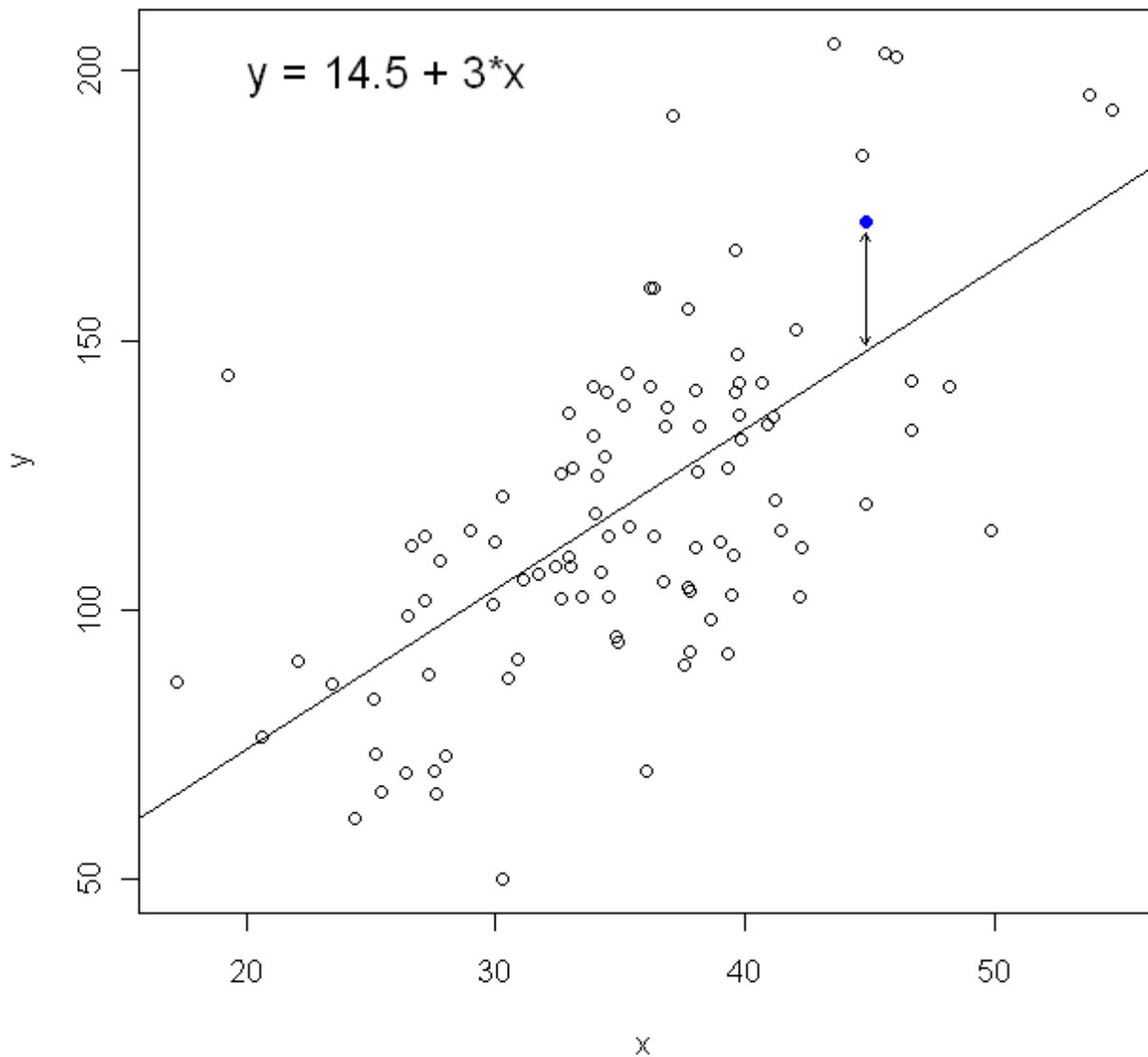
The best way to think about this is to imagine a scatterplot of points with yy on the vertical axis and xx represented by the horizontal axis. Given this framework, you see a cloud of points, which may be vaguely circular, or may be elongated into an ellipse. What you are trying to do in regression is find what might be called the 'line of best fit'. However, while this seems straightforward, we need to figure out what we mean by 'best', and that means we must define what it would be for a line to be good, or for one line to be better than another, etc. Specifically, we must stipulate a [loss function](#). A loss function gives us a way to say how 'bad' something is, and thus, when we minimize that, we make our line as 'good' as possible, or find the 'best' line.

Traditionally, when we conduct a regression analysis, we find estimates of the slope and intercept so as to minimize the **sum of squared errors**. These are defined as follows:

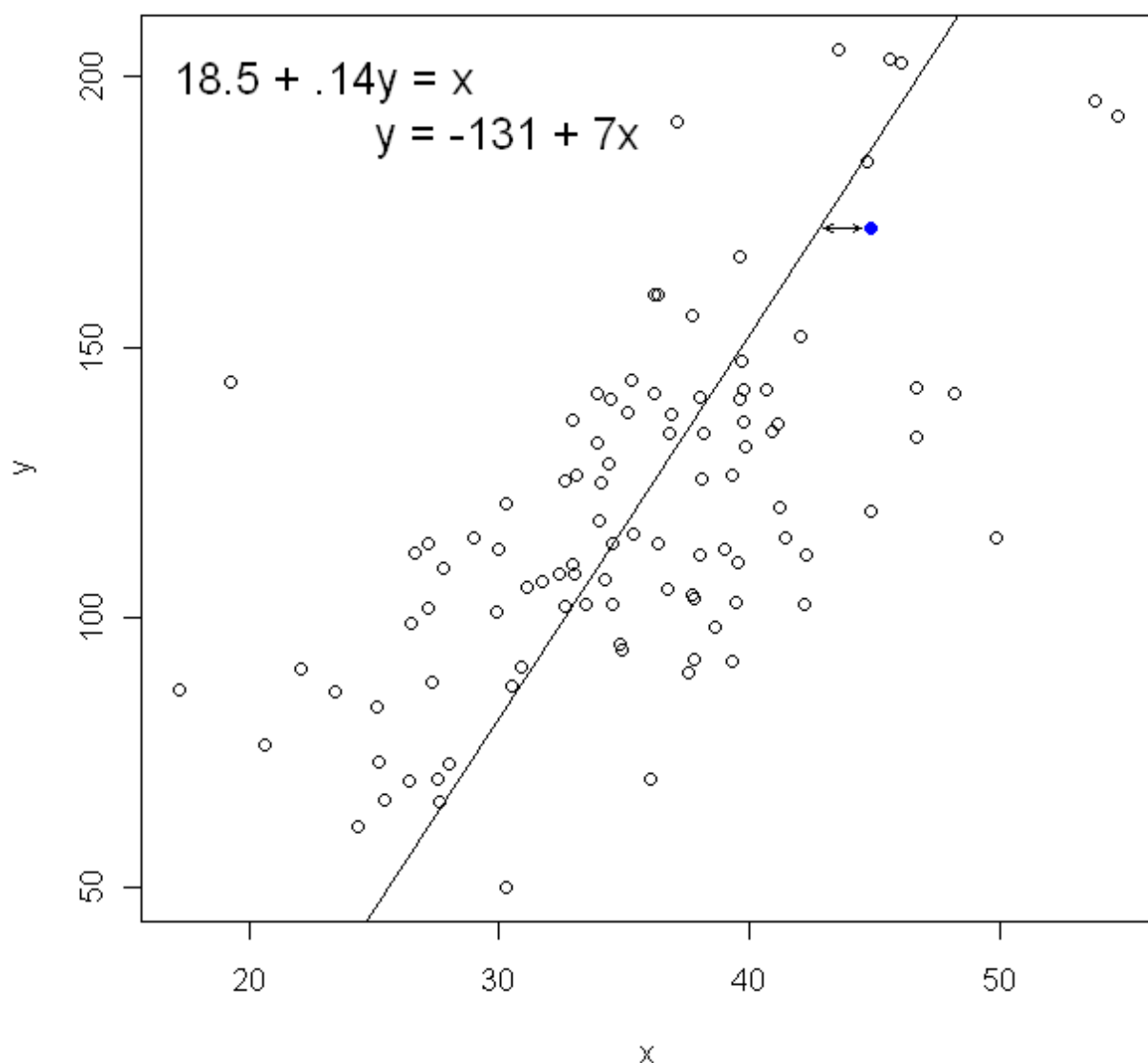
$$SSE = \sum_{i=1}^N (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$

$$SSE = \sum_{i=1}^N (y_i - (\beta^0 + \beta^1 x_i))^2$$

In terms of our scatterplot, this means we are minimizing the (sum of the squared) *vertical distances* between the observed data points and the line.

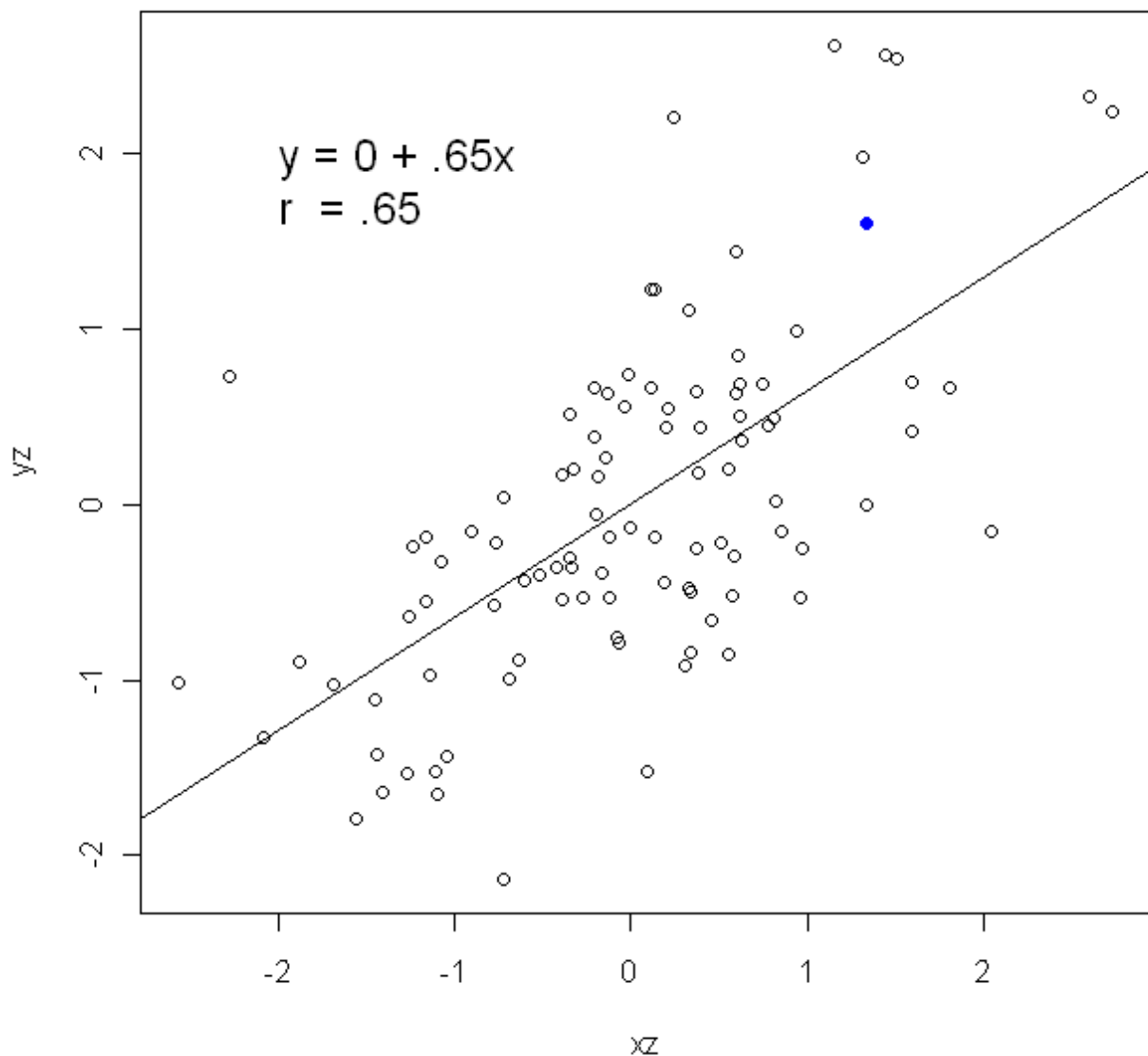


On the other hand, it is perfectly reasonable to regress  $xx$  onto  $yy$ , but in that case, we would put  $xx$  on the vertical axis, and so on. If we kept our plot as is (with  $xx$  on the horizontal axis), regressing  $xx$  onto  $yy$  (again, using a slightly adapted version of the above equation with  $xx$  and  $yy$  switched) means that we would be minimizing the sum of the *horizontal distances* between the observed data points and the line. This sounds very similar, but is not quite the same thing. (The way to recognize this is to do it both ways, and then algebraically convert one set of parameter estimates into the terms of the other. Comparing the first model with the rearranged version of the second model, it becomes easy to see that they are not the same.)



Note that neither way would produce the same line we would intuitively draw if someone handed us a piece of graph paper with points plotted on it. In that case, we would draw a line straight through the center, but minimizing the vertical distance yields a line that is slightly *flatter* (i.e., with a shallower slope), whereas minimizing the horizontal distance yields a line that is slightly *steeper*.

A correlation is symmetrical;  $xx$  is as correlated with  $yy$  as  $yy$  is with  $xx$ . The Pearson product-moment correlation can be understood within a regression context, however. The correlation coefficient,  $r$ , is the slope of the regression line when both variables have been *standardized* first. That is, you first subtracted off the mean from each observation, and then divided the differences by the standard deviation. The cloud of data points will now be centered on the origin, and the slope would be the same whether you regressed  $yy$  onto  $xx$ , or  $xx$  onto  $yy$  (but note the comment by @DilipSarwate below).



Now, why does this matter? Using our traditional loss function, we are saying that all of the error is in only *one* of the variables (viz.,  $yy$ ). That is, we are saying that  $xx$  is measured without error and constitutes the set of values we care about, but that  $yy$  has *sampling error*. This is very different from saying the converse. This was important in an interesting historical episode: In the late 70's and early 80's in the US, the case was made that there was discrimination against women in the workplace, and this was backed up with regression analyses showing that women with equal backgrounds (e.g., qualifications, experience, etc.) were paid, on average, less than men. Critics (not just people who were extra thorough) reasoned that if this was true, women who were paid equally with men would have to be more highly qualified, but when this was checked, it was found that although the results were 'significant' when assessed the one way, they were not 'significant' when checked the other way, which threw everyone involved into a tizzy. See [here](#) for a famous paper that tried to clear the issue up.

(Updated much later) Here's another way to think about this that approaches the topic through the formulas instead of visually:

The formula for the slope of a simple regression line is a consequence of the loss function that has been adopted. If you are using the standard Ordinary Least Squares loss function (noted above), you can derive the formula for the slope that you see in every intro textbook. This formula can be presented in various forms; one of which I call the 'intuitive' formula for the slope. Consider this form for both the situation where you are regressing  $y$  on  $x$ , and where you are regressing  $x$  on  $y$ :

$$\begin{array}{c} \text{y on x} \\ \text{|||||} \end{array} \quad \hat{\beta}_1 = \frac{\text{Cov}(x, y)}{\text{Var}(x)}$$

$\beta^1 = \text{Cov}(x, y) \text{Var}(x)^{-1}$  y on x

$$\begin{array}{c} \text{x on y} \\ \text{|||||} \end{array} \quad \hat{\beta}_1 = \frac{\text{Cov}(y, x)}{\text{Var}(y)}$$

$\beta^1 = \text{Cov}(y, x) \text{Var}(y)^{-1}$  x on y

Now, I hope it's obvious that these would not be the same unless  $\text{Var}(x) = \text{Var}(y)$ . If the variances are equal (e.g., because you standardized the variables first), then so are the standard deviations, and thus the variances would both also equal  $\text{SD}(x)\text{SD}(y)$ . In this case,  $\hat{\beta}_1$  would equal Pearson's  $r$ , which is the same either way by virtue of the principle of commutativity:

$$\begin{array}{c} \text{correlating x with y} \\ \text{|||||} \end{array} \quad r = \frac{\text{Cov}(x, y)}{\text{SD}(x)\text{SD}(y)}$$

$r = \text{Cov}(x, y) \text{SD}(x)\text{SD}(y)^{-1}$  correlating x with y

$$\begin{array}{c} \text{correlating y with x} \\ \text{|||||} \end{array} \quad r = \frac{\text{Cov}(y, x)}{\text{SD}(y)\text{SD}(x)}$$

$r = \text{Cov}(y, x) \text{SD}(y)\text{SD}(x)^{-1}$  correlating y with x

edited Aug 31 '18 at 18:41