This tutorial is going to teach you how to find the least-squares line of a data set. So let's start with an example that deals with the airfare prices for certain destinations for Minneapolis/St. Paul Airport. Boston is 1,266 miles from St. Paul. And it has an airfare of $263. Et cetera for the rest of these.

The scatter plot looks like this. And the mean of the miles in this data set is 882.4. And the average airfare was $234 per ticket. You can also find the standard deviation for the miles and for the airfare. We'll need all of this information as we move forward.

There are two key terms that we're going to need to know as we go through and find the equation of the least-squares line. Two key facts-- first, the point x bar, y bar is a point on the line. Now, x's and y's don't do much for me. What do those mean in this case?

x bar is the mean of the explanatory variable. What's the explanatory variable here? It's the number of miles. And y bar is the mean of the response variable. That means it's the mean of the distance traveled.

So where is x bar, y bar? It's going to be 882.4, 234 on the graph. Where is that located on the graph? Right there. So that's a point that this line is going to pass through for sure, for sure, for sure.

Second key fact-- the slope of the line is equal to the correlation times the standard deviation of the y values, the response, over the standard deviation of the explanatory variable. So let's try and find all of those values. The standard deviation of the response variable is here, 68. Because airfare is the response variable.

The standard deviation of the explanatory variable is right here, 393. The only thing we don't have is the correlation coefficient. The correlation coefficient is easy enough to find. It's 0.794. So we're going to need these three values, the ones in red, to determine the slope of the line. And all we have to do is plug them in.

The slope is going to be 0.794 times 68 over 393. The result of that is 0.137. So what is that 0.137? That's the change in y, the change in airfare, which is measured in dollars, over a change in one of the miles. This is in dollars per mile. It's about 13.7 cents per mile.

And this is what it looks like when graphed on the graph here. And it does appear to go right through the pack of points like it's supposed to. Now, this is the slope. And we knew a point on the line to begin with. That's in fact all the information we need to algebraically determine the line, the equation of the line.

So the equation of the line was airfare hat equals b0 plus b1 times miles. Now, we just found on the previous page that b1, the slope, is $0.137 per mile. So we're just going to replace that.

Now we need to find b0. That's the only other constant here. And we don't know it. But we do know a value for miles and airfare hat currently. We know x bar, y bar. Average number of miles, average value of airfare is going to be on the line.

So we know the airfare is predicted to be 234 when the miles is 882.4. When we substitute those numbers in temporarily for miles and airfare, we can solve the rest and get 113.11 equals b0. Let's put that all together.

Now we know the slope. And we know the y-intercept. Airfare hat equals 113.11 plus 0.137 times the number of miles traveled. This might seem a little frustrating or confusing. But the least-squares line is typically found using technology-- a calculator or spreadsheet or some kind of internet applet.

So if you get frustrated, reach for your calculator. And your calculator can do it. Or we can use technology. We can use something like a spreadsheet. This is the correlation over here. On Excel, the easiest way to do it is to highlight your data and create a chart that is a scatter plot. So we're going to find a scatter plot. Right here.

And when you do this, you have to actually right-click or control-click onto the data points themselves. Do you see how they got highlighted? And say, Add Trendline. And we want the line, the linear trendline. And under Options, we can say, Display Equation.

And there it is. Same as we had. Due to rounding, that's where we got the 113, and this one says 112.9. But ultimately, it's essentially the same idea. So don't get too frustrated. Because technology can rescue you here. Especially for larger data sets, finding this by hand can be a pain. But we're going to mostly use technology to do it.

And so to recap, calculation of the least-squares line involves two key facts-- the point x bar, y bar-- mean of explanatory variable, mean of response-- is a point on the line And second, that the slope is a calculable value from the correlation and the standard deviations that you have.

So we talked about the least-squares line and calculating the least-squares line. And we used all of these values plus correlation in order to find it. Good luck. And we'll see you next time.