

# Linear Regression

One process used in experiments is making a measurement and using a physics equation to make another measurement. For example, I could measure height and fall time and calculate the acceleration due to gravity by using

$$h = \frac{1}{2}at^2 \tag{1}$$

which comes from  $y = y_0 + v_0t + \frac{1}{2}at^2$  with  $v_0 = 0$  for being dropped and letting the change from  $y_0$  to  $y$  be the height  $h$ .

With one measurement of  $h$  and  $t$ , I could calculate  $a$ , but it would be only be as accurate as my worst measurement. To get a more reliable answer, I could make multiple measurements and use a method called *linear regression*.

Imagine I were to take all my data and put it into a spreadsheet, like Excel<sup>®</sup>. I could make a graph out of the data. However, it turns out that I can do a lot more interesting stuff with a linear relationship, so instead I can mess with my data a little bit to make it a linear equation.

Recall the equation for a line:  $y = mx + b$ . We can match this to the equation above by comparing them:

$$y = mx + b \tag{2}$$

$$h = \frac{1}{2}at^2. \tag{3}$$

Comparing the equations, if my two variables being measured are  $h$  and  $t$ , it looks like I have  $y$  matching up with  $h$  and  $x$  matching up with  $t^2$ , which leaves  $\frac{1}{2}a$  corresponding to the slope  $m$  and no  $y$ -intercept (so  $b = 0$ ).

So if I were to plot a set of numbers whose  $x$ -values are  $t^2$  and  $y$ -values are  $h$ , the points should (hopefully) form a straight line, whose slope must be (hopefully)  $\frac{1}{2}a$ . If I can figure out that line's slope value, I can calculate my acceleration.

This is easy to do in Excel<sup>®</sup> or similar spreadsheet software (OpenOffice or Google Docs).

### IN EXCEL<sup>®</sup>, YOU CAN DO THE FOLLOWING:

- 1 Put the  $h$  and  $t^2$  values into two columns. (You can just put the  $t$  values in and have Excel<sup>®</sup> calculate the squares if you want to save time.)
- 2 Highlight the two columns of data and click **Charts** on the ribbon or **Insert Chart** from the **Insert** menu. Choose **Scatter (Marked Scatter)** if you need to specify—nothing that connects the dots).
- 3 Right-click (or Cmd+click for Mac) on a point and select **Add Trendline**. In the **Trendline** options, choose **Linear** and click **Show Equation on Graph**.
- 4 You should now get a straight line that roughly corresponds with your data and its equation.

In this case, since we're plotting  $h$  versus  $t^2$ , the slope must be equal to  $\frac{1}{2}a$ , and we can calculate  $a$ .

The same calculation can be done using a command called **LINEST** in Excel<sup>®</sup> (it does an estimation of the line). Using **LINEST** is a little strange, but it gives the same output as the graph method. Use whatever method you like.

### TO USE LINEST:

- 1 Highlight four empty cells in a  $2 \times 2$  grid.
- 2 In the upper left cell, type “=linest(”.
- 3 Highlight the  $y$ -values ( $h$  column), and type a comma.
- 4 Highlight the  $x$ -values ( $t^2$  column), and type a comma.
- 5 Type “1, 1”.
- 6 Hold Ctrl+Shift and press Enter on Windows or hold Cmd+Shift and press Enter on a Mac. (Just hitting Enter may mess it up.)
- 7 You should see four numbers. The upper left is the slope and the upper right is the  $y$ -intercept (which should ideally be zero but likely won't be for real data). The two numbers below them are the uncertainty values of the slope and intercept values, which are statistically calculated.

The slope in this method should be exactly the same as the graphical method. Use whichever you like.