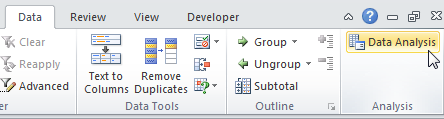
**Steps to running linear regression in EXCEL**

|  |  |
| --- | --- |
| # hrs studied (x) | exam score (y) |
| 56 | 91 |
| 41 | 82 |
| 27 | 76 |
| 44 | 84 |
| 25 | 68 |
| 35 | 77 |
| 29 | 83 |
| 31 | 72 |
| 50 | 97 |
| 22 | 66 |
| 40 | 85 |
| 48 | 93 |
| 38 | 80 |
| 24 | 79 |
| 49 | 96 |
| 27 | 81 |
| 26 | 72 |
| 45 | 88 |
| 47 | 93 |
| 29 | 78 |

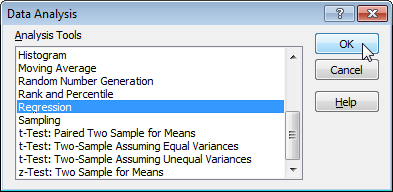


1.To run a regression, you first need to stack your data as shown to the right.

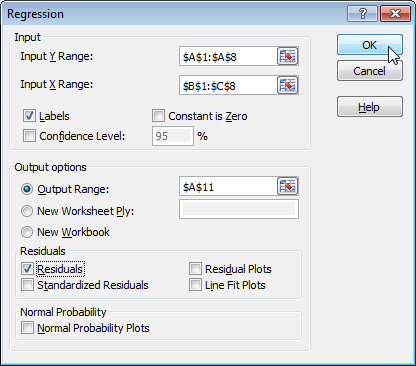
Independent variable goes on the left; response/outcome variable on the right.

In effect, your setting your data up as stacks of ordered pairs (X,Y).

2. On the data tab select ***Data Analysis…*** and choose “Regression.” OK.



3. Excel asks you for two ranges of cells, one containing the “Y”



values(i.e., your response variable),

and one containing the “X” values (i.e.,your independent variable).

Click each icon, then highlight the appropriate

columns of cells, *including their headings.*

4. Check the **Labels** box, so Excel knows you included headings atop

each column.

You can also check the **Line Fit Plots** to generate a graph of your

data and a best fit line.

5. Generated findings. You can quickly resize the columns by double clicking up top between the A & B, between B & C, etc. You can change the number of decimal places shown by clicking on the decrease decimal icon in the number area. There’s lots of info, but four pieces of major interest to you:

* The **slope coefficient** (identified by the response variable; in this case “Depth”) and the **intercept coefficient**. These respectively correspond the slope (b) and the yintercept (a) of your best fit line, and you can plug them into y = a +b x to get the equation of that line.
* The **p value** for the *slope* (*not the pvalue for the y intercept, which you usually don’t care about*). If p < .05, then you can reject the null hypothesis that the independent variable has no effect on the response variable. After all, a positive or negative slope is what you were after, and *the steeper it is, the stronger*

*the relationship*.

* The **RSquare value**. This is a number ranging from 0 to 1, and is a measure of *how tightly* your data points fit the best fit line. An Rsquare of 1.0 is a perfect fit, with every point falling right on the line, and zero means there’s absolutely no pattern or fit whatsoever.

6. If you also clicked Line Fit Plots, Excel also gave you a graph of the data and the best fit line, but it’s probably all scrunched together. Grab a corner and drag to make it bigger. To widen your plot even more, go ahead and delete the legend (click it, then hit delete). Finally, double click one of the best fit points, then give it a solid line under the Patterns tab.