

Regression with Microsoft Excel

Excel is one component of the Microsoft Office Suite of useful software programs. This appendix is oriented toward entering data into an Excel spreadsheet, constructing a statistical scatter plot, and performing regression analysis on the data.

Data Entry:

Cells can hold text, numbers, a combination of the two, or formulas.

Each cell is named with a column letter then row number (e.g., the A1 cell describes the cell in the A column and row 1).

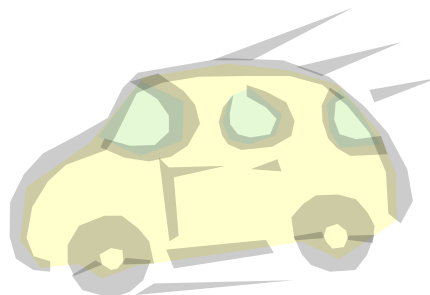
There are several ways to navigate around the spreadsheet. You can use arrow keys (right →, left ←, down ↓, up ↑); you can use your mouse and a single click to select a certain cell; and you can use the tab key, shift + tab, return, or shift + return to move right, left, down and up.

There are ways to change font, font size, font color, make letters bold, italic, or underlined, center- or left- or right-justify, round numbers up or down, etc. Excel is similar to any word processing software in many of these features.

You can also change the width of a column by clicking on the border between columns and dragging right or left.


The chart below shows data from a student's records of cumulative miles traveled by car over a 10-day period. This data is entered into columns A and B of an Excel spreadsheet. We are showing row numbers and column letters for your convenience.

	A	B
1	Day #	Cumulative Miles
2	1	12
3	2	27
4	3	49
5	4	72
6	5	86
7	6	103
8	7	129
9	8	158
10	9	184
11	10	212



Constructing statistical graphs (scatter plot):

First select (with a click and drag) the cells that contain data you want included in the graph.

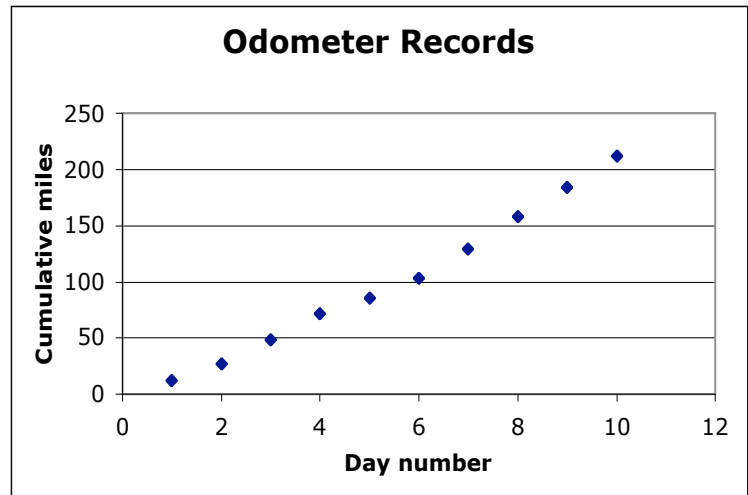
Next, click on the Chart Wizard icon. 

Find and select the XY (Scatter) plot, then choose the only true scatter plot (top left sub-type).

Regression with Microsoft Excel

Then follow the directions through several stages. Click on **Next** to advance to the next choice (you're fine-tuning the graph), and click on **Finish** to take your impressive graph back to the worksheet. We expect the axes to be clearly labeled and the graph to have a reasonable title. You can move your graph by clicking on the graph and dragging. You can resize your graph by clicking on a border button on the graph and dragging. You can also change fonts, colors, borders, backgrounds, and other features of the graph.

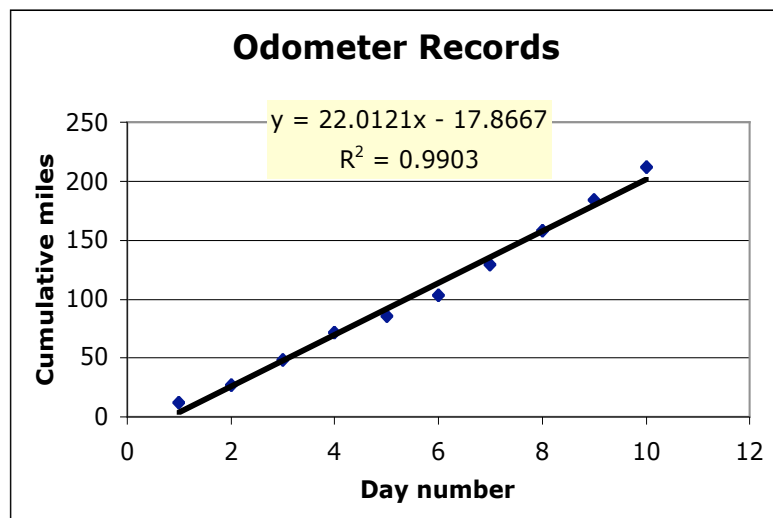
Your scatter plot may look like this:



The critical steps for regression analysis start with selecting the graph. After you single-click on the graph to select it, click on "Chart" on the menu bar and drag down to "Add Trendline". You'll notice two tabs: "Type" and "Options". Under "Type", choose "Linear", and under "Options", choose "Display equation on chart" and "Display R^2 value on chart" by clicking in the radio boxes . The software does the rest!

It is typical to have the software round the constants in the formula (in this case, the slope and y-intercept for the linear function) to a consistent number of decimal places. We double click on the formula and use the "Number" tabs to format the formula to have any number of decimal places we choose. In our example, we are showing 4 decimal place precision.

The results of this linear regression analysis should look like this:



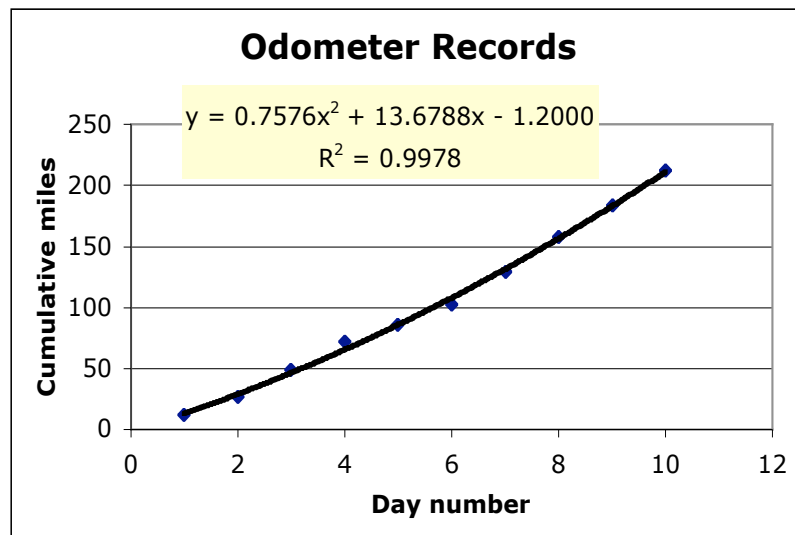
Regression with Microsoft Excel

One of the main points of regression analysis is to use the equation to make predictions. With the linear regression equation, in order to predict the number of miles the student may travel in a 31-day month or a 365-day year, you would type the command `=22.0121*A12 - 17.8667` or `=22.0121*A13 - 17.8667`, assuming that the A12 and A13 cells hold the numbers 31 and 365, as shown below. The results of evaluating the formula with $x = 31$ and $x = 365$ are 665 miles and 8,017 miles, respectively.

	A	B
1	Day #	Cumulative Miles
2	1	12
3	2	27
4	3	49
5	4	72
6	5	86
7	6	103
8	7	129
9	8	158
10	9	184
11	10	212
12	31	665
13	365	8017

The correlation coefficient, given by the symbol R , is found simply by taking the square root of the given R^2 value. In any cell of the spreadsheet, we may type `=SQRT(0.9903)`. Since $\sqrt{R^2} = R$ or $-R$, we must consider the sign of our correlation coefficient; a positive slope indicates a positive correlation, and a negative slope indicates a negative correlation. A more precise method that works only for linear correlation is the `=CORREL(A2:A11,B2:B11)` command. Using either method, Excel finds R to be approximately 0.9951.

In a very similar way, Excel handles quadratic regression (using polynomial, order 2), cubic regression (using polynomial, order 3), and other polynomial regression functions (orders 4, 5, and 6), logarithmic regression, exponential regression, and power regression. The quadratic regression equation of best fit is shown below.



Regression with Microsoft Excel

For quadratic regression, we find the correlation coefficient to be 0.9989 $\boxed{=SQRT(0.9978)}$. And the predictions for 31 days and 365 days are 1,150 miles and 105,922 miles, respectively. Incidentally, we are using the formulas $\boxed{=0.7576*A12^2+13.6788*A12-1.2000}$ and $\boxed{=0.7576*A13^2+13.6788*A13-1.2000}$, with 31 and 365 in the A12 and A13 cells, as before.

While the R value for quadratic regression is close to 1 (slightly closer than the linear correlation coefficient), indicating a very close fit to the scatter plot, it doesn't seem to predict well for large day numbers. The number of decimal places in the constants also can make a dramatic difference in the reasonableness of the predictions.

A few notes about printing in Excel —

In order to print, it's best to click and drag over all cells you'd like to print. Then under "File", select "Set Print Area". This establishes the page contents.

You can use "Page Setup" under "File" to orient the page vertically ("Portrait") or horizontally ("Landscape"). You can also go under "File" to "Print Preview" to see how the page(s) will look before you commit to print. This is highly recommended!

Then, you're ready to click "Print" (under "File"), or click on the printer icon button. 

P.S. There are shortcuts and alternative methods for just about anything in Excel as in other software programs. Explore and enjoy the power of this spreadsheet software!