Economics 91 Spring 2021
Linus Yamane Pitzer College

Excel

Microsoft Excel is a spreadsheet which features calculations and graphing tools. It can be used to store and organize many data sets. Using its features and formulas, you can also use the tool to make sense of your data. For example, you could use a spreadsheet to track data and automatically see sums averages and totals. Since EXCEL is available on most every computer in the world, it is worth putting EXCEL in your toolbox.

Calculations

You should think of each cell on the spreadsheet as a calculator. You can put numbers in a cell, a formula, or text. All formulas in Excel must begin with an equals sign (=). This is because the cell contains, or is equal to, the formula and the value it calculates.

Excel uses standard operators for formulas, such as a plus sign for addition (+), a minus sign for subtraction (-), an asterisk for multiplication (*), a forward slash for division (/), and a caret (^) for exponents.

While you can create simple formulas in Excel manually (for example, =2+2 or =5*5), most of the time you will use cell addresses to create a formula. This is known as making a cell reference. Using cell references will ensure that your formulas are always accurate because you can change the value of referenced cells without having to rewrite the formula.

For example, A1 refers to the cell at the intersection of column A and row 1; B2 refers to the second cell in column B, and so on. The range A1:C2 includes 6 cells from A1 through C2.

To add up the values in cells A1 and A2, you use this formula: =A1+A2

There are a number of mathematical functions built into Excel.

=sum(number1, number2, …)	adds up all the numbers
=sqrt(number)	returns the square root of the number
=ln(number)	returns the natural log of the number
=exp(number)	returns e raised to a given power
=pi()	returns the constant value of 3.141592654
=int(number)	rounds a number down to the next integer
=rand()	generates a random number between 0 and 1

There are a number of built in statistical functions.

= average(number1, number2,…)	mean
= stdev.p(number1, number2,…)	standard dev of population
= stdev.s(number1, number2,…)	standard deviation of sample
= var.p(number1, number2,…)	variance of population
= var.s(number1, number2,…)	variance of sample
= covariance.p(array1, array2)	covariance of population

= covariance.s(array1, array2)	covariance of sample
= correl(array1, array2)	correlation coefficient
= combin(number, number_chosen)	combinations
= permut(number, number_chosen)	permutations

Standard probability distributions are also available.

= norm.dist(x, mean, std dev, cumulative)	normal distribution
= chisq.dist(x, degree freedom, cumulative)	chi square distribution
= t.dist(x, degree freedom, cumulative)	Student's t distribution
= f.dist(x, degree freedom 1, degree freedom 2, cumulative)	Snedecor's F distribution

And logical expressions can be used as well.
=IF(logical test, value if true, value if false)
For example,
=IF(250+B1>=A1,"OK","Low")

Use the Fill Handle

There may be times when you need to copy the content (either numbers or formulas) of one cell to several other cells in your worksheet. You can use the fill handle to quickly copy and paste content to adjacent cells in the same row or column.

Select the cell(s) containing the content you want to use. The fill handle will appear as a small square in the bottom-right corner of the selected cell(s). Click, hold, and drag the fill handle until all of the cells you want to fill are selected. Release the mouse to fill the selected cells. Excel is intelligent about recognizing patterns to repeat or adjusting the relative cell references.

Graphing

You can make a lot of graphs in Excel. The most common graphs are pie graphs, column graphs, line graphs, area graphs, and scatter graphs. Once you have your data on your spreadsheet, highlight the data, and go to Insert/Recommended Charts. You can change colors, change the thickness of dots or lines, change the scale of your axes, add horizontal and vertical titles, and add error bars or a trendline.

Analysis ToolPak

In order to do "statistics", we will be using the Analysis ToolPak in EXCEL. You may need to load the toolpak before you use it. The instructions on loading the Analysis ToolPak are on this link:

https://support.microsoft.com/en-us/office/load-the-analysis-toolpak-in-excel-6a63e598-cd6d-42e3-9317-6b40ba1a66b4

You generally need to go to File/Options/Add-Ins and click on Analysis ToolPak. After you load the Analysis Toolpak, you should go toe the Data tab in Excel. On the far right you should see a "Data Analysis" option.

If you click on Data Analysis, a menu will open up with a bunch of options including Anova: Single Factor, Correlation, Covariance, Descriptive Statistics, Histogram, Random Number Generation, and Regression among others.

Descriptive Statistics

If you highlight the Descriptive Statistics option, and click on OK, you will see a new window. Indicate the location of the data under "Input Range." If you have labels (variable names) in the first row, click the box. Click on Output Range and indicate where you would like the output to appear. Click on Summary statistics. Click OK and you will see output which looks something like this.

height	
Mean	68.55998
Standard Error	0.71749
Median	69
Mode	72
Standard Deviation	4.183652
Sample Variance	17.50295
Kurtosis	0.443898
Skewness	0.144221
Range	20
Minimum	60
Maximum	80
Sum	2331.039
Count	34

Regression

If you want to run a regression in Excel, go to Data/Data Analysis. Choose Regression, and click OK. The Regression window will appear. Input Y Range is for your dependent variable Y. Input X Range is for all of your X variables. If you have variable names in the first row of the data, click on Labels. Click on Output Range and indicate where you would like the output to appear. There are some additional options if you want to see a residual plot. Click on OK, and Excel will generate your regression estimates. The regression may look something like this:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.40156355							
R Square	0.16125328							
Adjusted R Square	0.15907755							
Standard Error	0.67687753							
Observations	774							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2	67.91290559	33.95645	74.11432	3.62816E-30			
Residual	771	353.2438252	0.458163					
Total	773	421.1567308						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	9.74396065	0.146975146	66.29666	0	9.45544173	10.03247956	9.45544173	10.03247956
edu	0.09911083	0.008218514	12.05946	8.42E-31	0.082977517	0.115244151	0.082977517	0.115244151
ex	0.00367339	0.002158822	1.701571	0.089239	-0.000564477	0.007911253	-0.00056448	0.007911253

TSS = ESS + RSS

$$\Sigma(Y_i - \bar{Y})^2 = \Sigma\left(\hat{Y}_i - \bar{Y}\right)^2 + \Sigma\left(Y_i - \hat{Y}_i\right)^2$$

TSS = 421.1567308
ESS = 67.91290559
RSS = 353.2438252

N = 774
N-1 = 773
N-k = 771

$$LNWAGE_i = \beta_1 + \beta_2 EDU_i + \beta_3 EX_i$$

$$LNWAGE = 9.74396065 + 0.09911083EDU + 0.00367339EX$$

MS = SS / df

$R^2 = {ESS}/{TSS}$ = 0.16125328 and $\bar{R}^2 = 1 - \frac{{RSS}/{(N-k)}}{{TSS}/{(N-1)}} = 0.15907755$

$F = \frac{{ESS}/{(k-1)}}{{RSS}/{(N-k)}}$ = 74.11432

$$t\ Stat = \frac{Coefficient}{Standard\ Error}$$

Root MSE = square root of the mean squared error = standard error of the regression = 0.67687753

$$root\ MSE = \sqrt{\frac{RSS}{N-k}} = \sqrt{\frac{\Sigma e_i^2}{N-k}}$$