

How to use Excel

- 1) Go to class web-page.
- 2) Download applicable data set. - Right-click, choose "Save Target As".
Remember the directory where you saved the file.
E.g. Download the file height.
This data set contains the gender and height of student who took econometrics in the Fall 2007. It also contains information about the height of their parent.
- 3) Open data set in Microsoft excel.
- 4) Select an empty cell. Then choose "insert" and "function". Pick the desired function. The result will be displayed in the selected cell.
Choose 'average' (calculates the mean) and select all the numbers under the category height. The result should be 70.94
- 5) Select the entire data set. Then choose "data" and "sort".
Make sure to tell excel whether your data set contains a header row.
If you select sort by height the data will be sorted with the shortest student on top and the tallest on the bottom. Each observation (row) will remain unchanged.
- 6) Choose "tools" and "data analysis"
(You may have to activate this option, by choosing "Tools", "Add-Ins", "Analysis Tool Pack")
 - a) Choose "correlation", and then select the numbers below 'father' and 'mother' as the input range. The result should be .54.
 - b) Choose "Descriptive Statistics", select the variables 'gender', 'height', 'father' and 'mother'. Select the option 'summary statistics'.
- 7) Play with excel and the various options.

Random Number Generation

(Not required but maybe instructive)

A good way to improve your understanding of statistical concepts is to generate data that follows in part certain patterns, but is also influenced by random components.

Excel allows you to generate random numbers in two different ways:

1. Choose “Data Analysis” and “Random Number Generation”
2. Select an empty cell and choose “Insert” , “Function” , “Rand”

E.g. $Y_i = a + bX_i + e_i$ where e is a random number, and subscript i refers to the respective observation, a and b are constants. Generate two random variables (each in a column), call them X and e . Generate a third variable, Y , (a new column) using:

$Y_i = a + bX_i + e_i$. Now, you can apply the various techniques we talked about It's fun* and instructive to play around a little bit with some data you generated. You can vary the constants, sample size, variance of the random numbers ... Then see what happens when you apply the various tools.

* Otherwise I am sane.