Data analysis 1
Week 3 practical

There are two exercises to do this week:

1. the *'Week 3 project'* practical in the computing cluster. It is designed to help you learn about graphing in Excel and MATLAB. In order to do this practical you should be familiar with the bullet points below.
2. the assessment *'Introduction to statistics'* on Blackboard. You should be able to do this in your own time, but may get to finish it in class.

- How to do an x-y scatter plot in Excel 2007—see `http://www.youtube.com/watch?v=6drKbhJA2EM`.
- How to do an x-y scatter plot in Matlab—see `http://www.youtube.com/watch?v=KiASzo8Yukc`.
- How to do a histogram in Excel 2007—see `http://www.youtube.com/watch?v=mdixXI5hAW0`.
- How to do a histogram in Matlab—see `http://www.youtube.com/watch?v=0T_ZzHQkc98`.
- Where to download the Excel spreadsheets from on Blackboard: Go to Course content→Spreadsheets and data.

In this practical you will be required to input your answers into Blackboard. You may wish to do the exercise in the spreadsheet or in MATLAB first though before opening the test.
First download the ELECTRIC.xls and FHEALTH.xls and MHEALTH.xls spreadsheets. Also if you haven't done so, download the notes for the course to your p-drive.

1. In the spreadsheet ELECTRIC.xls, four columns of data are available (energy consumption in kWh; cost of electricity; degree days; and average temperature in Fahrenheit) (Course content→Spreadsheets and data) use the 22 average daily temperatures and the corresponding 22 amounts of energy consumption (kWh) to construct a scatterplot in Excel or MATLAB. Based on the result, is there a relationship between the average daily temperatures and the amounts of energy consumed? Try to identify at least one reason why there is or isn't a relationship.
2. Construct a histogram of energy consumptions (kWh) with bin mid-points for every 500 kWh. Label your axes.
3. Manually constructed graphs are generally unsuitable for publications and presentations. Computer-generated graphs are much better for such purposes. From the Blackboard website for this course download the xls files FHEALTH.xls and MHEALTH.xls. These contain pulse rates (labelled PULSE) of females and males, and they are available as files that can be opened by Excel or imported into MATLAB. Using either Excel or MATLAB, analyse the FHEALTH and MHEALTH xls files, generate three histograms:
   (a) a histogram of the pulse rates of females with bin spacing of 5 BPM;
   (b) a histogram of the pulse rates of males with bin spacing of 5 BPM;

(c) a histogram of the combined list of pulse rates of males and females.

Compare the histograms. Does it appear that the pulse rates of males and females have similar characteristics? (Later in this course, we will present more formal methods for making such comparisons.)

Save your spreadsheet or MATLAB workspace (e.g. type: `save <filename>`) somewhere for later reference.