Assignment 3

Statistics exercises

Please report, for all tests you perform: which test you used, the relevant statistics (t-values, F-values, z-values, correlation coefficients, intercept and slope etc.) and the significance level. Also include all relevant plots (with a brief explanation). For completeness' sake, specify which data file you used.
If you make plots, pay attention to details (axes labels, readability etc.).

If you use R, you may also specify the R-command you used (but this is *only* supplementary information to allow me to check what you did exactly).

Go to http://ai.vub.ac.be/~bart/statsnumbers.html and type in your student number. This will be converted to four numbers between 1 and 4
If you receive the numbers: 1, 3, 2, 2, use :
Question1_1.csv, Question2_3.csv, Question3_2.csv, Question4_2.csv
The data files themselves are all found in MWO3Data.zip on the Pointcarre site.

The deadline is December 18, midnight.

Question 1

The data file contains reaction times for 100 users using user interfaces with blue elements and with red elements.

- Make histograms and boxplots of distributions of both colors.
- Make quantile-quantile plots comparing them with the normal distribution. Does this look OK?
- Test whether the reaction times for the two user interfaces are different, and report the effect size

Question 2

The data file contains numbers of bugs found in the same programming project performed by 90 different students, each one of which used either C++, C or Scheme.

- Investigate whether there are significant differences between these groups, and figure out which groups are significantly different. Make sure you test the assumptions of the test you are using.
- Assuming a normal distribution, and aiming for a power of 0.8 and a p-value of 0.05, how large would your sample need to be to know whether there is a significant difference between the number of bugs in C and Scheme?

Question 3

The data file contains information about participants using an interactive training tool to improve their programming skills. It contains their age, the number of hours they spent training and whether they passed the final test (0 indicates failure, 1 indicates success).

- Investigate whether the number of hours has predictive value for passing the test or not (in other words, does training longer improve performance?). Carefully explain your method.
- Does the age of the participants have a significant influence on the outcome of the test (controlling for the effect of the number of training hours)?

Question 4

The data file contains the number of times ICT projects with and without version management finished on time or over time.

- What are the percentages of finishing on time for the two conditions?
- Test if version management has a significant influence on whether a project runs on time or not.
- Explain why you used the test you used
- Do you find the result surprising? Can you think of a way this study could be improved?