**AC50002 PROGRAMMING LANGUAGES FOR DATA ENGINEERING**

**MATLAB ASSIGNMENT**

**Lecturer:** *Emanuele (Manuel) Trucco*

**Lab tutors:** *Roberto Annunziata, Andrew McNeil*

**Problem description**

1. Generate 200 real numbers from a Gaussian distribution, with mean 0 and standard deviation 1. Generate a histogram of the sample, choosing the number of bins with the Freedman-Diaconis rule.
2. Run a Chi-square test for Gaussianity to test the hypothesis that the numbers are a sample from a Gaussian distribution. Report the result in terms of answer (yes or no) and significance level.
3. Eliminate 10 numbers, chosen at random, from the initial set, and replace them with numbers drawn from a uniform distribution of mean 0 and standard deviation 3.
4. Repeat step 2.
5. Repeat steps 3 and 4 for 10 times, each time removing 10 numbers in random positions from the *original* Gaussian sample, and replacing them with a *new* set of uniform deviates.
6. Write a concise report (max 2 pages, Word document, Times New Roman, pointsize 11) on your observations about the results of steps 3, 4 and 5. Do results change? What conclusions can you draw?

**Marking scheme**

The maximum number of marks for the assignment is 50, divided as follows:

* 10 marks for point 1 (of which 4 for good MATLAB code, 6 for correct code behaviour),
* 15 marks for point 2 (of which 4 for good MATLAB code, 5 for correct theory solutions and choices, 6 for correct code behaviour),
* 15 marks for points 3-5 (mark division as per previous point),
* 10 marks for report covering point 6 and point 2 (2 for clear explanation, 8 for correct conclusions).

**What to submit where**

Please send the MATLAB code and the report to e.trucco@dundee.ac.uk.

The MATLAB code must be

* tested and ready to run on our laboratory machines;
* the source code only in ASCII text (no MATLAB files please).
* accompanied by concise instructions on how to run the code, and what to expect.

**Deadline**

Friday week 5 (19th Feb 2016).