MA4413 Assignment (10% of Module)

Details

- This assignment must be your own work. Evidence of copying will result in a score of zero.
- The deadline for this assignment is 25/11/2014:
  - Hard copy to be submitted during the lecture time.
  - Electronic copy and `R` script file containing all code to be submitted using the "Assignments" section on sulis.
- Each student has their own set of data. You will find this in the "MA4413-R-Assignment-Datasets" excel file (row containing your ID).

Report

- Results must be accompanied by concise description. Do not waffle.
- Your report will be a word document:
  - Text size: 12pt.
  - Main Font: arial, calibri or times new roman.
  - Font for `R` output: `courier`.

Hypothesis Tests

- When you see the phrase "carry out the test", you are required to:
  - carry out the test using `R`;
  - copy the `R` output into your report (use `courier` font);
  - clearly state the null and alternative hypotheses;
  - provide conclusion based on the p-values and confidence intervals.

Independent Research

- There are two items highlighted as being "independent research". These items have not been covered in the lectures; it is up to you to find out how to do them yourself.

Comparing Two CPU Designs **(30 marks)**

A manufacturer of CPUs wants to compare two designs in terms of their typical operating temperature (degrees Celsius). A sample of each type was selected and the temperature was recorded after 1 hour of normal use.

You have been assigned the task of analysing this data. Based on the analysis, you will decide whether or not there is a difference between the two designs and, if there is a difference, state which CPU design is superior (i.e., lower temperature is better).

The measurements for CPU 1 and CPU 2 can be found in the "MA4413-R-Assignment-Datasets" excel file. Each student has their own data. Find the row in the excel file containing your ID.

The marking scheme is as follows:
(complete the tasks in the order shown below)

Presentation **(5 marks)**

- There are 5 marks going for overall presentation of your report.

1) **Graphical and Numerical Summaries** (Lecs 1 and 2) **(8 marks)**

- Copy the data into an `R` script file and name the data vectors `cpu1` and cpu2 respectively.
- Plot histograms for each CPU type. Define the breakpoints in `hist` yourself so that there are 5 classes that span the data. The code `seq(min(x), max(x), length=6)` achieves this where `x` is a vector of data.
- Plot the two boxplots on the same graph.
- Calculate the mean, standard deviation, quartiles, IQR, minimum and maximum temperature for both CPUs. Present these in a table with two columns - one for each CPU - and round all numbers to two decimal places.
- Comment on all of the above output with reference to the shape of the distributions, centre and spread etc. but **do not waffle**. This must be concise and to the point.

2) **Check for Normality of Data** (Lec 10) **(4 marks)**

- Use Q-Q plots to determine whether or not the two datasets are approximately normally distributed (also refer to the histograms and boxplots).
- *(independent research)* There also exists a hypothesis test of normality called the Shapiro-Wilk test. Carry out this test in `R`.

3) **Difference Between Population Means** (Lecs 13,14,15,16) **(9 marks)**

- While the summaries from part (1) above are useful for describing a *sample* of data, we cannot make statements about the *whole population* without constructing confidence intervals and performing hypothesis tests.
- Carry out the F-test.
- Carry out the t-test for two independent samples (assuming equal variances if possible).
- *(independent research)* When the samples are small, the t-test requires the assumption that the data is normally distributed. If this is not the case *non-parametric methods* can be used. For comparing two independent groups, the Wilcoxon rank-sum test (also known as the Mann-Whitney U test) can be used. Carry out this test in `R`. (note: do not confuse with the Wilcoxon signed-rank test for paired data)

4) **Final Summary** **(4 marks)**

- Briefly summarise the main results of your analysis. Also, provide your final conclusion in non-statistical language. **Do not waffle**. A few key sentences is sufficient (and no more than half a page).
