L2 Statistics Assignment (Michaelmas 2016)

This assignment should take about 10 hours to complete (from data familiarisation, code development through to report writing).

CER Smart Meter Trial data

The Irish Commission for Energy Regulation (CER) undertook an electrical smart metering trial and installed smart meters with about 6000 customers. These meters measured electrical energy demand at half hourly intervals over a period of about 2 years.

As part of this assignment, data from 2 days has been extracted: Day 1, representing a summer day and Day 200, representing a winter day. The meter data for each day is given as a matrix. Within that matrix, each row represents a single customer's demand for that day (so the 48 columns represent half hour demand for that day).

You are to investigate the customer consumption differences between these two days and determine if there is any significant difference, and perform some basic analysis on outlier customers. The consumption can be computed by adding up the 48 half hour demands as recorded by the smart meter.

Data source

The Matlab matrix file can be found on duo under:

Level 2 Engineering Mathematics > Assessment > Statistics Assignment > CER_smartmeters.mat.

You should copy this file into your Matlab working directory

This can be loaded into your matlab environment using:

```
load CER_smartmeters.mat
```

The data will then be placed into two matrices: `day1_meter` and `day200_meter`. These are 6256x48 matrices of half hourly demand (kW).

You will also find a template Matlab function (`cer_stats.m`) which you should use as the basis for your code. Do not change either the file name or the first line in this file. This is needed for the automated code analysis.

Report and code

You will need to write some Matlab code to be able to answer the following questions. Your report should include three sections, each briefly answering the following questions:

1. Provide a brief summary of the data you are analysing.

2. Demand patterns:

   a) Plot the consumption distributions (frequency plot with 50 bins) for both Day 1 and 2 (separate graphs).

   b) What is the average consumption for Day 1, and include the 90% and 95% confidence intervals for these loads.

   c) What is the average consumption for Day 200, and include the 90% and 95% confidence intervals for these loads.

   d) Compare these two consumption averages. What can you conclude?

3. Outliers: (working at 95% CI using Day 1 consumption)

   a) Based on Day 1 consumption, how many customers are there outside of the 95% CI?

   b) Remove all the outliers that are greater than the Day 1 mean + 95% CI (leave the customers below the lower limit in the dataset). Plot the frequency of the consumption of these remaining customers (use 50 bins).

   c) Using the new dataset: compute the revised mean consumption and 95% CI.

   d) How many customers are now outside the new 95% CI?

   e) Comment on these results.

Helpful hint:

You might find the following statistics library function useful for computing confidence intervals

```
z = norminv(p);
```

This can then be used, for eg:

```
norminv(0.975)
ans =
    1.9600
```

Hint: check this against the normal probability distribution table to familiarise yourself.

Submission instructions

You need to submit the following via duo:

1. Brief report (as indicated above), submitted in PDF format.

   - The report should be no more than 2 pages with a minimum of 10 point font.

   - Do not forget to include your anonymous Z-code at the top of the document!

   - Use PDF format, as other file formats can get incorrectly displayed due to different software versions. This can include graphs and/or equations not displaying, and this will count against you.

2. Your Matlab code (save as `cer_stats.m`).

3. These must be submitted (along with the regression assignment files) as a single zip file. The zip file should not create subfolders. Name your zip file using your anonymous Z-code (eg: `Z0123456` should submit as `Z0123456.zip`). This ensures that all submissions have a unique file name.

Your report will be read and marked in conjunction with your code. The code will be tested and automated scoring will be applied (syntactic correctness, execution time), but may also be manually reviewed. Therefore, you should use an appropriate amount of commenting to guide the reader through your code!

Submission deadline: 14:00 on 16 January 2017 using TurnItIn via DUO.