School of Mathematics

MATH5741M: Statistical Theory and Methods

Continuous Assessment, Part I

Note

This is the first of two pieces of work which count towards your *continuous assessment* component of this module. Overall, the continuous assessment counts for 20% of the total. This piece of work counts for 8% of the total (i.e. 40% of the continuous assessment).

Work to hand in

Please submit your work **as a PDF file** to the VLE (through "turnitin"). Your report should use no smaller than 12 point fonts, be no less than single-spaced, and no more than 6 pages in total (including any title pages, table of contents, apppendices...).

The deadline is midnight on Friday 3rd November, 2017.

Data

A large dataset (around 6 million records) has been collected about crimes in Chicago, USA. The source is available from: https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2

Background to the data, as given on the above website, is as follows:

> *"This dataset reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent seven days. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. In order to protect the privacy of crime victims, addresses are shown at the block level only and specific locations are not identified."*

If you hover your mouse over the "ⓘ" symbols on the web page, you can find out some information about each of the variables.

Note: Although you might try to use the website to create plots and do your analysis, I do not receommend this. All the work you hand in should be done using R.

A random sample of 500,000 of these data was selected, and most of the columns saved in a local csv file, which can be read into R using the command:

```
dd=read.csv("http://www1.maths.leeds.ac.uk/    #type these two as one line
   ~charles/math5741/crime.csv",header=T)
```

Assignment

Summarize this smaller dataset using appropriate graphical and numerical summaries.

Any data cleaning or processing should be explained.

Please highlight each key finding in a succinct, but non-ambiguous sentence before showing the relevant tables/numbers or figures to illustrate/confirm.

(You should not perform any hypothesis tests.)