ST201 Data Analysis

Continuous Assessment – Assignment Sheet 1

Instructions

- Answer the question below for continuous assessment. Due: Before 5pm on 15th October 2021. Upload a Word Document or PDF to Moodle.
- Include your name, student number and tutorial time with your work.

Question

You will find the file `popIreland.rds` in the ST201 folder on the Rstudio server. Open a new R script and load the file into R by copying the following code into your R script:

```
popIreland <- readRDS("~/SharedFiles/ST201/popIreland.rds")
```

In this dataset you will find the population for areas of Ireland (mostly counties) in units of thousands, for the year 2011 and the year 2016. For this assignment you should analyse these data to look at the differences in the 2011 population numbers and the 2016 population numbers.

- The first step in your analysis should be to create and store a variable of population differences, which is the difference between the 2011 population data and the 2016 population data.
- Then you should analyse these differences and produce a 1-page (max) report of your findings. Presenting the report in a word document or equivalent.
- You should upload your Rcode as a separate file or as an appendix.

To achieve full marks the 1-page report should be organised as follows:

Aim

Briefly describe the aim of your analysis.

Data

Introduce the data. Provide a brief description of the data and the variables you intend to analyse.

Analysis

This should include the following:

1. Summary statistics of the **population differences**: mean, median, standard deviation, with an interpretation.
2. A boxplot of the **population differences**, with an interpretation. Note: be sure to label your plot axis appropriately.
3. An ecdf plot of the **population differences**. Comment on some specific percentile values (e.g., 80th percentile) based on the ecdf plot (you shoud confirm what you see using the quantile function).

4. Comment overall on the % of areas (counties) that saw an increase/decrease in population between 2011 and 2016.

Conclusions

Briefly mention what you have concluded from your analysis.