Data-Driven Decision-Making

Assignment Two (Re-Assessment), 2020/21

Assignment Information

Upload your solution to Moodle by the deadline indicated on Moodle. You risk being penalized if you submit late without prior agreement. Please see further submission information on Moodle.

This is an individual piece of assessment and the work submitted should be entirely your own and prepared specifically for this assessment. You are not allowed to collaborate with other people or to copy the work of others.

Your coursework will be electronically checked for plagiarism. In the event of any doubt about authorship, you will be interviewed by the School of Computing Academic Conduct Officer and may be asked questions about any aspect of the work.

Background

You have been hired by the World Health Organisation to help understand what factors influence a country's life expectancy. They have gathered a dataset with information on countries from the year 2000 to 2015. Data on the economics, the healthcare, the immunisation levels and other information was collected for each country and in each year.

Your analysis will be used to give recommendations to governments on how they can improve life expectancy. When analysing the data, think carefully about which factors are likely to be under a government's control. What meaningful advice can you give?

You will have to make decisions about how to clean the data: in particular, how to deal with missing values.

The data is available in the file `life_expectancy.csv`, which is posted on Moodle.

Below is a description of some of the variables in the data.

- **status** Developed or Developing status
- **life_expectancy** Life Expectancy in years
- **adult_mortality** Adult Mortality Rates of both sexes (number of people dying between 15 and 60 years per 1000 population)
- **infant_deaths** Number of Infant Deaths per 1000 population
- **alcohol** Alcohol, recorded per capita (15+) consumption (equivalent to liters of pure alcohol)
- **percentage_expenditure** Expenditure on health as a percentage of Gross Domestic Product per capita (%)
- **hepatitis_b** Hepatitis B (HepB) immunization coverage among 1-year-olds (%)
- **measles** Measles - number of reported cases per 1000 population
- **bmi** Average Body Mass Index of entire population
- **under_five_deaths** Number of under-five deaths per 1000 population
- **polio** Polio (Pol3) immunization coverage among 1-year-olds (%)
- **total_expenditure** General government expenditure on health as a percentage of total government expenditure (%)
- **dtp3** Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-yar-olds (%)
- **hiv_aids** Deaths per 1 000 live births HIV/AIDS (0-4 years)
- **gdp** Gross Domestic Product per capita (in USD)
- **population** Population of the country
- **schooling** Number of years of Schooling

For this assignment you should produce the following:

1. A report on the data set you have been given which analyses the drivers behind life expectancy. This report is for other analysts and can use technical language. You will be expected to use appropriate statistical methods in support of your argument. (**Max 2000 words)**
2. An executive summary for other people in The World Health Organisation. The people reading the executive summary will not be experts on data analysis, so this report should be clear to a layperson. The summary should discuss what you have found in the data, and also discuss the importance of the organisation using data. **(Max 1000 words)**
3. The code used to generate the conclusions found in the data analysis report (part 2). This code will be checked for correctness and clarity. Note you do not have to submit any code used to generate part 3.

When you submit, please:

1. Combine parts 1 and 2 into a single PDF and submit at the document submission upload link on Moodle.
2. Upload part 3, either as a .R file or a .Rmd file at the code upload link on Moodle.

We strongly recommend consulting the guidance in the marking scheme below.

Marking Breakdown

Data analysis report – 30 marks

Executive summary – 20 marks

Code for Data analysis report – 20 marks

Total – 70 marks

Guidance on how your reports will be marked

Part 1

The mark breakdown is as follows

- correctness of analysis (7)
- completeness analysis (8)
- story from the data (8)
- depth of analysis (7)

We are looking for:

- The analysis will be fully correct
- A notably thorough exploration of all variables that could answer the question, with appropriate emphasis on the most important variables.
- Strong conclusions but with appropriate levels of uncertainty well explained.
- Report focuses on the aspects of the analysis which are most important.
- Conclusions are logical and well thought through.
- One model explained in detail, or several models but explained in less detail. Clear that the you have an excellent understanding of the model. The results are explained very well.

Part 2

This part will be marked not on the content of the analysis – this is covered in prior parts – but on the focus and clarity of explanation. Pay attention especially to the trade-off between presenting more detailed results versus ease of comprehension.

The breakdown of marks is as follows:

- clarity of presentation (7)
- focus on key priorities (6)
- the of graph / visualisations in support of arguments (7)

We are looking for:

- Exceptionally clear, report easy to read and understand
- Clear identification of priorities and strong focus throughout
- Good use of visualisations, well selected, tailored to arguments, clearly integrated

Part 3

The mark breakdown is as follows

- correctness of code [6]
- code matches analysis [6]
- code style [8]

We are looking for:

- A particular effort has been made to make the code easy to run. All code delivered runs correctly on a computer with the same packages installed. Appropriate amount of code.
- Effortless to see how code matches analysis. Probably uses Rmarkdown.
- Code very easy to read. Code has a strong and clear structure. Formatting is consistent. Comments where needed.