MXB344 Assignment 1 Task Description

Due: Friday 21st April 2023

Introduction

This document describes the context of your first MXB344 assignment. Also see the assessment marking criteria in the Task Criteria document on Canvas.

Scenario

You are an analyst working for a global industrial manufacturing company. Your company employs over 10,000 workers across a number of large factories internationally. Your production network spans Asia, Europe and South-America each with their own unique industrial regulations. A recent unfortunate accident in South America has put your company's safety record under the microscope. Journalists have been sending your CEO difficult questions about your company's workplace safety practices, which have been identified as highly variable across your network.

Your CEO has requested you drop everything and perform some analysis on workplace injury data to help inform your company's response to this growing crisis. She would like to know:

1. Of the various safety regimes in place across your company, which one would you recommend become the international standard for your company, based solely on injury prevention performance?
2. It has been suggested by senior management that industry experience is more important than the safety regime when it comes to preventing injuries. His idea is that a policy directed instead at lowering employee turnover to reduce injury rates. Does available data support this assertion?

The Data

You have obtained a CSV file called `injury.csv` (see Canvas). It contains counts of injuries and hours worked aggregated by the experience level of the workers and the workplace safety regime in place at their factory. The data are for the last 12 months of operation.

Specifically the variables are:

- `Injuries` - count of injuries in group
- `Safety` - the safety regime in place for group
- `Hours` - total hours worked by this group
- `Experience` - the experience level of group

Tasks

Task 1 Statistical analysis (70 marks total)

Conduct a regression analysis for the injury counts. Motivate this analysis using your CEO's queries. Draw conclusions that clearly address the queries. Document and develop your analysis in a single rmarkdown document. The audience of this document is another analyst, so you should clearly outline the question/s being addressed, the methods applied and the conclusions drawn, with every step/decision being justified.

Base all of your conclusions on a single fitted statistical model. Validate the assumptions of the model you have found, including a fixed or estimated overdispersion parameter based on the mean-variance relationship of your observations. Appropriately assess the fit of your model.

Task 2 Summary on a Page (SOAP) (30 marks total)

Produce a 1 - 2 page summary for your CEO that addresses her two queries directly. This must include at least one plot. Utilise graphics to make your points clear wherever possible. Some considerations:

- Nominate the methods used but do not describe them in detail.
- Base your assertions and recommendations on evidence from your analysis.
- Do not present the effect of a covariate without communicating the uncertainty around that effect. State confidence intervals and show confidence bounds on plots.
- Be concise. Dot points are appropriate.
- This is not the work, it is the advertisement for your work in Task 1. In the real world, people are unlikely to look at the work if the advertisement isn't clear and engaging.

I would encourage you to use RMarkdown for this document, however, html and pdf are also acceptable.

Submission

Submission of this assessment with via the Assessment 1 Submission page on Canvas. Please note that this assessment item is due at 11.59pm Friday, Week 7, 2023. This is a strict deadline, and only files submitted by Sunday 23rd April (48 hours automatic extension applies to this assessment) will be marked. Hence, it is worth submitting your assignment early and double checking that you have attached the correct files.

Submission Format

Keeping your submission neat and tidy will assist in grading. Create a README.txt file if you need to give me some instructions. Ideally your set of files will contain only:

1. The report of your analysis in single Rmarkdown document e.g. `Injury_Analysis_2023.Rmd` with the knitted pdf file e.g. `Injury_Analysis_2023.pdf` .
2. The data file, `injury.csv` , from Canvas
3. Your SOAP e.g. `Injury_SOAP_2023.pdf` .
4. Your README.txt file. (Optional)

Good Luck!