

MXB344 Assignment 1 Task Description

James McGree

31 March 2025

Introduction

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

Scenario

You are a data 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 an analysis on workplace injury data to help inform your company's response to this 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 should be developed that is directly related to lowering employee turnover will reduce injury rates. Your CEO would like to know if the available data support this assertion.
3. Your CEO is also interested in if there appears to be any relationship between:
 - Injuries and the annual bonuses a proportion of employees received
 - Injuries and whether staff have received any formal external qualifications e.g. external safety training or a university degree.

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 group
- `Experience` - the experience level of group
- `bonus` - proportion of group who received an annual bonus last year
- `training` - proportion of group who have completed external safety training
- `university` - proportion of group who have at least one university degree

Tasks

Task 1 Statistical analysis (70 marks total)

Conduct a regression analysis using a generalised linear model for the injury counts. Motivate this analysis using your CEO's queries such that all conclusions you draw clearly address the queries. Base all of your conclusions on a single fitted generalised linear model that is only as complex as it needs to be. Validate the assumptions of the model you are proposing, including a fixed or estimated overdispersion parameter. Appropriately assess the goodness-of-fit of your model. Ensure uncertainty is communicated throughout the results, interpretation of results and in your conclusions.

Document and develop your analysis in a single Rmarkdown document. The audience of this document is another data analyst, so you should clearly outline the question/s being addressed, the methods applied and the conclusions drawn, with every step/decision being justified.

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

Produce a 1 - 2 page summary for your CEO that addresses her 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 is not 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 via Canvas. Please note that this assessment item is due at 11.59pm Friday, Week 9. This is a strict deadline, and only files submitted by this time will be marked (unless you have received an extension). 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. Your analysis in Rmarkdown form. e.g. Injury_Analysis_2025.Rmd
2. Your compiled/knitted report as a pdf e.g. Injury_Analysis_2025.pdf
3. The data file, injury.csv, which you got from Canvas.
4. Your SOAP e.g. Injury_SOAP_2025.Rmd or Injury_SOAP_2025.pdf etc.
5. Your README.txt file. (Optional)

Good Luck!