MXB344: Assessment 1 Criteria

Task	Modelling Task
Unit Learning Outcomes Addressed	1. Expertly and critically carry out statistical analysis using statistical models in the analysis of various data sets and examples. 2. Use R to carry out statistical analyses. 3. Communicate statistical conclusions clearly and concisely both in written form and orally.
Due date	11.59pm, Friday, Week 7
Weighting	20%
Specifications	Individual

Overview

This first assessment task is designed to give you the chance to apply regression modelling to count data in a real industry scenario. It will introduce you to tools and artefacts relevant to applied statistical problems in industry.

What you will do

1. Conduct an analysis of data describing workplace injuries with a view to answering specific industry questions.
2. Create a 'Summary On a Page' (SOAP) document, telling a story with your analysis, results and communicating your conclusions.
3. Conduct your analysis using R and document your analysis using Rmarkdown, so that the code and the analysis are held together and are reproducible.
4. Submit your assignment electronically via Canvas

What you will submit

Main deliverables:

1. A statement on a page which describes your analysis and conclusions at executive level detail.
   a) Should contain at least one plot that summarises your results as they relate to the original queries.
   b) Appropriately communicate conclusions (with uncertainty) in a way that is accessible to non-statisticians.
2. A report describing your analysis, methodology and conclusions.
   a) Should be written in Rmarkdown to ensure reproducibility.
   b) Should include exploratory analysis plots with comments.
   c) Analysis should use appropriate model formulation and model checking procedures:
      i. Justification of likelihood.
      ii. Analysis of residuals.
      iii. Justification for choice of fixed or estimated overdispersion parameter.

Resources and Useful References

1. Poisson Regression Lecture Notes
2. Cross validated and Stack Overflow websites. E.g: http://stats.stackexchange.com/questions/66791/where-does-the-offset-go-in-poisson-negative-binomial-regression
3. Rmarkdown documentation: http://rmarkdown.rstudio.com/
4. Story Telling with Data (http://www.storytellingwithdata.com/) https://www.youtube.com/watch?v=X79o46W5pII
5. Canvas page for this Project.

Task / Grade	7	6	5	4	3	2-1
Task 1 (Analysis): Content	Motivates analysis clearly using scenario context. Conducts exploratory analysis to identify unusual observations and relationships likely to be useful in generalised linear modelling. Assesses quality of statistical model fit and validity of model assumptions using plots and formal procedures where appropriate. Appropriately assesses model fit. States conclusions and recommendations addressing motivating queries based on evidence from modelling. Communicates uncertainty around evidence/effects during discussion of them. Identifies interesting unresolved	Motivates analysis clearly using scenario context. Conducts exploratory analysis to identify relationships likely to be useful in generalised linear modelling. Assesses quality of statistical model fit and validity of model assumptions using plots and formal procedures where appropriate. Appropriately assesses model fit. States conclusions and recommendations addressing motivating queries incompletely based on evidence from modelling. Communicates uncertainty around evidence/effects at some stage.	Analysis not clearly motivated using scenario context. Conducts some exploratory analysis without explicitly stating bearing on analysis decisions. Assesses quality of statistical model fit and validity of model assumptions using plots and formal procedures. Assesses model fit. States conclusions and recommendations addressing motivating queries incompletely based on evidence from modelling. Communicates uncertainty around evidence/effects at some stage.	Analysis not clearly motivated using scenario context. Conducts some exploratory analysis without explicitly stating bearing on analysis decisions. Assesses quality of statistical model fit and validity of model assumptions only informally. Informally assesses model fit. States conclusions and recommendations addressing motivating queries incompletely based on evidence from modelling. Communicates uncertainty around evidence/effects at some stage.	Analysis not clearly motivated using scenario context. Conducts some exploratory analysis without explicitly stating bearing on analysis decisions. Fails to assess quality of statistical model fit or assumptions in coherent way. States conclusions and recommendations addressing motivating queries without consideration to evidence from modelling. Fails to Communicate uncertainty around evidence/effects.	Analysis not clearly motivated using scenario context. No exploratory analysis. Fails to assess quality of statistical model fit or assumptions in coherent way. States conclusions and recommendations addressing motivating queries that are incorrect. Fails to Communicate uncertainty around evidence/effects.

	questions that arise from analysis and makes suggestions as to further data that could be acquired to explore them.					
Task 1 (Analysis): Format	Analysis format is an R markdown document. All required R code to complete analysis is embedded in document in relevant places, but does not make inordinate amounts of code visible in final output. The document is inherently reproducible, it can be knitted to html in a clean R environment without errors.	Analysis format is an R markdown document. All required R code to complete analysis is embedded in document but placement may make it difficult to locate. Does not make inordinate amounts of code visible in final output. The document is inherently reproducible, it can be knitted to html in a clean R environment without errors.	Analysis format is an R markdown document. Some R code required to complete analysis is missing. Some R code blocks or output disrupt flow of commentary in final output. The document is inherently reproducible, it can be knitted to html in a clean R environment without errors.	Analysis format is an R markdown document. Some R code required to complete analysis is missing. Many R code blocks or output disrupt flow of commentary in final output. The document is not inherently reproducible, it cannot be knitted to html in a clean R environment without fixing errors.	Some R code required to complete analysis is missing. Many R code blocks or output disrupt flow of commentary in final output. The document is not inherently reproducible. it either cannot be knitted to html in a clean R environment or is not an R markdown document.	R code required to complete analysis is missing or in a separate file. The document is not inherently reproducible. it either cannot be knitted to html in a clean R environment or is not an R markdown document.
Task 2 (SOAP): visualisation	Clear evidence in design of visualisations with intent to communicate information relevant to queries driving analysis. The design is engaging and transmits information in easy to understand	Clear evidence in design of visualisations with intent to communicate information relevant to query driving analysis. The design is engaging and transmits information in easy to understand way. The	The information that visual elements are intending to communicate is not evident at first sight. The design is apparently engaging however does not transmit information in easy to understand	The visualisation chosen is not appropriate for the combination of audience and information. The design is not engaging and does not transmit information in an easy to understand way.	The visualisation chosen is not appropriate for the combination of audience and information. The visualisation is misleading and confusing. The visualisation does not	No evidence of design in considering audience or information. The visualisation is misleading and confusing. The visualisation does not use appropriate axes and legends. The

	way. The design uses appropriate axes and legends.	design attempts to use appropriate axes and legends, however some minor flaws with colours, labels, or scales makes the overall visualisation harder to understand at first sight.	way (might confuse the reader). The design attempts to use appropriate axes and legends, however some minor flaws with colours, labels, or scales make the overall visualisation harder to understand at first sight.	The design attempts to use appropriate axes and legend, however some evident flaws with colours and/ or labels and /or scales make the overall visualisation harder to understand. The visualisation is technically correct and uses correctly selected data, however assumes too much knowledge to interpret the visualisation correctly.	use appropriate axes and legends.	visualisation is technically incorrect.
Task 2 (SOAP): Communicating Conclusions	Actionable recommendations are made that address the queries driving analysis. Recommendations are linked to evidence from analysis. The uncertainty around effects is addressed in a clear way through both visualisation and description. Caveats or debatable assumptions from analysis are stated.	Actionable recommendations are made that address the queries driving analysis. Recommendations are linked to evidence from analysis, but that link may be unclear at first sight. The uncertainty around effects is addressed though it may be difficult to interpret. Caveats or debatable assumptions from analysis are	Actionable recommendations are made that address the queries driving analysis. Recommendations are presented alongside evidence from analysis, with links to be drawn by reader. The uncertainty around effects is addressed though it may be difficult to interpret. A caveat or debatable assumption	Actionable recommendations are made that do not fully address the queries driving analysis. Recommendations are presented alongside evidence from analysis, with links to be drawn by reader. The uncertainty around effects is not addressed. A caveat or debatable assumption from analysis is not stated. Level of	Actionable recommendations are made that do not fully address the queries driving analysis. Some recommendations are presented without supporting evidence. The uncertainty around effects is not addressed. A caveat or debatable assumption from analysis is not stated. Level of technical detail and volume of	Actionable recommendations are not made or do not address the queries driving analysis. Recommendations are presented without supporting evidence. The uncertainty around effects is not addressed. Caveats or debatable assumptions from analysis is not stated. Level of technical detail and volume of content is

	Level of technical detail and volume of content is appropriate for CEO level (Non-technical decision maker with limited attention span).	stated. Level of technical detail and volume of content may be slightly inappropriate for CEO level (Non-technical decision maker with limited attention span).	from analysis is not stated. Level of technical detail and volume of content may be slightly inappropriate for CEO level (Non-technical decision maker with limited attention span).	technical detail and volume of content is inappropriate for CEO level (Non-technical decision maker with limited attention span).	content is inappropriate for CEO level (Non-technical decision maker with limited attention span).	inappropriate for CEO level (Non-technical decision maker with limited attention span).