MXB334: Assessment 1 Description

|  |  |
| --- | --- |
| 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 | Week 5 |
| Weighting | 40% |
| Specifications | Individual |
| Overview | |
| This first assessment task is designed to give you the chance to apply generalised liner modelling to count data in a realistic 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 into data describing workplace injuries with a view to answering a specific query. 2. Create a ‘Summary On a Page’ (SOAP), telling a story with your analysis results and communicating your conclusions. 3. Conduct your analysis using R and document your analysis using R markdown, so that the code and the analysis are held together and are reproducible. 4. Commit the analysis in stages to a github repository. | |
| What you will submit | |
| Main deliverables:   1. A statement on a page which describes your analysis and conclusions at executive level detail.    1. Should contain at least one plot that summarises your results as they relate to the original query. 2. A report describing your analysis methodology and conclusions.    1. Should be written in Rmarkdown to ensure reproducibility.    2. Should include exploratory analysis plots with commentary on contrasting or unusual obersations.    3. Analysis should use apropriate model formulation and model checking procedures:       1. Justification of likelihood and link function selection.       2. Analysis of residuals.       3. Justification for choice of fixed or free dispersion parameter.          * An answer to: What is the resulting probability mass function if the fixed dispersion parameter of the poisson distribution is allowed to vary? | |
| 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=X79o46W5plI> 5. Blackboard Folder for this Project. | |

# A Top Level Effort

## SOAP

* Clear evidence in design of visualisations with intent to communicate information relevant to query driving analysis.
  + Axis labels, axis scales, legends that support this.
* Clear explanation or visualisation of the uncertainty around analysis outputs that are relent to the query.
* May name methods used but avoids inappropriate explanations of them.
* Unambiguously states the outcome of the analysis in a way that directly relates to initial query.
* Contains link or reference to the repository of the analysis report.

## Report

* Articulates aim of investigation.
* Exploratory analysis creates intuition for results of statistical modelling.
* Makes justification for assumptions in terms of model form.
* Clearly assesses quality of statistical model fit and validity of assumptions
  + Plots of deviance residuals
  + Pearson Residual Chi-Sq test or Test of residual deviance (equivalent)
  + Leverage plots
  + Highlights issues and makes sound arguments for their resolution either backed up by theory or references.
* Explores the issue of under and over dispersion.
  + Fits free and fixed dispersion parameter. Makes an augment for free or fixed.
  + Clearly shows how the quasi-poison probability mass function can be obtained from the canonical exponential family form of the poisson.
* Has all required R code embedded in Rmarkdown document in relevant places, but does not make inordinate amounts of code visible in final output.
* Is inherently reproducible.
  + Can be executed and knitted from an empty environment without errors or numerous warnings.
* Clearly maps results of modelling to conclusions reached relating to initial query that flow through to SOAP.
* Identifies interesting unresolved questions that arise from analysis and makes suggestions as to further data that could be acquired to explore them.

# A Passing Effort:

## SOAP

* The visualisation lacks elements that make it immediately obvious what it is communicating.
* Uncertainty is expressed in a way that is technically correct but requires some mental translation by executive.
* Includes details of analysis not relevant to executive.
* States outcome of analysis using too many caveats just too many words.

## Report

* Clearly articulates report aims.
* Exploratory steps are lacking commentary that clearly relates them to aims.
* Ignores or glosses over issues that arise with checking statistical assumptions and model fit
* Explores issues of dispersion but provides weak justification for selected dispersion parameter.
* Is inherently reproducible but code organisation may be lacking. Code may not be relevant to section of report it is in.
* May display too much code in the report in a way that breaks up logical flow.
* Ignores or glosses over questions arising from analysis.
* Omits no more than 1 standard measure or plot relating to testing assumptions or model fit.
* Clearly maps results of modelling to conclusions reached relating to initial query that flow through to SOAP.