<div align="center">

ACSC/STAT 3740, Actuarial Models II
WINTER 2023
Toby Kenney
Instructions for Project.

</div>

General Instructions

The project should consist of a thorough analysis of a dataset, going through all the stages covered in the course. It should be written up as a report for a target audience with limited data science background (e.g. a manager in a company, or a scientist analysing an experiment). You are free to choose a data set to analyse, and a reasonable question that might plausibly be of interest. The question can be vague or specific.

Choice of Data

There are a large number of publically available data sets. Some are included by default in `R`. Other websites with data sets include `http://lib.stat.cmu.edu/datasets/` (this site also has links to other sites with data sets available) and `https://www.kaggle.com/datasets` (this appears to need a free registration to download data sets). There are also many websites that provide data sets on particular topics.

The data set should be rich enough to allow a detailed analysis — so for example, a data set with only one variable, or with so few observations that only the simplest model can be fitted would not be ideal. You can choose any topic that interests you, but I strongly recommend avoiding sports, since there is a danger that projects based on sports will make assumptions that are non-obvious to people not familiar with the sport. There is also the danger that preconceived opinions may make it more difficult for me to grade the project fairly.

Many available data sets have already been analysed in the literature. You may discuss previous analysis of the data, but make sure that your analysis is different from previous work. A good way to achieve this in some cases is to analyse the data to answer a different question.

Report

You should write up a full report as if you were analysing the data for an employer/client/collaborator. It should be written for an audience with limited data science background, but the technical parts should be detailed enough that another data scientist could repeat your analysis. The report should be long enough to perform a complete data analysis and explain your analysis and main conclusions. If the report (excluding tables and figures) is less than 10 pages in length, then it is likely that either the data set you have chosen is too simple, or you have not given enough consideration to the project.

Grading

The grading of this project will be based on demonstrating good judgement and creative approaches to the challenges presented by the data, rather than technical accuracy. Therefore, a dataset that is too

simple will limit the scope for receiving top grades. A good treatment of a challenging dataset will receive more credit than a "perfect" treatment of an easy dataset.