STSB6816 Test 2 of 2024

Mathematical Statistics and Actuarial Science; University of the Free State

2024/05/23

## Time: 170 minutes; Marks: 45

# Instructions

* Answer all questions in a single R Markdown document. Please knit to PDF or Word at the end and submit both the PDF/Word document and the “.Rmd” file for assessment, in that order.
* Label questions clearly, as it is done on this question paper.
* All results accurate to about 3 decimal places.
* Show all derivations, formulas, code, sources, and reasoning.
* Intervals should cover 95% probability unless stated otherwise.
* No communication software, devices, or communication capable websites may be accessed prior to submission. You may not (nor even appear to) attempt to communicate or pass information to another student.
* Use of AI tools must be disclosed and summarised.

# Introduction

The data is provided at <https://ufs.blackboard.com>. It is a manipulated and distorted sample of events for a group life insurer, that is an insurer who sells insurance to a company to cover the lives of all their employees as a group. Every employee in a company pays the same premium and then when an individual dies their beneficiaries and the company get a small payout to help cover the resulting expenses.

The sample was distorted be removing about 200000 non-events randomly. This should not affect the statistical significance of independent variables, but does massively inflate the rate of deaths. Pretend that this insurance is for mercenary companies in a war zone, then the numbers will make sense.

Your objective is to fit an appropriate generalised linear mixed effects model on the target events. The choice of explanatory variables is largely up to you, but must include at least age last birthday (ALB) and gender as fixed effects, along with clients as random intercepts. Use of log salary is optional, as is interactions, but do not use any of the other variables.

# Question 1

**1.1)** You should fit two of the following models and compare them to determine which best fits the data:

1. Poisson regression (log link)
2. Logistic regression (logit link)
3. Negative Binomial (Type II) regression (log link)

Additionally, you must use at least two selections of explanatory variables and compare those. So, in total, you must fit at least 3 models and compare them. Explain each model structure used briefly and explain what the result of the model comparison implies. **[20]**

*[Hint: if you have trouble comparing models for any reason then use any one reasonable fit to answer the rest of the test, as if it was the best model.]*

**For the best model:**

**1.2)** Analyse the fixed effect parameter summaries in depth. **[5]**

**1.3)** Compare the distributions of the predicted event (death) rates of the different clients visually (adjusted for client size). State which client has the highest expected event rate (on any basis). **[7]**

[Hint: The key steps are: obtaining an explanatory matrix, multiplying by the parameter matrix, and applying the appropriate transformation. You only need to do it for Client N in order to answer the questions to come. ]

**1.4)** Compare the distribution of the standardised predicted payouts for Client N versus the average client (per individual) as follows: subtract the distributions and illustrate the resulting density. **[7]**

[Hint: the standardised predicted payout for a client is the dot product of the event probabilities and the sum assured values, divided by the number of individuals within the client.]

**1.5)** Assume that premiums are set for each client at the 60% quantile of the expected payout for that client, taking the fixed effects into account but not the random effects. Calculate the premium for Client N, standardised by client size. **[3]**

**1.6)** Consider again Client N. If they were a new random client instead of an existing client, how much higher would you set their premium? Repeat the premium calculation including random draws from the between client distribution and produce an exact figure per individual. **[3]**