

Assessment 4: Generalised Linear Models (Topic 6)**Due 11th May****Total marks: 100**

Refer to the **Assessment Criteria and Guidelines** in the Assessment block of the Moodle site. **Five marks** will be awarded for the following:

- Clear expression including correct use of terminology and notation
- Presentation of figures and tables including relevant R output.
- Clearly and concisely annotated R code.

Question 1**[45 marks]**

A longitudinal study was conducted to assess whether olfactory dysfunction is a potential marker for dementia. The study recorded odour identification scores in 1,055 randomly selected hospital outpatients aged over 65 years of age, excluding any patients who had been diagnosed with dementia or cognitive decline. Patients were asked to identify 10 odours and the score was determined as the number of correctly identified odours out of 10. After a 5 year period, the odour scores of these patients were then linked to whether the patients has been subsequently diagnosed with dementia.

The data are given below:

Odour score	0	1	2	3	4	5	6	7	8	9	10	Total
patients with dementia	24	20	26	25	27	29	27	29	30	23	18	278
Total	75	64	86	87	93	96	92	105	115	112	130	1055

- State and fit a generalised linear model for the proportion of patients who were diagnosed with dementia against the odour score recorded. Explain all components of the model. [15 marks]
- Is the model a good fit? Refer to the relevant information from the analysis of deviance table. [10 marks]
- Give an informative interpretation of the summary table, including a practical interpretation of the coefficient of odour score. [10 marks]
- Estimate the odour score value that corresponds to 25% of patients being diagnosed with dementia. Show your calculations. [10 marks]

Question 2: Who survived the Titanic?

[50 marks]

The Titanic was a British luxury passenger liner that sank when it struck an iceberg about 640km south of Newfoundland on April 14-15, 1912, on its maiden voyage to New York City from Southampton, England. Of 2201 known passengers and crew, only 711 are reported to have survived. The data are given in `Titanic1912.txt`. People are classified by their **sex**, **age** (adult or child) and **class** (either first, second, third class or crew). For each age/sex/class combination, the number of people (**total**) and the number of surviving (**survive**) are also included.

- (a) Calculate the proportion of survival in each age/sex/class combination. On the basis of examination of the data, give a brief summary of how the survival rate is affected by **class**, **age** and **sex**. [10 marks]

- (b) Fit a logistic regression model that includes all the factors **class**, **age** and **sex**, plus all of the two-way interactions. Use appropriate testing procedures to decide if any interaction terms can be eliminated.

State the equation for the final model and summarise the results (maximum 1 page) you have obtained by interpreting the regression coefficients, describing the difference in survival rates for various factor combinations. [40 marks]

(Hint: How does the survival of the crew differ from the passengers? First class from third class? Males from Females? Children vs adults? Did children in first class have higher survival rate compared to the children in third class ?)