

MAB624 Applied Statistics 2

Assignment 4

Generalised Linear Models: Counts (weight 15%)
Due: 5pm Friday 16 October

Background

The table below summarises the results of a prospective study conducted in Canada on smoking and health. One of the purposes of the study was to investigate the relationships between smoking habits and mortality from chronic diseases, particularly lung cancer. The study was initiated by a questionnaire and data on the deaths occurring among the respondents were also collected over a six-year follow-up period. The table shows the male populations and the numbers of deaths in each of several age groups, classified according to whether or not they were smokers and, if so, what type of smoker. The data are stored in the data set `smoking.csv`.

Age	Non-smokers		Cigar and pipe only		Cigarette and other		Cigarette only	
	Deaths	Pop.	Deaths	Pop.	Deaths	Pop.	Deaths	Pop.
40–44	18	656	2	145	149	4531	124	3410
45–49	22	359	4	104	169	3030	140	2239
50–54	19	249	3	98	193	2267	187	1851
55–59	55	632	38	372	576	4682	514	3270
60–64	117	1067	113	846	1001	6052	778	3791
65–69	170	897	173	949	901	3880	689	2421
70–74	179	668	212	824	613	2033	432	1195
75–79	120	361	243	667	337	871	241	436
80+	120	274	253	537	189	345	63	113

A description of the variables in the data set is as follows:

Variable	Description
Age	Age group (ordinal or continuous)
Smoke	Type of smoker: no = non-smoker cigarPipeOnly = cigar and pipe only cigarettePlus = cigarette and other cigaretteOnly = cigarette only
Deaths	Number of male deaths
Pop	Number of males in respective population (exposure variable)

Task

Carry out an analysis of the data to investigate how the rate of deaths depends on age and smoking status by developing an appropriate Poisson generalised linear model for the **rate of deaths**, with **Age** and **Smoke** as explanatory variables. Use the **GENMOD** procedure in SAS with a **log link** to carry out your analysis.

In developing your model you are **not** expected to:

1. Investigate the use of alternative link functions. You have been instructed to use the logarithmic link function (only).
2. Investigate possible transformations of the explanatory variables. If your model diagnostic plots suggest that transformation(s) may be required, then you just need to comment on this.
3. Remove observations from the data set. But you should comment on any unusual observations.

Reports can be submitted individually or as a group of no more than three students. Write the **names**, **student numbers** and **email addresses** of all contributors on your report so that marks can be awarded and feedback provided to all contributors.

You are required to write up your investigation as a **word processed report**. You may use any word processor you like. However, you will need to choose a word processor that can write equations, e.g., Microsoft Word or LaTeX. You must submit your report as a **PDF file**. See the sections on Report Specifications and Submitting your Report for further details on setting out and submitting your report. Note that it is not sufficient to simply give your final model and SAS code, you must communicate your modelling approach, model assessment and model interpretations within your report, see the section on Suggested Approach for guidance.

Suggested Approach

To get you started, below are suggestions as to how you might approach this data analysis problem. Please note that these are only *suggestions*, they do not form an exhaustive list of all steps that may be required in your investigation.

1. Carry out some exploratory data analyses by producing an appropriate plot(s) of the sample data.
2. Develop an appropriate generalised linear model relating the response to the explanatory variables. In assessing statistical significance of terms you should allow for under- or over-dispersion if appropriate.
3. Assess the goodness of fit of your models.
4. Examine and interpret diagnostic statistics and plots to check the validity of your models.
5. **Age** can be treated as an ordinal categorical variable or a continuous variable in this data set. In developing your model, consider whether a model that includes **Age** treated as a continuous variable is a better/worse model than a model that includes **Age** as a categorical variable.
6. Your findings should include, amongst other things, a clear statement of your final model, interpretation of your model, an appropriate measure of goodness of fit of your model, and a graphical summary if appropriate.

Report Specifications

1. As with all scientific reports, your report should be written in sections. One example might be

Report

- 1 Introduction
- 2 Model
 - 2.1 Distribution
 - 2.2 Link
 - 2.3 Model equation
- 3 Model Fitting
 - 3.1 Model selection
 - 3.2 Parameter estimates
 - 3.3 Goodness of fit and model validity
- 4 Conclusions

Appendix

A SAS Code

2. Your **SAS code** should be placed **in an appendix** at the end of your report. Your code should be **written in a logical order**. You must briefly **comment your SAS code** so that it is clear what you are trying to do in each segment of code. You **should not include any irrelevant code**.
3. Only put **relevant SAS output in the body of your report** along with your written comments. Put the **remainder of the output in a separate PDF file as supplementary material**.

Submitting your Assessment

By submitting your assessment you affirm that the work is your own except where the words of others are specifically acknowledged through the use of inverted commas and in-text references and that this assessment has not been submitted in whole or in part for any other unit at QUT or any other institution.

Individual assessment submission: If you completed the assessment by yourself, and not in a group, then submit your assessment as per the instructions below.

Group assessment submission: Only one submission is required per group. Nominate one group member to submit your assessment as per the instructions below.

1. Give your report PDF file, that contains your report, the name **report.pdf** and give your supplementary material PDF file (if you have one), that contains your additional SAS output (if any), the name **supplementary.pdf**
2. Zip your files into one zip file. Give the zipped file a name with the format

studentnumber__ surname__ firstname.zip

For example, Mary Smith has student number 1234567 so she would name her zipped file

1234567__ smith__ mary.zip

3. Submit your zipped file through the MAB624 Blackboard site by going to

Assessment > Assignment 4

Note that you need to choose the 'submit' button and not the 'save' button. If multiple submissions are received, only the most recent submission will be marked.