Assignment 8 - Poisson and Logistic Regression. Due December 1, 11:59pm 2021

EPIB607 - Inferential Statistics[a]

[a]McGill University

This version was compiled on November 23, 2021

In this assignment you will practice poisson and logistic regression. State all assumptions. Provide confidence intervals with appropriate units. Answers should be given in full sentences (DO NOT just provide the number). All graphs and calculations are to be completed in an R Markdown document using the template from previous assignments. You are free to choose any function from any package to complete the assignment. Concise answers will be rewarded. Be brief and to the point. Please submit the compiled pdf report to Crowdmark. You need to save your answers to each question in separate pdf files. You also need to upload your code separately to Q3. See https://crowdmark.com/help/ for details.

Rates | Parameter contrasts | Regression

1. (50 points) Population mortality rates in Denmark

The following table contains mortality data for males and females in Denmark for 4 age groups over three time periods.

Year	Age	Female_deaths	Female_PT	Female_rate	Male_deaths	Male_PT	Male_rate
1980-1984	70-74	15989	586882.8	0.0272439	23810	456908.21	0.0521111
1980-1984	75-79	20838	454142.7	0.0458843	24707	300318.92	0.0822692
1980-1984	80-84	24073	297678.6	0.0808691	20319	167303.51	0.1214499
1980-1984	85-89	20216	147771.7	0.1368057	13524	74295.83	0.1820291
2000-2004	70-74	13912	521561.9	0.0266737	17360	436994.92	0.0397259
2000-2004	75-79	19731	471945.5	0.0418078	22477	341362.82	0.0658449
2000-2004	80-84	25541	369989.9	0.0690316	22992	217929.72	0.1055019
2000-2004	85-89	27135	226798.1	0.1196439	17444	104009.58	0.1677153
2005-2009	70-74	12179	540568.6	0.0225300	15782	472012.84	0.0334355
2005-2009	75-79	17273	444474.2	0.0388616	19547	344351.34	0.0567647
2005-2009	80-84	23513	363534.1	0.0646789	21781	230530.24	0.0944822
2005-2009	85-89	26842	237877.3	0.1128397	17811	114485.04	0.1555749

a. (15 points) Come up with a suitable regression model for this data. Write down the regression equation in terms of parameters and determinants.
b. (15 points) Estimate the parameters of this model using the data in the table above. Provide the fitted regression equation. The data is provided in the `denmark.csv` file in myCourses.
c. (10 points) Interpret the parameter for gender. Are mortality rates significantly different in males compared with females?
d. (10 points) Perform a goodness of fit test for the fitted model in part (b). Is this a good fit?

2. (50 points) Survival of patients following admission to an adult intensive care unit (ICU)

The ICU study data set consists of a sample of 200 subjects who were part of a much larger study on survival of patients following admission to an adult intensive care unit (ICU). The major goal of this study was to develop a logistic regression model to predict the probability of survival to hospital discharge of these patients. A code sheet for the variables in this data is provided in `icu_codebase.pdf`. The primary outcome variable is vital status at hospital discharge (`sta`). Clinicians associated with the study felt that a key determinant of survival was the type of admission, `type`. The dataset can be loaded into `R` as follows:

```
load("icu.rda")
```

a. Write down the equation for the logistic regression model of `sta` on `type`. What characteristic of the outcome variable, `sta`, leads us to consider the logistic regression model as opposed to the usual linear regression model to describe the relationship between `sta` and `type`? What is the parameter of interest and what does it represent?

b. Use the `plot` function (in base R) to plot the relationship between `sta` and `type`. Interpret the plot.

c. Using a logistic regression routine of your choice, obtain the estimates of the parameters of the logistic regression model in part a. Interpret the estimate for the parameter of interest in the context of the problem and provide a 95% confidence interval for this parameter. State your assumptions.

d. Using these estimates, write down the equation for the fitted values, that is, the estimated probabilities of the response.

e. Plot the fitted values as a function of `sta` using a boxplot. Interpret the boxplot.

f. Fit a model with `type` as the response and `sta` as the determinant. Comment on the differences and similarities of this model vs. the one in part c. Would you prefer one model over the other? Explain.