

Stat 536: Assignment 5

Due: noon, Thursday, April 29

Q1.

The file "DCCT.csv" (available on the course web-site) provides a subset of the DCCT dataset as described in example 8.4, p. 331 of Lachin. The variables provided are *nevents*, *ftime*, *age* (years) and *group* ("EXPERIMENTAL", "STANDARD")

- a. Fit a poisson regression model including *group* and *ftime* as explanatory variables. Assess this fit for overdispersion using Pearson's X^2 and the score test described briefly in lecture (based on observed average of $z_i = \frac{(y_i - \hat{\mu}_i)^2 - y_i}{\hat{\mu}_i}$).
- b. Obtain a quasi-likelihood fit to the same data (this can be done in *R*, using *glm* and *family=quasipoisson*). In addition, fit a negative binomial model. Compare the fits in relationship to that provided in (a). What key differences do you note? Are there any key similarities between the quasi-likelihood and negative binomial fits? How would you interpret the apparent effect of *ftime* as an explanatory variable? Can you think of a potential clinical explanation for this effect? c. Add the variable *age* to both models. Explain the impact of adding *age* to the model on the *ftime* coefficient. It may help to consider using the *adult* variable instead of *age*.
- d. Conduct a Monte Carlo simulation to compare the power of the Pearson X^2 test and the score test as follows.
 - i. Obtain fitted values from negative binomial model including *age* and *group* as explanatory variables.
 - ii. Simulate samples of 715 with the same age and group assignments as the original data by generating response values following

- a Poisson distribution and
- negative binomial distributions, with α set to .5, 1, and $2 \times$ the estimated value from the fit to the actual data.

Q2.

a. The data set on the web as "PTCA.csv" was collected over several years by following 600 patients who had undergone angioplasty to correct blockages in cardiac blood vessels. All patients had at least two affected vessels. The columns in the spreadsheet correspond to:

age = Age in years

sex = Male/Female

ua = History of unstable angina prior to angioplasty (yes/no)

prevcabg = Previous coronary arterial bypass graft surgery (yes/no)

mostofu = Months of follow-up

death = Status at follow-up (yes/no)

Examine the relationship of survival to each of the variables above, separately. Construct a multiple regression model for predicting outcome (be sure to describe carefully the rationale for choosing your model). Interpret your findings.

b. Suppose that one observes independent counts following a Poisson distribution in the presence of censoring. Assume for censored cases that we know only that the actual value exceeds that recorded. Derive expressions for the score and observed information for the Poisson mean and simplify them to show that they may be calculated from the usual Poisson probability and cumulative distribution functions. Implement a Newton-Rafson scheme to compute the maximum likelihood estimate and associated standard error from the following sample of 20 observations (*-ed values are censored).

7	11	7	11*	10	8*	6	12	8	8
18	9	9*	11	7	8	10	11	14	4