Homework5

BIOS6643 Fall 2021

8/20/2021

By far the three most common cases for generalized linear models are for normal, binary, and count outcomes. Since you have seen so many examples of normal (standard linear models) and binary (logistic regression), these questions involve count outcomes. (Where appropriate include brief and annotated SAS or R output)

Question 1

The Cereal2.csv contains data from a nutrition study where several members of each of a number of families recorded the number of servings of breakfast they ate each week starting at baseline and continuing for 14 weeks. Some families were in an experimental program (cond=1), the others did not receive anything special (cond=0). Family members were coded $1 = Mom$, $2 = Dad$, $3 = Kid1$, $4 = Kid2$, $5 = Kid3$, $6 = Kid4$, $7 = Kid5$. Sex is coded $0 = Male$, $1 = Female$. Weight at baseline has been recoded in units of $100lbs$ (so $1.5 = 150lbs$) to avoid numerical problems in some procedures.

Note: To get agreement between GENMOD and NLMIXED you need to use a new variable Cond2 = 1 - Cond for the condition variable. This just gives a different parameterization.

You can get the data into SAS using

```
Proc import
   Datafile = 'C:\yourdir\cereal.csv'
   Out = cereal
   Dbms = CSV
   Replace;
Run;
```

Or into R using

```
cereal <- read.csv("C:/yourdir/cereal.csv")
```

In this question you will analyze only the week 1 (baseline) data, C1. There was some question as to whether the experimental and control groups were comparable at this time because some aspects of the intervention may have been done before week 1. Carry out the following analyses to compare the groups at baseline, using data only from $Kid1$ ($FamMem = 3$) - families were sampled based on this child. For each model, write the model equation, and write a sentence describing the results, understandable by dietitians.

a. Use a Poisson GzLM (i.e. Poisson regression) to estimate the association between condition and number of breakfast servings, adjusting for sex and weight.

b. Repeat **(a)** allowing for overdispersion by using quasilikelihood with the Poisson GzLM. Use the Pearson method for estimating the scale parameter. Show algebraically the relation between the QL SEs for the betas and those from the Poisson model in **(a)**.

c. Repeat **(a)** allowing for overdispersion by adding a random normal error to the linear predictor in the Poisson GzLM and using maximum likelihood estimation. I am not aware of an algebraic relation between these ML SE's and those from the models in **(a)** or **(b)**.

d. Repeat **(a)** allowing for overdispersion by using a Negative Binomial GzLM estimated with maximum likelihood. I am not aware of an algebraic relation between these ML SE's and those from the models in **(a)**, **(b)**, or **(c)**.

Question 2

As before use data for $Kid1$ only, at baseline only (C1), as in **question 1** above. You don't need to fit any new models.

a. Use the chart below to summarize the models you fit in **question 2** comparing Conditions and adjusting for Sex and Weight. In each entry except 'Intercept' and 'Other param' give the estimate of the rate ratio and a 95% CI (these are easy to calculate from the beta and SE). For 'Intercept' give the beta and its SE. For 'Other param', give the Scale parameter for QL, the SD for the normal error model, or the dispersion parameter for NB. $QL = Quasilikelihood$, $NB = Negative\ Binomial$.

b. Write a short paragraph summarizing the results of comparing conditions, e.g. which condition gives higher consumption and by how much.

c. Write a short paragraph comparing the model fits, e.g. differences between parameter estimates across models.

	"Poisson Regression"	"Poisson QL"	"Poisson + Normal error"	"NB NLMIXED"
Intercept				
Cond				
Sex				
Wt				
Other	NA			