HW2_answer

Guojing Wu | UNI: gw2383

10/25/2019

Marginal Models. For this homework you will need the R package gee.

Part A

Read the Analysis requirement in part B. For the both analysis parts (1) and (2) in part B, clearly set up the model i.e. clearly state

1. Distribution assumption for the outcome

2. The systematic component of the model

3. The canonical link function

4. The variance function

5. The correlation structure

Part B

In a clinical trial of patients suffering from epileptic seizures, patients were randomized to receive either a placebo or the drug progabide, in addition to standard therapy. A baseline count of the number of epileptic seizures in a 8-week period prior to randomization was obtained. In addition, counts of the number of epileptic seizures in each of the four successive 2-week (post baseline) treatment periods were obtained. The goal of the analysis is to make a comparison between the two treatment groups in terms of changes in the rates of epileptic seizures throughout the duration of the study. The data epilepsy.dat is available on Courseworks. The variables in the data set are

- Patient ID

- Treatment (0=Placebo, 1=Progabide)

- Age at baseline

- Baseline 8 week seizure count

- First 2 week seizure count

- Second 2 week seizure count

- Third 2 week seizure count

- Fourth 2 week seizure count

1.

(a) Consider a model for the seizure rate that includes the main effects of treatment and time (time is regarded as a ordinal variable with 5 levels) and their interaction. (library gee, command gee()).

Let y_{ij} = the number of epileptic seizures in each of the four successive 2-week treatment periods. Then we assume $y_{ij} \sim Pos(\lambda_{ij})$. The systematic component will be $\theta_{ij} = X_{ij}^T\beta$, canonical link function: $\theta_{ij} = log(\lambda_{ij})$, variance function: λ_{ij} and the correlation function as unstructured.

```
          (Intercept)              Treatment1                 TimeFirst
           3.42705076              0.02753449                -1.19091077
           TimeSecond               TimeThird                TimeFourth
          -1.31251790             -1.25392373                -1.34760922
 Treatment1:TimeFirst Treatment1:TimeSecond  Treatment1:TimeThird
          -0.11416538             -0.01153415                -0.10521964
Treatment1:TimeFourth
          -0.19862899
```

(b) What conclusions do you draw about the effect of treatment on changes in the seizure rate.

The log rate ratio of number of epileptic seizures between Treatment and Control group is 0.028 in baseline time point.

(c) Construct a variable Ptime such that Ptime $= 0$ if baseline and Ptime $= 1$ if post baseline. Repeat the above analysis using Ptime instead of the 5 categorized time variables. Compare the two models.

```
     (Intercept)         Treatment1            Ptime Treatment1:Ptime
      3.42705076         0.02753449      -1.27445834      -0.10472579
```

The log rate ratio of number of epileptic seizures between Treatment and Control group is 0.028 in baseline time point. Compare to (b), it's the same.

(d) Are there any potential outliers in the dataset. If so do they have any effect on the analysis.

(e) Summarize your results and provide relevant interpretation.

2.

Construct a new outcome variable NS at each time point such that NS $= 1$ if number of seizures is greater than 20 and NS $= 0$ If the number of seizures is less than or equal to 20.

(a) Repeat the analysis in (1) above to model this new outcome.

Let $y_{ij} =$ the indicator of whether the number of epileptic seizures > 20 or not in each of the four successive 2-week treatment periods. Then we assume $y_{ij} \sim binomial(n, p_{ij})$. The systematic component will be $\theta_{ij} = X_{ij}^T \beta$, canonical link function: $\theta_{ij} = log(\frac{p_{ij}}{1-p_{ij}})$, variance function: $p_{ij}(1 - p_{ij})$ and the correlation function as unstructured.

```
          (Intercept)              Treatment1                 TimeFirst
          -0.1431008               0.6026332                -1.9771627
           TimeSecond               TimeThird                TimeFourth
          -1.9771627              -1.9771627                -1.6486586
 Treatment1:TimeFirst Treatment1:TimeSecond  Treatment1:TimeThird
          -1.1565183              -1.1565183                -0.7159619
Treatment1:TimeFourth
          -2.2120711
```

The log odds ratio of number of epileptic seizures between Treatment and Control group is 0.603 in baseline time point.

```
     (Intercept)          Treatment1              Ptime Treatment1:Ptime
     -0.1431008          0.6026332          -1.8870696          -1.2466113
```

The log odds ratio of number of epileptic seizures between Treatment and Control group is 0.603 in baseline
time point, same as above

Appendix

```r
knitr::opts_chunk$set(echo = FALSE, message = FALSE, warning = FALSE, comment = "")
library(tidyverse)
library(gee)  # for gee()
options(knitr.table.format = "latex")
theme_set(theme_bw())
# load data
epilepsy = read.table("epilepsy.dat") %>%
  rename(ID = V1, Treatment = V2, Age = V3, Baseline = V4, First = V5,
         Second = V6, Third = V7, Fourth = V8) %>%
  gather(Baseline:Fourth, key = "Time", value = "Count") %>%
  mutate(Treatment = as.factor(Treatment),
         Time = factor(Time, levels = c("Baseline", "First", "Second", "Third", "Fourth"))) %>%
  as.tibble()

# fit GEE with AR1
fit.gee1 = gee(Count ~ Treatment * Time, data = epilepsy, family = "poisson", id = ID, corstr = "unstru
epilepsy.ptime = epilepsy %>%
  mutate(Ptime = ifelse(Time == "Baseline", 0, 1))

fit.gee2 = gee(Count ~ Treatment * Ptime, data = epilepsy.ptime, family = "poisson", id = ID, corstr = 
epilepsy.ind = epilepsy %>%
  mutate(Count = ifelse(Count > 20, 1, 0))

fit.gee3 = gee(Count ~ Treatment * Time, data = epilepsy.ind, family = "binomial", id = ID, corstr = "u
epilepsy.ind.ptime = epilepsy.ind %>%
  mutate(Ptime = ifelse(Time == "Baseline", 0, 1))

fit.gee4 = gee(Count ~ Treatment * Ptime, data = epilepsy.ind.ptime, family = "binomial", id = ID, corst
```