Biostat 653 Homework 5

David (Daiwei) Zhang

December 11, 2017

1 Solutions

1. Problem 1

   (a) The GEE is

   $$\sum_{i=1}^{N} D_i^T V_i^{-1}(Y_i - \mu_i) = 0$$

   Here $D_i = \frac{\partial \mu}{\partial \mu} = I_2$, $V_i = I_2$, and $\mu_i = \mu$, so

   $$\hat{\mu} = \frac{1}{N}\sum^{N} Y_i$$

   and

   $$\hat{\delta} = L\hat{\mu} = \frac{1}{N}\sum^{N}(Y_{i2} - Y_{i1})$$

   where $L = [-1, 1]$.

   (b) We have the empirical variance estimator

   $$Cov(\hat{\beta}) = F^{-1}GF^{-1}$$

   where

   $$F = \sum^{N} D_i^T V_i^{-1} D_i = NI_2$$
   $$G = \sum^{N} D_i^T V_i^{-1}(\hat{Y}_i - \mu_i)(\hat{Y}_i - \mu_i)^T V_i^{-1} D_i = \sum^{N}(Y_i - \mu)(Y_i - \mu)^T$$

   Then

   $$\hat{Cov}(\hat{\mu}) = \frac{1}{N^2}\sum^{N}(Y_i - \mu)(Y_i - \mu)^T$$

   and

   $$\hat{Cov}(\hat{\delta}) = \frac{1}{N^2} L[\sum^{N}(Y_i - \mu)(Y_i - \mu)^T]L^T$$

(c) The variable

$$\gamma = \beta_2 - \beta_1 = \log(\frac{\mu_2}{\mu_1})$$

is the log ratio of the expected number of new cancers.

(d) For GLM, the usual variance function is $v(\mu) = \mu$, so in our case, we can use

$$\begin{aligned} V_i =& \phi A^{1/2} R(\alpha) A^{1/2} \\ =& \phi \begin{bmatrix} \sqrt{\exp\beta_1} & 0 \\ 0 & \sqrt{\exp\beta_2} \end{bmatrix} \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} \sqrt{\exp\beta_1} & 0 \\ 0 & \sqrt{\exp\beta_2} \end{bmatrix} \\ =& \phi \begin{bmatrix} \exp\beta_1 & \alpha \exp\frac{1}{2}(\beta_1+\beta_2) \\ \alpha \exp\frac{1}{2}(\beta_1+\beta_2) & \exp\beta_2 \end{bmatrix} = V \end{aligned}$$

(e) For the GEE, we now have

$$D_i = \frac{\partial \mu}{\partial \beta} = \begin{bmatrix} \exp\beta_1 & 0 \\ 0 & \exp\beta_2 \end{bmatrix} = D$$

so we need to solve

$$0 = \sum^N D_i W (Y_i - \mu) = \sum^N DW(Y_i - \mu) = DW \sum^N (Y_i - \mu).$$

A solution of this equation is $\hat{\mu} = \frac{1}{N}\sum^N Y_i$.

(f) Finally, for the variance estimate,

$$\begin{aligned} F =& NDV^{-1}D \\ G =& DV^{-1}[\sum^N (Y_i - \mu)(Y_i - \mu)^T]V^{-1}D \end{aligned}$$

so

$$\begin{aligned} \hat{Cov}(\hat{\beta}) =& F^{-1}GF^{-1} \\ =& \frac{1}{N^2}(D^{-1}VD^{-1})(DV^{-1}[\sum^N (Y_i - \mu)(Y_i - \mu)^T]V^{-1}D)(D^{-1}VD^{-1}) \\ =& \frac{1}{N^2} D(\hat{\beta})^{-1}[\sum^N (Y_i - \exp\hat{\beta})(Y_i - \exp\hat{\beta})^T]D(\hat{\beta})^{-1} \end{aligned}$$

2. Problem 2

(a) See SAS output (proc genmod). The estimated treatment effect is -0.0777 with a p-value of 0.1485.

(b) Increase in the expected log odds of the probability of having moderate or severe symptom per month in the control group.

(c) Difference in the expected rate of increase in log odds of the probability of having moderate or severe symptom per month between the case and the control group.

(d) Since the p-value for the intersection term is $0.1485 > 0.05$, the effect of the treatment is not significance.

(e) See SAS output (proc glimmix). The estimated treatment effect is -0.1424 with a p-value of 0.0284.

(f) We have $\sigma_b^2 = 16.0349$. The magnitude of this value represents the variability of the log odds among all the individuals.

(g) Difference in the expected rate of increase in log odds of the probability of having moderate or severe symptom per month between the case and the control group, given that the individual's random intercept is fixed.

(h) Since the p-value for the intersection term is $0.1485 > 0.05$, the effect of the treatment is not significance, given that the individual's random intercept is fixed.

(i) The estimated treatment effect in the GLMM model, compared to that in the GEE model, is greater in magnitude and has a lower p-value. The GEE model treats all the inter-individual variation as measurement error, which adds noise to the analysis and reduces the effect and significance of the treatment. This issue is fixed in the GLMM model, which takes into account the inter-individual variation.

(j) We find that the number of quadrature points has virtually no effect on the estimation and p-value of the treatment effect. See the table below. (SAS output truncated due to the excessive number of pages.)

Quad Pts	Est Effect	p-val
2	-0.1303	0.0271
5	-0.1387	0.0276
10	-0.1432	0.0286
20	-0.1424	0.0284
30	-0.1424	0.0284
50	-0.1303	0.0271

2 SAS code

```
libname bs653 "~/biostat653";
data toe;
set bs653.toenail;
run;

proc genmod data=toe descending;
class id;
```

```
model y = month trt*month/d=bin;
repeated subject=id/type=exch;
run;

PROC GLIMMIX METHOD=QUAD(QPOINTS=50);
title "QP = 50";
CLASS id;
MODEL y= month trt*month /DIST=BINOMIAL LINK=LOGIT S;
RANDOM INTERCEPT / SUBJECT=id TYPE=UN;
run;

PROC GLIMMIX METHOD=QUAD(QPOINTS=2);
title "QP = 2";
CLASS id;
MODEL y= month trt*month /DIST=BINOMIAL LINK=LOGIT S;
RANDOM INTERCEPT / SUBJECT=id TYPE=UN;
run;

PROC GLIMMIX METHOD=QUAD(QPOINTS=5);
title "QP = 5";
CLASS id;
MODEL y= month trt*month /DIST=BINOMIAL LINK=LOGIT S;
RANDOM INTERCEPT / SUBJECT=id TYPE=UN;
run;

PROC GLIMMIX METHOD=QUAD(QPOINTS=10);
title "QP = 10";
CLASS id;
MODEL y= month trt*month /DIST=BINOMIAL LINK=LOGIT S;
RANDOM INTERCEPT / SUBJECT=id TYPE=UN;
run;

PROC GLIMMIX METHOD=QUAD(QPOINTS=20);
title "QP = 20";
CLASS id;
MODEL y= month trt*month /DIST=BINOMIAL LINK=LOGIT S;
RANDOM INTERCEPT / SUBJECT=id TYPE=UN;
run;

PROC GLIMMIX METHOD=QUAD(QPOINTS=30);
title "QP = 30";
CLASS id;
MODEL y= month trt*month /DIST=BINOMIAL LINK=LOGIT S;
RANDOM INTERCEPT / SUBJECT=id TYPE=UN;
run;
```