3000c-106-1

Chapter 10. Random Effects : Generalized Linear Mixed Models

Sangun Park

Yonsei University

sangun@yonsei.ac.kr

November 15, 2020

10.1 Random effects model (Conditional model)

- Example : Consider i represents the ith subject and j represents the jth treatment. Then there may be covariances between the measurements within the sample subject.

Subject	Treatment				Average
	None	Tablet	Capsule	Coated	
1	44.5	7.3	3.4	12.4	16.9
2	33.0	21.0	23.1	25.4	25.6
3	19.1	5.0	11.8	22.0	14.5
4	9.4	4.6	4.6	5.8	6.1
5	71.3	23.3	25.6	68.2	47.1
6	51.2	38.0	36.0	52.6	44.5
Average	38.1	16.5	17.4	31.1	25.8
Var	505.10	177.44	167.38	588.74	

- Two-Way model : $y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$ where ϵ_{ij}'s are independent.

10.1 Random effects model (Conditional model)

- Fixed effect vs random effect :
  1. Fixed effect model (misleading) : μ, α_i, β_j are constants to be estimated.

$$
\begin{aligned}
Cov(y_{i1}, y_{i2}) &= Cov(\mu + \alpha_i + \beta_1 + \epsilon_{i1}, \mu + \alpha_i + \beta_2 + \epsilon_{i2}) \\
&= 0
\end{aligned}
$$

  2. Random effect model : μ, β_j are constants to be estimated. α_i's are random components from $N(0, \sigma_\alpha^2)$ and σ_α^2 needs to be estimated.

$$
\begin{aligned}
Cov(y_{i1}, y_{i2}) &= Cov(\mu + \alpha_i + \beta_1 + \epsilon_{i1}, \mu + \alpha_i + \beta_2 + \epsilon_{i2}) \\
&= Cov(\alpha_i + \epsilon_{i1}, \alpha_i + \epsilon_{i2}) = \sigma_\alpha^2
\end{aligned}
$$

  We reduced the number of parameters to be estimated, but we need to consider the covariance in comparing $\bar{y}_i$'s.

- Generalized linear mixed model :
  α_i : random effect for subject i, β_j : fixed effect for treatment j, y_{ij} : observation j in subject i, x_{ij} : the corresponding explanatory variable.

$$
logitPr(y_{ij} = 1) = \mu + \alpha_i + \beta_j + \beta x_{ij}
$$

3000c-106-4

- Input Data : Question = 1 for "Pay Higher Tax", 0 for Cut living standards, Y = 1 for "Yes" , 0 for "No"

```
> dat[c(1:2,453:456,717:720,931:934,2287:2288),]
     person question y
1         1        1 1
2         1        0 1
453     227        1 1
454     227        0 1
455     228        1 1
456     228        0 0
717     359        1 1
718     359        0 0
719     360        1 0
720     360        0 1
931     466        1 0
932     466        0 1
933     467        1 0
934     467        0 0
2287   1144        1 0
2288   1144        0 0
```

- Summarized Data

Table 10.1. Opinions Relating to Environment

	Cut Living Standards		
Pay Higher Taxes	Yes	No	Total
Yes	227	132	359
No	107	678	785
Total	334	810	1144

- Ignoring the subject effect (misleading)
  - Odds ratio under independence : $(359/785)/(334/810) = 1.11$
  - Marginal model under independence :

$$\begin{aligned} logitP(Y_1 = 1) &= \mu + \beta \\ logitP(Y_2 = 1) &= \mu \end{aligned}$$

- Considering the subject effect (fixed effect model) : We have too many parameters to be estimated.
- Considering the subject effect (random effect model) : OR can be obtained from the following conditional model as exp(0.21) which is equal to 132/107.

$$\begin{aligned} logitP(Y_{i1} = 1) &= \mu + \alpha_i + \beta \\ logitP(Y_{i2} = 1) &= \mu + \alpha_i \end{aligned}$$

where $\alpha_i \sim N(0, \sigma_\alpha^2)$.

R Code : Ignoring the subject effect

3000c-106-6

```
> dat=read.table("opinions.dat", header=T)
> res=glm(y~question, family=binomial,data=dat)
> summary(res)

Call:
glm(formula = y ~ question, family = binomial, data = dat)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-0.868  -0.868  -0.831   1.522   1.569

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.886      0.065  -13.62   <2e-16 ***
question       0.103      0.091    1.14     0.26
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2806.4  on 2287  degrees of freedom
Residual deviance: 2805.1  on 2286  degrees of freedom
AIC: 2809

Number of Fisher Scoring iterations: 4

> exp(res$coefficients[2])
question
    1.11
```

R Code : Considering the subject effect (random effect model)


```
> library(lme4)
> res=glmer(y~(1|person) + question, family=binomial, nAGQ=50, data=dat)
> summary(res)
Generalized linear mixed model fit by maximum likelihood (Adaptive Gauss-Hermite Quadrature, nAGQ = 50) ['glmerMod']
 Family: binomial  ( logit )
Formula: y ~ (1 | person) + question
   Data: dat

     AIC      BIC   logLik deviance df.resid
    2527     2544    -1260     2521     2285

Scaled residuals:
   Min     1Q Median     3Q    Max
-0.887 -0.269 -0.242  0.465  1.252

Random effects:
 Groups Name        Variance Std.Dev.
 person (Intercept) 8.14     2.85
Number of obs: 2288, groups:  person, 1144

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.834      0.162  -11.30   <2e-16 ***
question       0.210      0.130    1.61     0.11
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
```

Example :

3000c-106-8

- Example (Agresti, Table 9.1, 3rd Ed) : Under three situations, 1850 subjects choose "Y" or "N" to the legalized abortion; 1 for Yes, 0 for N.; 1 for Female, 0 for Male

```
> head(dat,6)
  person gender situation response
1      1      1         1        1
2      1      1         2        1
3      1 matrix(data = NA, nrow = 1, ncol = 1, byro
4      2      1         1        1
5      2      1         2        1
6      2      1         3        1
```

- Summarize data :

Gender	YYY	YYN	NYY	NYN	YNY	YNN	NNY	NNN
Male	342	26	6	21	11	32	19	356
Female	440	25	14	18	14	47	22	457

- Marginal probabilities : P(Y)

Gender	1st Situation	2nd Situation	3rd Situation
Female	0.5072	0.4793	0.4725
Male	0.5055	0.4859	0.4649

Example :

- Ignoring the subject effect : Misleading
- Two right approaches
  1. Marginal model : GEE
  2. Conditional Model : mixed model

$$logitP(Y_{it} = 1) = \alpha_i + \beta_1 Gender + \beta_2 Situation1 + \beta_3 Situation2$$

where $\alpha_i \sim N(0, \sigma_\alpha^2)$.

R Code and Result : Analysis under independence


```
> dat=read.table("abortion.dat", header=T)
> head(dat,3)
  person gender situation response
1      1      1         1        1
2      1      1         2        1
3      1      1         3        1
> dat$situation=factor(dat$situation,levels=c(3,1,2))
> res=glm(response~gender+situation, family=binomial,data=dat)
> summary(res)

Call:
glm(formula = response ~ gender + situation, family = binomial,
    data = dat)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.189  -1.148  -1.125   1.207   1.231

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.125408   0.055601  -2.255   0.0241 *
gender       0.003582   0.054138   0.066   0.9472
situation1   0.149347   0.065825   2.269   0.0233 *
situation2   0.052018   0.065843   0.790   0.4295
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7689.5  on 5549  degrees of freedom
Residual deviance: 7684.2  on 5546  degrees of freedom
AIC: 7692.2
```

R Code and Result : GEE(exchangeable)

3000c-106-11

```
> res=gee(response~gender+situation,id=person,family=binomial,corstr="exchangeable",data=dat)
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
running glm to get initial regression estimate
 (Intercept)       gender   situation1   situation2
-0.125407576  0.003582051  0.149347113  0.052017989
> summary(res)

 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
 Link:                      Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     Exchangeable

Call:
gee(formula = response ~ gender + situation, id = person, data = dat,
    family = binomial, corstr = "exchangeable")

Summary of Residuals:
       Min         1Q     Median         3Q        Max
-0.5068644 -0.4825396 -0.4687095  0.5174604  0.5312905


Coefficients:
                Estimate Naive S.E.     Naive z Robust S.E.    Robust z
(Intercept) -0.125325730 0.06782579 -1.84775925  0.06758212 -1.85442135
gender       0.003437873 0.08790630  0.03910838  0.08784072  0.03913758
situation1   0.149347107 0.02814374  5.30658404  0.02973865  5.02198729
situation2   0.052017986 0.02815145  1.84779075  0.02704703  1.92324179

Estimated Scale Parameter:  1.000721
Number of Iterations:  2

Working Correlation
          [,1]      [,2]      [,3]
[1,] 1.0000000 0.8173308 0.8173308
[2,] 0.8173308 1.0000000 0.8173308
[3,] 0.8173308 0.8173308 1.0000000
```

R Code and Result : Mixed model

3000c-106-12

```
> library(lme4)
> res=glmer(response~(1|person) + gender + situation, family=binomial, nAGQ=100, data=dat)
> summary(res)
Generalized linear mixed model fit by maximum likelihood (Adaptive Gauss-Hermite Quadrature, nAGQ = 100) ['glmerMod']
 Family: binomial  ( logit )
Formula: response ~ (1 | person) + gender + situation
   Data: dat

     AIC      BIC   logLik deviance df.resid
  4588.5   4621.6  -2289.3   4578.5     5545

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.7810 -0.1223 -0.1055  0.1396  1.7149

Random effects:
 Groups Name        Variance Std.Dev.
 person (Intercept) 76.49    8.746
Number of obs: 5550, groups:  person, 1850

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.61936    0.37847  -1.636    0.102
gender       0.01261    0.49001   0.026    0.979
situation1   0.83478    0.16008   5.215 1.84e-07 ***
situation2   0.29245    0.15670   1.866    0.062 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
           (Intr) gender sittn1
gender     -0.725
situation1 -0.218  0.000
situation2 -0.211  0.000  0.508
```