BST5220 Multilevel and Longitudinal Analysis Homework 4

Miao Cai [miao.cai@slu.edu](mailto:miao.cai@slu.edu)

3/4/2019

59 epileptic patients were randomized to treatment with Progabide (analog and prodrug of gamma-aminobutyric acid), or to placebo in addition to standard chemotherapy, so that 28 patients on placebo and 31 patients on progabide. The numbers of seizures were collected on each patient at baseline, 2, 4, 6 and 8 weeks after the baseline. Data is given on the blackboard.

Questions:

1. Convert these wide-form data to long-form. Refer to lecture 1, slide33.
2. Graphically examine the association between the number of seizures and week. What do the graphs indicate about the pattern of change of the numbers of seizures over time (positive or negative)?
3. Use the mixed model building strategies developed in lecture 7 to determine the best model. Report the G and V matrices at each step. For the final model, report the within-subject correlation and interpret all significant effects.
4. Fit the GEE model with an appropriate working correlation matrix (EXCH, UN, AR(1), or TOEP).

**Answers:**

1. Wide format to long format:

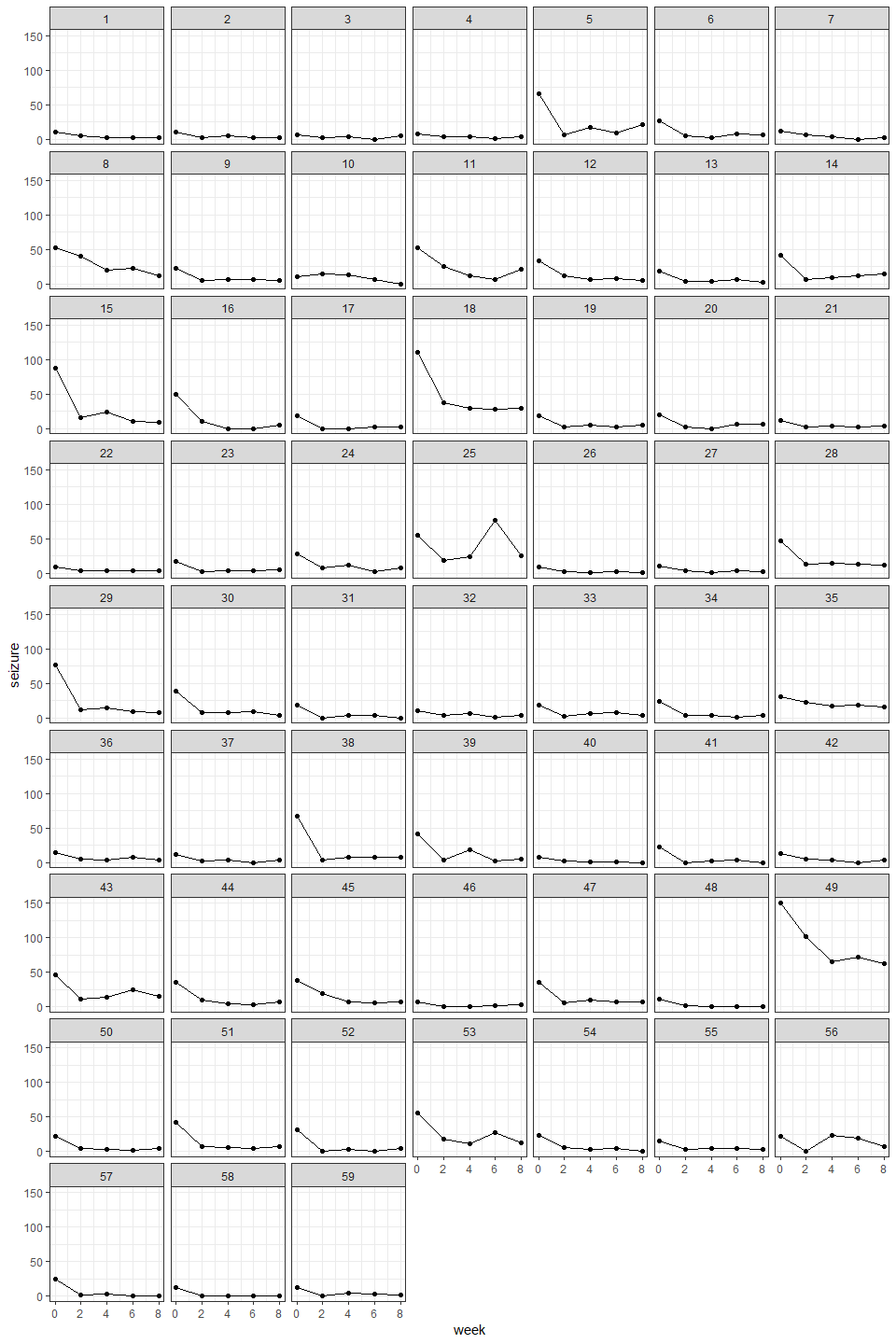
pacman::p\_load(tidyverse, tidyr)  
dat = haven::read\_sas('HW4/one.sas7bdat')  
  
datlong = dat %>%   
 gather(key = 'week', value = 'seizure', -ID, -Treatment, -Age) %>%   
 mutate(week = as.integer(substr(week, 8, 8))\*2 - 2)  
  
knitr::kable(head(datlong),   
 caption = 'First 5 observations of the converted long format data')

First 5 observations of the converted long format data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Treatment | Age | week | seizure |
| 1 | 0 | 31 | 0 | 11 |
| 2 | 0 | 30 | 0 | 11 |
| 3 | 0 | 25 | 0 | 6 |
| 4 | 0 | 36 | 0 | 8 |
| 5 | 0 | 22 | 0 | 66 |
| 6 | 0 | 29 | 0 | 27 |

1. Graphics on the association between the number of seizures and week

datlong %>%   
 ggplot(aes(week, seizure)) + geom\_point() + geom\_line() +   
 facet\_wrap(.~ID, ncol = 7) + theme\_bw()



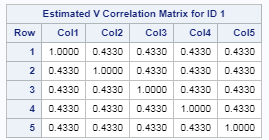
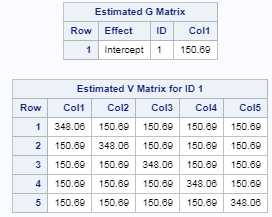
1. Hierarchical model building

**Model 1: the intercept only model**

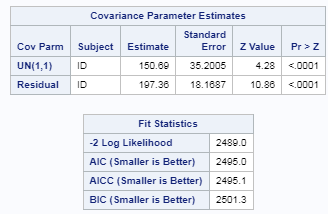
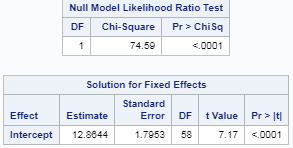
Level 1 : seizureij = b0j + eij (1)

Level 2 (intercept): b0j = g00 + u0j (2)

G matrix and V matrix:



Fit statistics and parameter estimates:

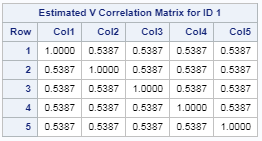
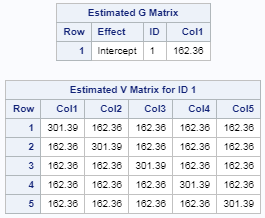
Since the random intercepst and between-subject variance are statistically siginificant, the random intercept model is good as a starting point.

**Model 2: add level 1 variables:**

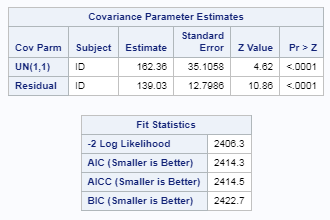
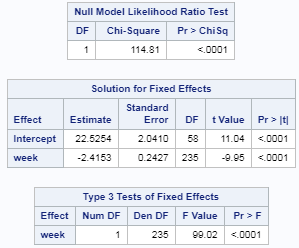
Level 1 : seizureij = b0j + b1weekij + eij (1)

Level 2 (intercept): b0j = g00 + u0j (2)

G matrix and V matrix:



Fit statistics and parameter estimates:

The Chi-square test shows that P-value is less than 0.001, which indicates that this model is significantly better than intercept-only model.

**Model 3: adding random coefficients of the time variable**

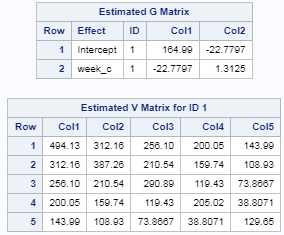
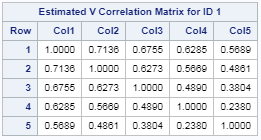
In this model, I added the random slopes for centered week (week\_c). Centered week is computed as week – mean(week). The model in matrix form is:

Level 1 : seizureij = b0j + b1j weekij + eij (1)

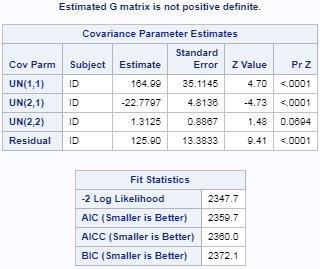
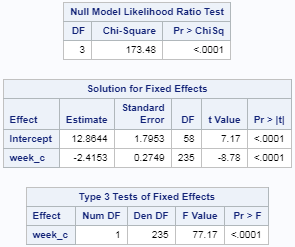
Level 2 (intercept): b0j = g00 + u0j (2)

B1j = g10 + u1j

G matrix and V matrix:

Fit statistics and parameter estimates:

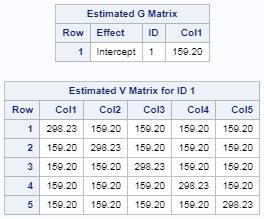
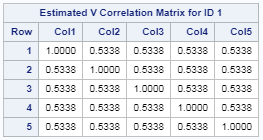
Since the estimated G matrix is not positive definite and level 2 random slope variance UN(2, 2) is not significant at the level of 0.05, this model is no better than Model 2. Therefore, we did NOT add random slopes on week.

**Model 4: adding level 2 variables**

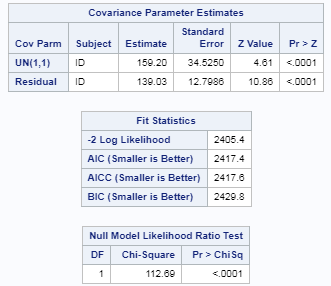
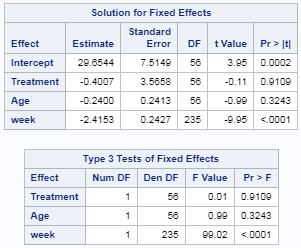
Level 1 : seizureij = b0j + b1j weekij + b2j Ageij + b3j Treatmentij + eij (1)

Level 2 (intercept): b0j = g00 + u0j (2)

G matrix and V matrix:

Fit statistics and parameter estimates:

Although these two patient level variables are not significant and deviance test showed no significance (P-value = 0.638), we still keep them since treatment is the variable of interest and age is an important demographic variable. This model turns out to be the optimal model.

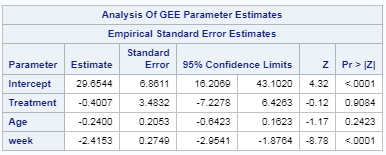
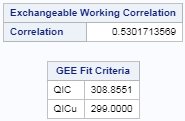
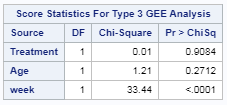
**Model 5: cross-level interaction**

Since there are no level-2 random slopes in the model, I did not test cross level interaction in the model building process.

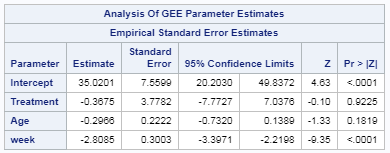
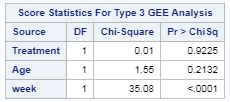
The optimal model is **model 4**, which suggests that the treatment did NOT produce significant effect on patient’s seizure.

1. The GEE model

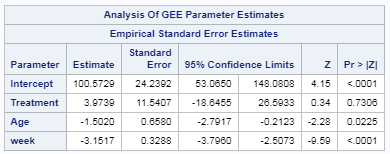
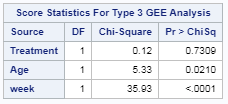
**Model 1: Exchangeable correlation matrix**

**Model 2: using AR(1) correlation structure**

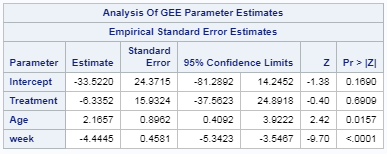
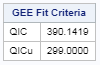
  

**Model 3: using TOEP correlation structure**

**Model 4: using unstructured correlation structure**

When using SAS to fit the model, it suggested that iteration limit exceeded. I suspect the algorithm did not converge.

Based on QIC statistics, the model using the exchangeable correlation matrix produces the minimum value. Therefore, model 1 is the optimal model, which also suggests that the treatment did NOT produce significant effect on patients.

SAS code:

**proc** **import** datafile = '/folders/myfolders/HW4/hw4long.csv' replace

out = HW4.hw4

dbms = CSV;

**run**;

/\* Problem 3 \*/

\* model 1 random intercept model;

**proc** **mixed** data=HW4.hw4 noclprint covtest noitprint method=ml;

class id;

model seizure = /solution ddfm=bw ;

random intercept /sub=id type=UN g v vcorr;

repeated / subject = id r;

**run**;

\* model 2 add level 1 variable;

**proc** **mixed** data=HW4.hw4 noclprint covtest noitprint method=ml;

class id;

model seizure = week/solution ddfm=bw ;

random intercept /sub=id type=UN g v vcorr;

repeated / subject = id r;

**run**;

**DATA** pvalue;

DF = **1**; CHISQ = **2489.0** - **2406.3**;

PVALUE = **1** - PROBCHI(CHISQ, DF);

**RUN**;

\* model 3 adding random coefficients of the time variable;

**proc** **mixed** data=HW4.hw4 noclprint covtest noitprint method=ml;

class id;

model seizure = week/solution ddfm=bw ;

random intercept week/sub=id type=UN g v vcorr;

repeated / subject = id r;

**run**;

**DATA** pvalue;

DF = **1**; CHISQ = **2406.3** - **2347.7**;

PVALUE = **1** - PROBCHI(CHISQ, DF);

**RUN**;

/\* center week and rerun the model\*/

**proc** **sql**;

create table HW4.weekC as

select \*, week - mean(week) as week\_c from HW4.hw4;

**quit**;

**proc** **mixed** data=HW4.weekC noclprint covtest noitprint method=ml;

class id;

model seizure = week\_c/solution ddfm=bw ;

random intercept week\_c/sub=id type=UN g v vcorr;

repeated / subject = id r;

**run**;

\* model 4: adding level 2 variables;

**proc** **mixed** data=HW4.hw4 noclprint covtest noitprint method=ml;

class id;

model seizure = Treatment Age week/solution ddfm=bw ;

random intercept /sub=id type=UN g v vcorr;

repeated / subject = id r;

**run**;

**DATA** pvalue;

DF = **2**; CHISQ = **2406.3** - **2405.4**;

PVALUE = **1** - PROBCHI(CHISQ, DF);

**RUN**;

/\* Problem 4 \*/

\* Model 1 Exchangeable correlation matrix ;

**proc** **genmod** data = HW4.hw4;

class id;

model seizure = Treatment Age week /type3 dist=normal;

repeated subject = id/type=exch corrw ;

**run**;

\* Model 2: Using the ar(1) correlation structure;

**proc** **genmod** data = HW4.hw4;

class id;

model seizure = Treatment Age week /type3 dist=normal;

repeated subject = id/type=ar(**1**) corrw ;

**run**;

\* Model 3: Using the toep correlation structure;

**proc** **genmod** data = HW4.hw4;

class id;

model seizure = Treatment Age week /type3 dist=normal;

repeated subject = id/type=toep corrw ;

**run**;

\* Model 4: Using the unstructured correlation structure;

**proc** **genmod** data = HW4.hw4;

class id;

model seizure = Treatment Age week /type3 dist=normal;

repeated subject = id/type=un corrw ;

**run**;