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:

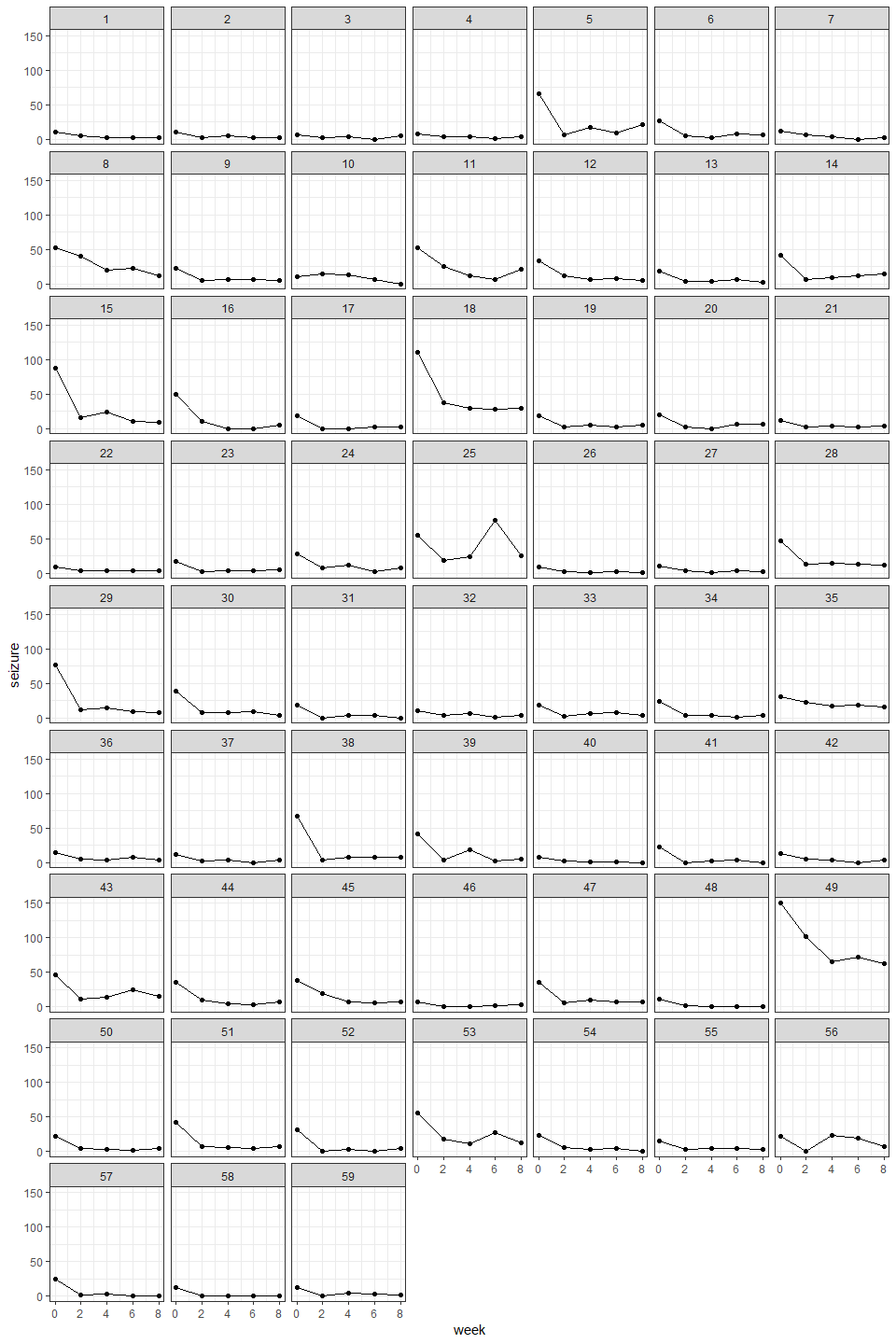
pacman::p\_load(tidyverse, tidyr)  
dat = haven::read\_sas('HW4/one.sas7bdat')  
  
datlong = dat %>%   
 gather(key = 'week', value = 'seizure', -ID, -Treatment, -Age) %>%   
 mutate(week = as.integer(substr(week, 8, 8))\*2 - 2)  
  
knitr::kable(head(datlong),   
 caption = 'First 5 observations of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

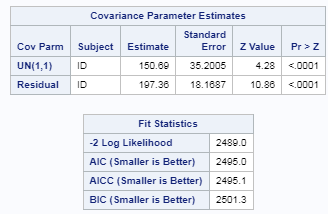
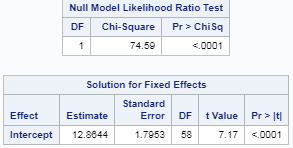


readr::write\_csv(datlong, 'HW4/hw4long.csv')

1. **Model 1: the intercept only model**

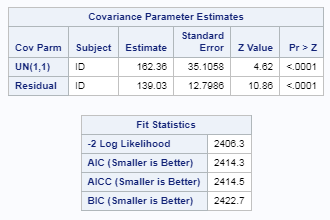
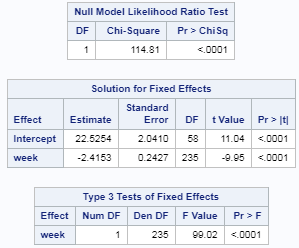
Level 1 : seizureij = b0j + eij (1)

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

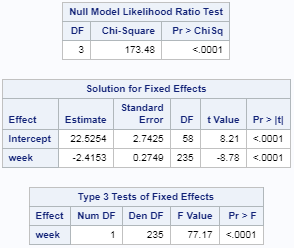
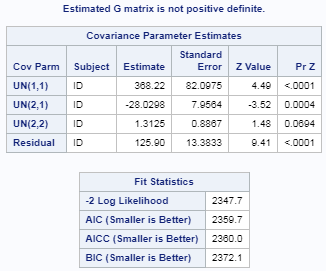
Since the random intercepst and between-subject variance are statistically siginificant, the random intercept model is good as a starting point.

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

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

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.

Model 4: adding level 2 variables

