JiayiShi_js6177_p8158hw5

Jiayi Shi

2023-03-07

Problem 1

1. colnames(data) = c('y1','y2','y3')

data = data %>% mutate(ID = row_number())

data_long = gather(data, index, y, y1:y3, factor_key = T) %>% arrange(ID) %>% mutate(time = case_when(index=='y1'~ 0, index=='y2'~ 1, index=='y3'~ 2))

set.seed(12345)

gmm1 <- hlme(y ~ time, subject = 'ID', random= ~ 1 + time, ng = 1, data=data_long)

gmm4 <- hlme(y ~ time, subject = 'ID', random= ~ 1 + time, ng = 4, data=data_long, mixture=~time, B=random(gmm1))

2. The five-class model was unparsimonious and unviable because it split one class into two parallel classes, creating a very small class (1.6%), and failed to converge when covariates were included in the model.

3. The slope for the low-stable group was 1.64 and significant due to the group's large size (83.1%) and small standard error (0.14); the slope for the high-stable group was -5.07 and non-significant due to the group's small size (2.2%) and large standard error (5.5).

4. For multiple deployers in the moderate-improving class, the adjusted odds of screening positive for heavy drinking is 2.03 times that of screening negative for heavy drinking, with 95% CI: (1.41, 2.94).

Problem 2

1. Fit a linear growth curve model with a random intercept and slope.

```
data = read.csv("data/hamd.csv", header = F)
colnames(data) = c('id','baseline','week1','week2','week3','week4','week6')

data_long = data %>%
  pivot_longer(
  baseline:week6,
  values_to = "HamD",
  names_to = "time"
) %>%
  mutate(
    time = case_when(time=='baseline'~ 0,
```

```
                          time=='week1'~ 1,
                          time=='week2'~ 2,
                          time=='week3'~ 3,
                          time=='week4'~ 4,
                          time=='week6'~ 6
                          )) %>%
   mutate_if(is.character, as.numeric)
```

```
## Warning: There was 1 warning in `mutate()`.
## ℹ In argument: `HamD = .Primitive("as.double")(HamD)`.
## Caused by warning:
## ! NAs introduced by coercion
```

```
library(nlme)
```

```
##
## Attaching package: 'nlme'
```

```
## The following object is masked from 'package:dplyr':
##
##     collapse
```

```
gmm = lme(HamD ~ time, random =~ time|id, data=data_long, method="ML", na.act
ion = na.exclude)
summary(gmm)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: data_long
##        AIC      BIC    logLik
##   27164.09 27202.61 -13576.04
##
## Random effects:
##  Formula: ~time | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 3.2724948 (Intr)
## time        0.8386899 0.167
## Residual    3.8677035
##
## Fixed effects:  HamD ~ time
##                 Value  Std.Error   DF  t-value p-value
## (Intercept) 19.394762 0.15175499 3758 127.8031       0
## time        -1.807012 0.04237932 3758 -42.6390       0
##  Correlation:
##      (Intr)
## time -0.265
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.44892190 -0.57116209  0.01962597  0.56370370  3.58758956
##
```

```
## Number of Observations: 4537
## Number of Groups: 778
```

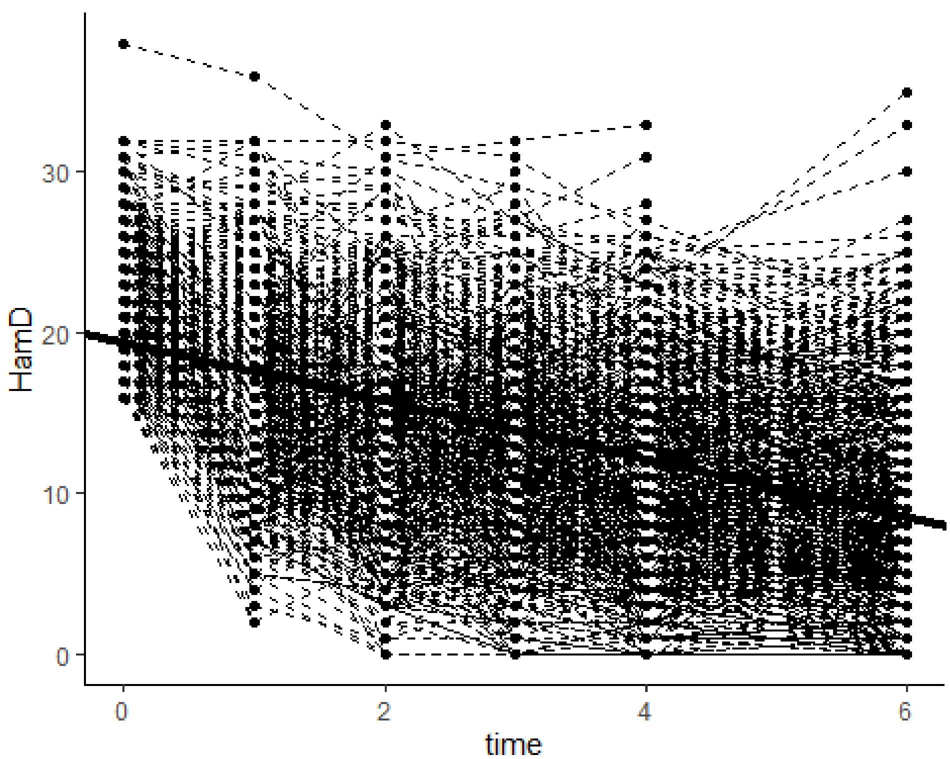
The overall estimated intercept is 19.39476 and slope is -1.80701. They are both statistically significant with p-value = 0.

```
data_long %>%
  ggplot(aes(x = time, y = HamD, group = id)) +
  geom_line(linetype = "dashed") +
  geom_point()+
  geom_abline(slope = -1.80701, intercept = 19.39476, size = 1.5) +
  theme_classic()

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.

## Warning: Removed 79 rows containing missing values (`geom_line()`).

## Warning: Removed 131 rows containing missing values (`geom_point()`).
```



2. 95% CI of slope is (-1.890082, -1.723941).

```
intervals(gmm)

## Approximate 95% confidence intervals
##
##  Fixed effects:
##                 lower      est.     upper
## (Intercept) 19.097298 19.394762 19.692227
```

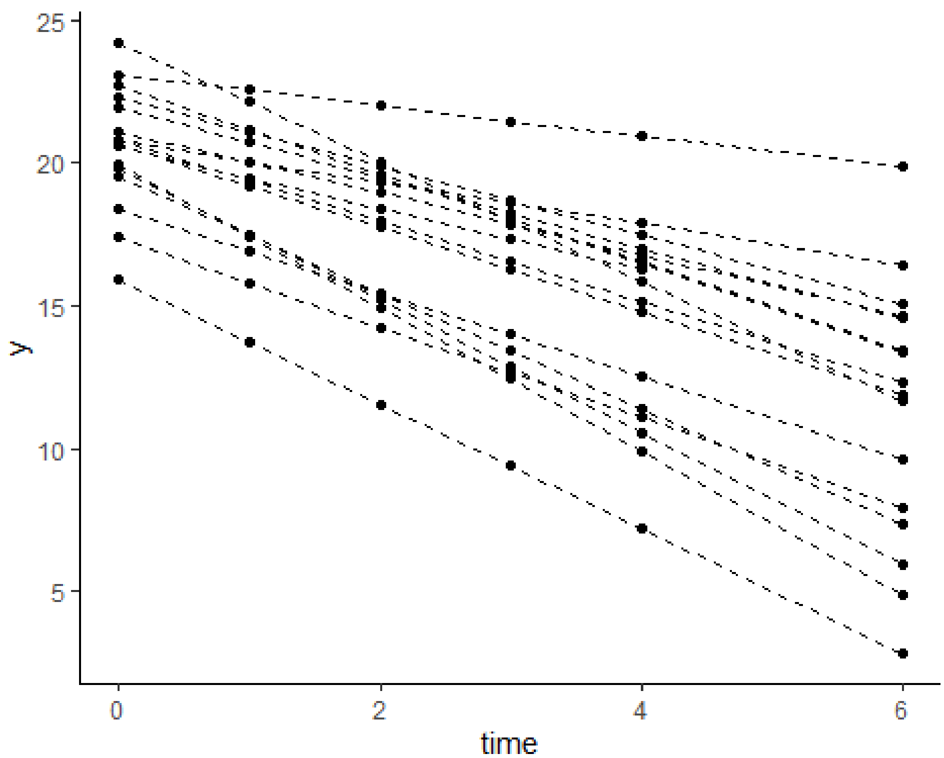
```
## time        -1.890082 -1.807012 -1.723941
##
##  Random Effects:
##   Level: id
##                            lower      est.     upper
## sd((Intercept))       3.0073485 3.2724948 3.5610181
## sd(time)              0.7570482 0.8386899 0.9291361
## cor((Intercept),time) 0.0188286 0.1669784 0.3079533
##
##  Within-group standard error:
##    lower     est.    upper
## 3.771185 3.867703 3.966693
```

A plot of the fitted Ham-D scores for 20 individuals:

```
data_long %>% filter(id<1021) %>%
  mutate(y = fitted(gmm)[1:102]) %>%
  ggplot(aes(x = time, y = y, group = id)) +
  geom_line(linetype = "dashed") +
  geom_point()+
  theme_classic()

## Warning: Removed 1 row containing missing values (`geom_line()`).

## Warning: Removed 1 rows containing missing values (`geom_point()`).
```



3. Fit a linear growth curve mixture model with K=2:

```
set.seed(12345)
gmm1 <- hlme(HamD ~ time, subject = 'id', random= ~ 1 + time, ng = 1, data=da
ta_long)

gmm2 = hlme(HamD ~ time, subject = 'id', random= ~ 1 + time, ng = 2, data=dat
a_long, mixture=~time, B=random(gmm1))

summary(gmm2)

## Heterogenous linear mixed model 
##      fitted by maximum likelihood method 
##  
## hlme(fixed = HamD ~ time, mixture = ~time, random = ~1 + time, 
##     subject = "id", ng = 2, data = data_long)
##  
## Statistical Model: 
##      Dataset: data_long 
##      Number of subjects: 778 
##      Number of observations: 4537 
##      Number of observations deleted: 131 
##      Number of latent classes: 2 
##      Number of parameters: 9  
##  
## Iteration process: 
##      Convergence criteria satisfied 
##      Number of iterations:  178 
##      Convergence criteria: parameters= 3.9e-06 
##                          : likelihood= 3.2e-05 
##                          : second derivatives= 2.8e-11 
##  
## Goodness-of-fit statistics: 
##      maximum log-likelihood: -13561.96  
##      AIC: 27141.92  
##      BIC: 27183.83  
##  
##  
## Maximum Likelihood Estimates: 
##  
## Fixed effects in the class-membership model:
## (the class of reference is the last class) 
## 
##                      coef      Se    Wald p-value
## intercept class1 -0.48836 0.16103  -3.033 0.00242
## 
## Fixed effects in the longitudinal model:
## 
##                      coef      Se    Wald p-value
## intercept class1 20.62729 0.42029  49.079 0.00000
## intercept class2 18.64019 0.25783  72.297 0.00000
## time class1      -0.89849 0.08381 -10.721 0.00000
```

```
## time class2      -2.36570 0.07969 -29.687 0.00000
##
##
## Variance-covariance matrix of the random-effects:
##           intercept    time
## intercept   9.82949
## time       -0.23870 0.18467
##
##                               coef      Se
## Residual standard error:  3.86862 0.05004
```

Fit a linear growth curve mixture model with K=3:

```
set.seed(12345)

gmm3 = hlme(HamD ~ time, subject = 'id', random= ~ 1 + time, ng = 3, data=dat
a_long, mixture=~time, B=random(gmm1))
summary(gmm3)
```

```
## Heterogenous linear mixed model
##      fitted by maximum likelihood method
##
## hlme(fixed = HamD ~ time, mixture = ~time, random = ~1 + time,
##     subject = "id", ng = 3, data = data_long)
##
## Statistical Model:
##      Dataset: data_long
##      Number of subjects: 778
##      Number of observations: 4537
##      Number of observations deleted: 131
##      Number of latent classes: 3
##      Number of parameters: 12
##
## Iteration process:
##      Convergence criteria satisfied
##      Number of iterations:  110
##      Convergence criteria: parameters= 1e-10
##                          : likelihood= 8.2e-07
##                          : second derivatives= 6.8e-05
##
## Goodness-of-fit statistics:
##      maximum log-likelihood: -13556.9
##      AIC: 27137.81
##      BIC: 27193.69
##
##
## Maximum Likelihood Estimates:
##
## Fixed effects in the class-membership model:
## (the class of reference is the last class)
```

```
##
##                        coef      Se    Wald p-value
## intercept class1  1.50356 0.60045   2.504 0.01228
## intercept class2  1.16778 0.53814   2.170 0.03000
##
## Fixed effects in the longitudinal model:
##
##                        coef      Se    Wald p-value
## intercept class1 17.92560 0.50250  35.673 0.00000
## intercept class2 20.28289 0.36910  54.952 0.00000
## intercept class3 23.12482 1.18492  19.516 0.00000
## time class1      -2.17671 0.12189 -17.858 0.00000
## time class2      -0.82945 0.08049 -10.305 0.00000
## time class3      -3.26922 0.25495 -12.823 0.00000
##
##
## Variance-covariance matrix of the random-effects:
##           intercept time
## intercept   7.74323
## time        0.47968 0.03
##
##                                coef      Se
## Residual standard error:  3.86543 0.04917
```

```
# K=2
tibble(
  class = c(1,2),
  proportion = c(summarytable(gmm2, which = "%class")[1],summarytable(gmm2, w
hich = "%class")[2]),
  intercept = c(coef(gmm2)[2:3]),
  slope = c(coef(gmm2)[4:5])
) %>% knitr::kable()
```

class	proportion	intercept	slope
1	37.01799	20.62729	-0.8984882
2	62.98201	18.64019	-2.3657015

```
# K=3
tibble(
  class = c(1,2,3),
  proportion = c(summarytable(gmm3, which = "%class")[1],summarytable(gmm3, w
hich = "%class")[2],summarytable(gmm3, which = "%class")[3]),
  intercept = c(coef(gmm3)[3:5]),
  slope = c(coef(gmm3)[6:8])
) %>% knitr::kable()
```

class	proportion	intercept	slope
1	55.912596	17.92560	-2.1767116
2	36.889460	20.28289	-0.8294487
3	7.197943	23.12482	-3.2692165

```
summarytable(gmm2)

##      G    loglik npm      BIC  %class1  %class2
## gmm2 2 -13561.96   9 27183.83 37.01799 62.98201

summarytable(gmm3)

##      G   loglik npm      BIC %class1  %class2  %class3
## gmm3 3 -13556.9  12 27193.69 55.9126 36.88946 7.197943
```

2-class model: Positive intercepts and negative slopes for both classes. 3-class model: split one class into two classes, creating a relatively small class (7.2%). Positive intercepts and negative slopes for all classes.

Model “K=2” is better than “K=3” with lower BIC and fewer parameters (parsimonious).