

# Homework 3

*Xinyi Lin*

*11/25/2019*

```
library(lme4)
library(nlme)
library(gee)
library(tidyverse)
```

## Question 1

```
toenail <- read.table("./toenail.dat", quote="\")
colnames(toenail) = c("Subject_ID", "Response", "Treatment", "Month", "Visit")
```

## Problem 1

Let  $y_{ij}$  be the onycholysis outcome for subject  $i = 1, 2, \dots, 294$  at visit  $j = 1, 2, \dots, 7$ .  $y_{ij}|b_i$  follows binomial distribution.

Model:  $\text{logit}(E[y_{ij}|b_i]) = b_{i0} + \beta_0 + \beta_1 * \text{time}_{ij} + \beta_2 \text{treatment}_{ij} + \beta_3 \text{treatment}_{ij} * \text{time}_{ij}$  where  $b_i \sim \text{MVN}(0, G)$ .

```
glm.model = glmer(Response ~ as.factor(Treatment) + Month + Month*as.factor(Treatment) + (1|Subject_ID)
summary(glm.model)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##   Family: binomial ( logit )
## Formula:
## Response ~ as.factor(Treatment) + Month + Month * as.factor(Treatment) +
##   (1 | Subject_ID)
##   Data: toenail
##
##           AIC          BIC    logLik deviance df.resid
##    1265.6     1293.4    -627.8   1255.6     1903
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.290 -0.149 -0.071 -0.006  47.215
##
## Random effects:
##   Groups      Name      Variance Std.Dev.
##   Subject_ID (Intercept) 20.76     4.557
## Number of obs: 1908, groups: Subject_ID, 294
##
## Fixed effects:
##
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.50986    0.76388  -3.286  0.00102 **
```

```
## as.factor(Treatment)1      -0.30484    0.68707   -0.444    0.65727
## Month                      -0.39973    0.04679   -8.543    < 2e-16 ***
## as.factor(Treatment)1:Month -0.13714    0.06949   -1.973    0.04846 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) as.(T)1 Month
## as.fctr(T)1 -0.374
## Month       0.130  0.220
## as.fc(T)1:M  0.181 -0.265 -0.541
```

## Problem 2

Interpretation:

Fixed effect:

$\beta_0$  : Log odds of the onycholysis outcome for a “typical” or “average” individual in treatment B group when month equals to 0 is -2.51.

$\beta_1$  : Log odds ratio of the onycholysis outcome for one unit increase in month in a given individual in treatment B group is -0.30.

$\beta_2$  : Log odds ratio of the onycholysis outcome for two individuals who have similar underlying propensity in treatment A group versus treatment B group is -0.40 when month equals to 0.

$\beta_3$  : Log odds ratio of the onycholysis outcome for two individuals who have similar underlying propensity in treatment A group versus treatment B group with one unit increase of month is -0.14.

Random effect:

$b_i$  : patient-to-patient difference. It indicates overall, difference of response level between given individual and population average level. The difference among patients have variance of 20.76.

## Problem 3

According to the model, we can find that the coefficient of interaction term of month and treatment is significant and is negative. This means over time, the log odds of the onycholysis outcome in treatment A group is increasingly smaller than that in treatment B groups, which means the effect of treatment on changes in the severity of onycholysis is increase over time.

## Problem 4

When interpreting the GEE model, we are interpreting in population level while we are interpreting in individual level when interpreting the mixed effect model.

```
gee.model <- gee(Response ~ as.factor(Treatment) + Month + Month*as.factor(Treatment), id = Subject_ID,
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27

## running glm to get initial regression estimate
```

```
##              (Intercept)          as.factor(Treatment)1
##          -0.5566272629          -0.0005816613
##              Month as.factor(Treatment)1:Month
##          -0.1703077869          -0.0672216211
```

```
summary(gee.model)
```

```
##
## 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 ~ as.factor(Treatment) + Month + Month *
##      as.factor(Treatment), id = Subject_ID, data = toenail, family = binomial,
##      corstr = "exchangeable")
##
## Summary of Residuals:
##      Min      1Q      Median      3Q      Max
## -0.36016219 -0.25999828 -0.11218173 -0.02710866  0.97982191
##
##
## Coefficients:
##              Estimate Naive S.E.    Naive z
## (Intercept)      -0.581850825  0.14027487 -4.14793340
## as.factor(Treatment)1      0.007190545  0.19493800  0.03688632
## Month            -0.171274122  0.02103731 -8.14144495
## as.factor(Treatment)1:Month -0.077723657  0.03571174 -2.17641758
##              Robust S.E.    Robust z
## (Intercept)      0.17204948 -3.38188070
## as.factor(Treatment)1      0.25945870  0.02771364
## Month            0.02999742 -5.70962885
## as.factor(Treatment)1:Month 0.05410892 -1.43642956
##
## Estimated Scale Parameter:  1.090085
## Number of Iterations:  4
##
## Working Correlation
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## [1,] 1.000000 0.421203 0.421203 0.421203 0.421203 0.421203 0.421203
## [2,] 0.421203 1.000000 0.421203 0.421203 0.421203 0.421203 0.421203
## [3,] 0.421203 0.421203 1.000000 0.421203 0.421203 0.421203 0.421203
## [4,] 0.421203 0.421203 0.421203 1.000000 0.421203 0.421203 0.421203
## [5,] 0.421203 0.421203 0.421203 0.421203 1.000000 0.421203 0.421203
## [6,] 0.421203 0.421203 0.421203 0.421203 0.421203 1.000000 0.421203
## [7,] 0.421203 0.421203 0.421203 0.421203 0.421203 0.421203 1.000000
```

$\beta_0$  : On average, log odds of the onycholysis outcome in treatment B group when month equals to 0 is -0.58.

$\beta_1$  : On average, log odds ratio of the onycholysis outcome for one unit increase in month for treatment B group is 0.007.

$\beta_2$  : On average, log odds ratio of the onycholysis outcome among treatment A group versus treatment B group is -0.171 when month equals to 0.

$\beta_3$  : On average, log odds ratio of the onycholysis outcome among treatment A group versus treatment B group with one unit increase of month is -0.078.

## Question 2

```
skin = read.table("./skin2.txt", quote="\"", comment.char="")
colnames(skin) = c("id", "center", "age", "skin", "gender", "exposure", "y", "treatment", "year")
```

### Problem 1

Let  $y_{ij}$  be the number of new skin cancers for subject  $i = 1, 2, \dots, 1683$  at visit  $j = 1, 2, \dots, 5$ .  $y_{ij}|b_i$  follows poisson distribution.

Model:  $\log(E[y_{ij}|b_i]) = b_{i0} + \beta_0 + \beta_1 * year_{ij} + \beta_2 treatment_{ij}$  where  $b_i \sim MVN(0, G)$ .

```
glm.model1 = glmer(y ~ as.factor(treatment) + year + (1|id), family = "poisson", data = skin)
summary(glm.model1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: poisson ( log )
## Formula: y ~ as.factor(treatment) + year + (1 | id)
## Data: skin
##
##      AIC      BIC   logLik deviance df.resid
## 8426.5   8453.9 -4209.2   8418.5     7077
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.7067 -0.3827 -0.2390 -0.2264  6.5960
##
## Random effects:
## Groups Name      Variance Std.Dev.
## id      (Intercept) 2.189    1.48
## Number of obs: 7081, groups: id, 1683
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.46461    0.09886 -24.929  <2e-16 ***
## as.factor(treatment)1  0.17309    0.09978  1.735   0.0828 .
## year              0.01827    0.01792  1.020   0.3080
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) as.()1
## as.fctr(t)1 -0.538
## year        -0.504  0.010
```

## Problem 2

Fixed effect:

$\beta_0$  : Log of the number of new skin cancers for a “typical” or “average” individual in the placebo group at baseline is -2.465.

$\beta_1$  : Log ratio of the number of new skin cancers for one unit increase in year in a given individual is 0.17.

$\beta_2$  : Log ratio of the number of new skin cancers for two individuals who have similar underlying propensity in beta carotene group versus placebo group in the same year is 0.018.

Random effect:

$b_i$  : patient-to-patient difference. It indicates overall, difference of response level between given individual and population average level. The difference among patients have variance of 2.189.

## Problem 3

According to results above, we can find that the coefficient of treatment is not significant, which means there are no significant result to conclude that beta carotene have significant effect on prevent skin cancers.

## Problem 4

```
glm.model2 = glmer(y ~ as.factor(treatment) + year + skin + age + exposure + (1|id), family = "poisson")
summary(glm.model2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: poisson ( log )
## Formula: y ~ as.factor(treatment) + year + skin + age + exposure + (1 |
## id)
## Data: skin
##
##      AIC      BIC   logLik deviance df.resid
##  8071.6   8119.6  -4028.8   8057.6     7074
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.7420 -0.3471 -0.2456 -0.2090  6.4932
##
## Random effects:
##  Groups Name      Variance Std.Dev.
##  id      (Intercept) 1.376    1.173
## Number of obs: 7081, groups: id, 1683
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -4.219075   0.316626 -13.325 < 2e-16 ***
## as.factor(treatment)1  0.125473   0.087064  1.441 0.149541
## year           0.020442   0.017870  1.144 0.252641
## skin           0.328813   0.087853  3.743 0.000182 ***
## age            0.018422   0.004632  3.977 6.97e-05 ***
## exposure       0.188588   0.010675 17.666 < 2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) as.()1 year   skin   age
## as.fctr(t)1 -0.146
## year        -0.164  0.011
## skin        -0.141 -0.003  0.016
## age         -0.938  0.003  0.003  0.006
## exposure    -0.105 -0.018  0.021 -0.106 -0.044
## convergence code: 0
## Model failed to converge with max|grad| = 0.0175183 (tol = 0.001, component 1)
## Model is nearly unidentifiable: very large eigenvalue
## - Rescale variables?
```

After adjusting for skin type, age and the count of the number of previous skin cancers, the coefficient of treatment still isn't significant, which means after adjusting for skin type, age and the count of the number of previous skin cancers, there are no significant result to conclude that beta carotene have significant effect on prevent skin cancers.

## Problem 5

When interpreting the GEE model, we are interpreting in population level while we are interpreting in individual level when interpreting the mixed effect model.

```
gee.model <- gee(y ~ as.factor(treatment) + year, id = id, data = skin, family = poisson(), corstr = 'e
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
```

```
## running glm to get initial regression estimate
```

```
##          (Intercept) as.factor(treatment)1          year
##          -1.393567643          0.143399996          0.004714574
```

```
summary(gee.model)
```

```
##
## GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
## gee S-function, version 4.13 modified 98/01/27 (1998)
##
## Model:
## Link:                      Logarithm
## Variance to Mean Relation: Poisson
## Correlation Structure:     Exchangeable
##
## Call:
## gee(formula = y ~ as.factor(treatment) + year, id = id, data = skin,
##      family = poisson(), corstr = "exchangeable")
##
## Summary of Residuals:
##          Min          1Q          Median          3Q          Max
```

```
## -0.3079685 -0.2923503 -0.2610760 -0.2478360 21.6973282
##
##
## Coefficients:
##               Estimate Naive S.E.      Naive z Robust S.E.
## (Intercept)    -1.41233647 0.10241441 -13.7904075  0.10797418
## as.factor(treatment)1  0.14783744 0.11087166   1.3334105  0.10942660
## year            0.01734823 0.02249224   0.7712987  0.02473525
##               Robust z
## (Intercept)    -13.0803165
## as.factor(treatment)1  1.3510192
## year            0.7013568
##
## Estimated Scale Parameter:  2.648264
## Number of Iterations:  3
##
## Working Correlation
##           [,1]      [,2]      [,3]      [,4]      [,5]
## [1,]  1.0000000  0.3776243  0.3776243  0.3776243  0.3776243
## [2,]  0.3776243  1.0000000  0.3776243  0.3776243  0.3776243
## [3,]  0.3776243  0.3776243  1.0000000  0.3776243  0.3776243
## [4,]  0.3776243  0.3776243  0.3776243  1.0000000  0.3776243
## [5,]  0.3776243  0.3776243  0.3776243  0.3776243  1.0000000
```

Interpretation:

$\beta_0$  : On average, log of the number of new skin cancers in the placebo group at baseline is -1.412.

$\beta_1$  : On average, log ratio of the number of new skin cancers for one unit increase in year is 0.148.

$\beta_2$  : On average, log ratio of the number of new skin cancers in beta carotene group versus placebo group in the same year is 0.017.