

# HW3 Answer

Guojing Wu / UNI: gw2383

11/25/2019

**Question 1.** In a randomized, double-blind, parallel-group, multicenter study comparing two oral treatments (denoted A and B) for toe-nail infection (De Backer et al., 1998; also see Lesaffre and Spiessens, 2001), patients were evaluated for the degree of onycholysis (the degree of separation of the nail plate from the nail-bed) at baseline (week 0) and at weeks 4, 8, 12, 24, 36, and 48 thereafter. The onycholysis outcome variable is binary (none or mild versus moderate or severe). The binary outcome was evaluated on 294 patients comprising a total of 1908 measurements. The main objective of the analyses is to compare the effects of oral treatments A and B on changes in the probability of the binary onycholysis outcome over the duration of the study. The raw data are stored in an external file: toenail.dat Each row of the data set contains the following five variables: ID,Y,Treatment,Month,Visit. The binary onycholysis outcome variable Y is coded 0 = none or mild, 1 = moderate or severe. The categorical variable Treatment is coded 1=oral treatment A, 0=oral treatment B. The variable Month denotes the exact timing of measurements in months. The variable Visit denotes the visit number (visit numbers 1-7 correspond to scheduled visits at 0, 4, 8, 12, 24, 36, and 48 weeks).

**1. Consider a random effects model with a random intercept for the log odds of moderate or severe onycholysis. Assuming linear trends and month as the time variable.**

Let  $y_{ij}$  = the binary outcome of the severity of onycholysis, then we assume  $y_{ij} \sim \text{binomial}(n, p_{ij})$ .

Variance function:  $\text{Var}(y_{ij}|b_i) = \phi v(E[y_{ij}|b_i]) = 1 \cdot E[y_{ij}|b_i] \cdot (1 - E[y_{ij}|b_i])$ .

We removed 5 subjects that has only baseline observation. Then add baseline as another feature, so the model becomes:

$$\begin{aligned} g(E[y_{ij}|b_i]) &= \text{logit}(E[Y_{ij}|b_i]) \\ &= (\beta_1 + b_i) + \beta_2 \text{Treatment}_i + \beta_3 \text{Month}_{ij} + \beta_4 \text{Baseline}_i + \beta_5 \text{Treatment}_i * \text{Month}_{ij} \end{aligned}$$

where  $b_i \sim \text{MVN}(\underline{0}, G)$ ,  $G = (g_{11})$  and  $b_i$  is independent of X.

**2. Provide Interpretations for the fixed effects coefficients in your model. Interpret the random effect parameter.**

For fixed effects:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-3.7364682	0.4720985	-7.9145949	0.0000000
Treatment1	-0.2127069	0.4788883	-0.4441681	0.6569211
Month	-0.3834852	0.0489840	-7.8287900	0.0000000

	Estimate	Std. Error	z value	Pr(> z )
Baseline1	5.7745242	0.5724672	10.0870823	0.0000000
Treatment1:Month	-0.1242952	0.0721727	-1.7221905	0.0850350

- $\beta_1$ : the log odds of outcome at month 0 is -3.736, for a typical individual who belongs to treatment B and whose baseline outcome is 0.
- $\beta_2$ : the log odds ratio of outcome at month 0 between two individuals, who belong to treatment A and treatment B respectively, but have similar underline propensity for response and have the same outcome at baseline, is -0.213.
- $\beta_3$ : the log odds ratio of outcome for 1 unit increase in Month is -0.383, given a specific individual belongs to treatment B.
- $\beta_4$ : the log odds ratio of outcome at a specific month between two individuals, who have different outcomes at baseline, but have similar underline propensity for response and belong to the same group, is 5.775.
- $\beta_3 + \beta_5$ : the log odds ratio of outcome for 1 unit increase in Month is -0.508, given a specific individual belongs to treatment A.
- $\beta_5$ : the difference in 'log odds ratio of outcome for 1 unit increase in Month' between two individuals, who belong to treatment A and treatment B respectively, is -0.124

For random effect:

grp	var1	var2	vcov	sdcor
Subject_ID	(Intercept)	NA	6.39713	2.529255

$\text{Var}(b_i) = g_{11} = 6.397$ , is the variance of intercept( $\beta_1$ ) random effect.

**3. From the results of your analysis what conclusions do you draw about the effect of treatment on changes in the severity of onycholysis over time? Provide results that support your conclusions.**

Get the p-value for  $\beta_2$  and  $\beta_5$ :  $p_{\beta_2} = 0.657$ ,  $p_{\beta_5} = 0.085$ . All greater than 0.05, so we accept the null hypothesis and state that there is enough evidence to support 'Treatment' has no effect on changes in the severity of onycholysis over time.

**4. How are the interpretations different from the GEE model.**

- GEE interpret the parameters as population average
- GLMM interpret the parameters as subject-specific

**Question 2.** The Skin Cancer Prevention Study was a randomized, double-blind, placebo-controlled clinical trial of beta carotene to prevent non-melanoma skin cancer in high-risk subjects (Greenberg et al., 1989, 1990; also see Stukel, 1993). A total of 1805 subjects were randomized to either placebo or 50 mg of beta carotene per day for 5 years. Subjects were examined once a year and biopsied if a cancer was suspected to determine the number of new skin cancers occurring since the last exam. The outcome variable is a count of the number of new skin cancers per year. The outcome was evaluated on 1683 subjects comprising a total of 7081 measurements. The main objective of the analyses is to compare the effects of beta carotene on skin cancer rates. The raw data are stored in an external file: skin.dat Each row of the data set contains the following 9 variables: ID, Center, Age, Skin, Gender, Exposure, Y, Treatment and Year.

**Note:** The outcome variable Y is a count of the of the number of new skin cancers per year. The categorical variable Treatment is coded 1=beta carotene, 0 =placebo. The variable Year denotes the year of follow-up. The categorical variable Gender is coded 1 male, 0 female. The categorical variable Skin denotes skin type and is coded 1 = burns, 0 otherwise. The variable Exposure is a count of the number of previous skin cancers. The variable Age is the age (in years) of each subject at randomization.

1. Set up a suitable random effects (random intercept) model for rate of skin cancers with Treatment and Year as covariates.

Let  $y_{ij}$  = the count of new skin cancer per year, then we assume  $y_{ij} \sim \text{Pos}(\lambda_{ij})$ .

Variance function:  $\text{Var}(y_{ij}|b_i) = \phi v(E[y_{ij}|b_i]) = 1 \cdot E[y_{ij}|b_i]$ .

the model looks like:

$$\begin{aligned} g(E[y_{ij}|b_i]) &= \log(E[Y_{ij}|b_i]) \\ &= (\beta_1 + b_i) + \beta_2 \text{Treatment}_i + \beta_3 \text{Year}_{ij} + \beta_4 \text{Treatment}_i * \text{Year}_{ij} \end{aligned}$$

where  $b_i \sim \text{MVN}(\underline{0}, G)$ ,  $G = (g_{11})$  and  $b_i$  is independent of X.

2. Provide Interpretations for the fixed effects coefficients in your model. Interpret the random effect parameter.

For fixed effects:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.4149720	0.1108829	-21.7794798	0.0000000
Treatment1	0.0798434	0.1383237	0.5772218	0.5637897
Year	-0.0002495	0.0261529	-0.0095418	0.9923869
Treatment1:Year	0.0348951	0.0358871	0.9723568	0.3308731

- $\beta_1$ : the log rate of outcome at year 0 is -2.415, for a typical individual who belongs to placebo group.
- $\beta_2$ : the log rate ratio of outcome at year 0 between two individuals, who belong to beta carotene and

placebo respectively, but have similar underline propensity for response, is 0.08.

- $\beta_3$ : the log rate ratio of outcome for 1 unit increase in year is 0, given a specific individual belongs to placebo group.
- $\beta_3 + \beta_4$ : the log rate ratio of outcome for 1 unit increase in year is 0.035, given a specific individual belongs to beta carotene.
- $\beta_4$ : the difference in ‘log rate ratio of outcome for 1 unit increase in year’ between two individuals, who belong to beta carotene and placebo respectively, is 0.111

For random effect:

grp	var1	var2	vcov	sdcov
ID	(Intercept)	NA	2.189255	1.479613

$\text{Var}(b_i) = g_{11} = 2.189$ , is the variance of intercept( $\beta_1$ ) random effect.

### 3. From the results of your analysis what conclusions do you draw about the effect of beta carotene on the rate of skin cancers? Provide results that support your conclusions.

Get the p-value for  $\beta_2$  and  $\beta_4$ :  $p_{\beta_2} = 0.564$ ,  $p_{\beta_4} = 0.331$ . All greater than 0.05, so we accept the null hypothesis and state that there is enough evidence to support ‘beta carotene’ has no effect on changes of the rate of skin cancers.

### 4. Repeat the above analysis adjusting for skin type, age, and the count of the number of previous skin cancers. What conclusions do you draw about the effect of beta carotene on the adjusted rate of skin cancers?

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-4.1672145	0.3205286	-13.0010699	0.0000000
Treatment1	0.0268258	0.1295852	0.2070125	0.8360001
Year	0.0008950	0.0260899	0.0343055	0.9726335
Skin1	0.3286485	0.0878530	3.7408891	0.0001834
Age	0.0184366	0.0046315	3.9807193	0.0000687
Exposure	0.1885961	0.0106753	17.6665325	0.0000000
Treatment1:Year	0.0367268	0.0357915	1.0261302	0.3048302

Get the p-value for  $\beta_2$  and  $\beta_4$ :  $p_{\beta_2} = 0.836$ ,  $p_{\beta_4} = 0.305$ . All greater than 0.05, so we accept the null hypothesis and state that there is enough evidence to support ‘beta carotene’ has no effect on changes of the rate of skin cancers.

### 5. How are the interpretations different from the GEE model.

- GEE interpret the parameters as population average
- GLMM interpret the parameters as subject-specific

## Appendix: code

```
knitr::opts_chunk$set(echo = FALSE, message = FALSE, warning = FALSE, comment = "")
library(tidyverse)
library(lme4) # for glmer
## Question 1
# load original data
toenail1 <- read.csv("toenail.dat", header = T, sep = ' ') %>%
  as_tibble() %>%
  mutate(Response = as.factor(Response),
         Treatment = as.factor(Treatment))

# remove subject only have 1 observation
toenail2 <- toenail1 %>%
  filter(!Subject_ID %in% names(which(table(toenail1$Subject_ID) == 1)))

# add baseline column
toenail3 <- toenail2 %>% filter(Month != 0)
toenail3$Baseline <- rep(subset(toenail2, Month == 0)$Response, as.numeric(table(toenail3$Subject_ID)))

toenail_glmm <- glmer(Response ~ Treatment + Month + Baseline + (1 | Subject_ID) + Treatment * Month,
                    data = toenail3, family = "binomial")

# fixed effects
toenail_sum = summary(toenail_glmm)
toenail_sum$coefficients %>% knitr::kable()

# random effect
toenail_sum$varcor %>% knitr::kable()

## Question 2
# load data
skin <- read.table("skin2.txt", header = F)
colnames(skin) = c("ID", "Center", "Age", "Skin", "Gender", "Exposure", "Y", "Treatment", "Year")
skin = skin %>%
  as_tibble() %>%
  mutate(Skin = as.factor(Skin),
         Gender = as.factor(Gender),
         Treatment = as.factor(Treatment))

skin_glmm <- glmer(Y ~ Treatment*Year + (1 | ID), data = skin, family = "poisson")

# fixed effects
skin_sum = summary(skin_glmm)
skin_sum$coefficients %>% knitr::kable()

# random effect
skin_sum$varcor %>% knitr::kable()

# adjusted model
skin2_glmm <- glmer(Y ~ Treatment*Year + Skin + Age + Exposure + (1 | ID), data = skin, family = "poisson")
skin2_sum = summary(skin2_glmm)
skin2_sum$coefficients %>% knitr::kable()
```