# P8157 HW2 yz4184

Yunlin Zhou

2022-10-17

```r
# import dataset for question 1
toenail <- fread("toenail.txt")
colnames(toenail) <- c("id", "y", "treatment", "month", "visit")
toenail$id <- as.factor(toenail$id)
toenail$treatment <- as.factor(toenail$treatment)

# import dataset for question 2
skin <- fread("skin.txt")
colnames(skin) <- c("id","center","age","skin","gender","exposure", "y", "treatment", "year")
skin$id <- as.factor(skin$id)
skin$treatment <- as.factor(skin$treatment)
skin$gender <- as.factor(skin$gender)
skin$skin <- as.factor(skin$skin)
```

## Question 1

### 1.

First, I set a model with month effect and treatment interaction.

```r
gee1 <- geeglm(y ~ treatment * (month + I(month^2)), id = id, data = toenail, family = binomial(link =
summary(gee1)
```

```
##
## Call:
## geeglm(formula = y ~ treatment * (month + I(month^2)), family = binomial(link = "logit"),
##     data = toenail, id = id, corstr = "exchangeable")
##
##  Coefficients:
##                        Estimate   Std.err   Wald Pr(>|W|)
## (Intercept)           -0.378812  0.176363  4.614  0.03172 *
## treatment1            -0.053047  0.251016  0.045  0.83263
## month                 -0.308201  0.053739 32.892 9.74e-09 ***
## I(month^2)             0.012364  0.004076  9.202  0.00242 **
## treatment1:month      -0.029879  0.081520  0.134  0.71398
## treatment1:I(month^2) -0.003161  0.006998  0.204  0.65145
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Correlation structure = exchangeable
## Estimated Scale Parameters:
##
##             Estimate Std.err
## (Intercept)   0.9988  0.2733
##   Link = identity
##
## Estimated Correlation Parameters:
##       Estimate Std.err
## alpha   0.4391  0.1405
## Number of clusters:   294  Maximum cluster size: 7
```

Then test if treatment interaction term is required.

```
L <- matrix(0,ncol=6,nrow=2)
L[1,c(5)]  <- c(1)
L[2,c(6)]  <- c(1)
L
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]    0    0    0    0    1    0
## [2,]    0    0    0    0    0    1
```

```
esticon(gee1,L=L,joint.test = TRUE)
```

```
##   X2.stat DF Pr(>|X^2|)
## 1   1.885  2     0.3896
```

As shown above, the p-value is large than 0.05. We fail to reject the null hypothesis and conclude that we don't need the treatment interaction term.

Finally, we build up a model without treatment interaction.

```
gee2 <- geeglm(y ~ treatment + (month + I(month^2)), id = id, data = toenail, family = binomial(link =
summary(gee2)
```

```
##
## Call:
## geeglm(formula = y ~ treatment + (month + I(month^2)), family = binomial(link = "logit"),
##     data = toenail, id = id, corstr = "exchangeable")
##
##  Coefficients:
##             Estimate  Std.err  Wald Pr(>|W|)
## (Intercept) -0.39889  0.17545  5.17  0.02300 *
## treatment1  -0.00653  0.25168  0.00  0.97929
## month       -0.32603  0.04039 65.17  6.7e-16 ***
## I(month^2)   0.01151  0.00326 12.43  0.00042 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation structure = exchangeable
## Estimated Scale Parameters:
```

```
##
##             Estimate Std.err
## (Intercept)    0.992   0.205
##   Link = identity
##
## Estimated Correlation Parameters:
##       Estimate Std.err
## alpha    0.442   0.113
## Number of clusters:   294  Maximum cluster size: 7
```

Since the P-values of month term and month^2 are smaller than 0.05, we conclude that we need the month term and month^2 terms. The final model is gee2.

### 2.

- beta0 = -0.39889 beta0 is the baseline log odds ratio between having moderate or severe onycholysis in population, holding all other variables constant.
- beta1 = -0.00653 beta1 is the log odds ratio between treatment A and treatment B of having moderate or severe onycholysis in population for every month in the study, holding all other variables constant.
- beta2 = -0.32603 beta2 is the log odds ratio between having moderate or severe onycholysis in population for every month in the study, holding all other variables constant.
- beta3 = 0.01151 beta3 is the log odds ratio between having moderate or severe onycholysis in population for every month^2 in the study, holding all other variables constant.

### 3.

As we can see from gee2 model, the coefficient of treatment (beta1) is negative but not significant (p-value = 0.97929).The coefficients of month (beta2 and beta3) are significant.

We can conclude that the treatment 1 might have negative effect on onycholysis but the effect is not significant. However, as time goes by, the severity of onycholysis might be affected.

### 4.

```
gee3 <- geeglm(y ~ treatment + (month + I(month^2)), id = id, data = toenail, family = binomial(link =
summary(gee3)
```

```
##
## Call:
## geeglm(formula = y ~ treatment + (month + I(month^2)), family = binomial(link = "logit"),
##     data = toenail, id = id, corstr = "unstructured")
##
##  Coefficients:
##              Estimate   Std.err   Wald Pr(>|W|)
## (Intercept) -1.53e+16  2.88e+14 2801.0  < 2e-16 ***
## treatment1  -1.25e+15  1.66e+14   56.3  6.2e-14 ***
## month        2.86e+15  8.11e+13 1244.9  < 2e-16 ***
## I(month^2)  -1.29e+14  5.90e+12  476.5  < 2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation structure = unstructured
## Estimated Scale Parameters:
##
##             Estimate  Std.err
## (Intercept) 1.38e+15 1.72e+37
##   Link = identity
##
## Estimated Correlation Parameters:
##           Estimate  Std.err
## alpha.1:2   1.0532 1.31e+22
## alpha.1:3   0.8468 1.06e+22
## alpha.1:4   0.5982 7.56e+21
## alpha.1:5   0.1918 2.39e+21
## alpha.1:6  -0.3609 4.49e+21
## alpha.1:7  -0.3653 4.56e+21
## alpha.2:3   0.8697 1.09e+22
## alpha.2:4   0.6217 7.85e+21
## alpha.2:5   0.2038 2.54e+21
## alpha.2:6  -0.3111 3.87e+21
## alpha.2:7  -0.3360 4.19e+21
## alpha.3:4   0.6804 8.58e+21
## alpha.3:5   0.1798 2.24e+21
## alpha.3:6  -0.2738 3.40e+21
## alpha.3:7  -0.2484 3.10e+21
## alpha.4:5   0.2038 2.54e+21
## alpha.4:6  -0.1742 2.17e+21
## alpha.4:7  -0.1607 2.01e+21
## alpha.5:6   0.0498 6.19e+20
## alpha.5:7  -0.0146 1.82e+20
## alpha.6:7   1.1834 1.48e+22
## Number of clusters:   294  Maximum cluster size: 7
```

The result of unstructured correlation structure is different from that using exchangeable correlation structure. In this model we can see that every coefficient is significant, but they are also very small.

```r
gee4 <- geeglm(y ~ treatment + (month + I(month^2)), id = id, data = toenail, family = binomial(link =
summary(gee4)
```

```
##
## Call:
## geeglm(formula = y ~ treatment + (month + I(month^2)), family = binomial(link = "logit"),
##     data = toenail, id = id, corstr = "ar1")
##
##  Coefficients:
##             Estimate  Std.err  Wald Pr(>|W|)
## (Intercept) -0.41343  0.16234  6.49    0.011 *
## treatment1  -0.12275  0.21801  0.32    0.573
## month       -0.32645  0.04054 64.85  7.8e-16 ***
## I(month^2)   0.01321  0.00312 17.94  2.3e-05 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation structure = ar1
## Estimated Scale Parameters:
##
##             Estimate Std.err
## (Intercept)    0.975   0.145
##   Link = identity
##
## Estimated Correlation Parameters:
##       Estimate Std.err
## alpha    0.699  0.0703
## Number of clusters:   294  Maximum cluster size: 7
```

The result of ar1 correlation structure is similar to that using exchangeable correlation structure.