

# 20201029-p8157\_hw3\_jsg2145

Jared Garfinkel

10/29/2020

## Read in data

```
data = read_delim(file = "./data/toenail.txt", delim = " ", col_names = c("id", "y", "treatment", "month", "visit"))
```

data

```
## # A tibble: 1,908 x 5
##       id      y treatment  month visit
##   <dbl> <dbl>    <dbl>  <dbl> <dbl>
## 1     1     1      1      1     0     1
## 2     1     1      1      1 0.857     2
## 3     1     1      1      1 3.54     3
## 4     1     0      1      1 4.54     4
## 5     1     0      1      1 7.54     5
## 6     1     0      1     10.0     6
## 7     1     0      1     13.1     7
## 8     2     0      0     0     1
## 9     2     0      0 0.964     2
## 10    2     1      0     2     3
## # ... with 1,898 more rows
```

## Question 1

### Part 1

Setup the random effects model.

$$\mu_{ij} = E[Y_{ij}|b_i]$$

$$\text{mean response: } \log\left(\frac{\mu_{ij}}{(1-\mu_{ij})}\right)$$

$$v(\mu_{ij}) = \mu_{ij}(1 - \mu_{ij})$$

$$\text{random effects: } b_i \sim N(0, \sigma_b^2)$$

```
lme1 = glmer(y ~ treatment + month + (1|id), data = data, family = binomial(link = "logit"))
summary(lme1)
```

```

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: y ~ treatment + month + (1 | id)
## Data: data
##
##      AIC      BIC   logLik deviance df.resid
## 1267.6   1289.9   -629.8   1259.6     1904
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.7315 -0.1399 -0.0740 -0.0088 30.1542
##
## Random effects:
## Groups Name      Variance Std.Dev.
## id      (Intercept) 20.73    4.553
## Number of obs: 1908, groups: id, 294
##
## Fixed effects:
##              Estimate Std. Error z value      Pr(>|z|)
## (Intercept) -2.31940    0.74728  -3.104      0.00191 **
## treatment   -0.68556    0.65901  -1.040      0.29820
## month        -0.46094    0.03963 -11.633 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) trtmnt
## treatment -0.364
## month      0.280  0.081

```

## Part 2

The treatment is associated with a log-odds ratio of -0.686 in the presence of moderate to severe onycholysis comparing someone in the treatment group to someone who is not in the treatment group with identical propensity for the outcome holding other covariates constant.

The month is associated with a log-odds ratio of -0.461 in the presence of moderate to severe onycholysis for each one unit increase in month for a typical individual holding other covariates constant.

The intercept is the log-odds of moderate to severe onycholysis for a typical individual at month = 0,

## Part 4

compared to the generalized estimating equation model in HW2, the sign of the association changes. That is, the treatment appears to work when the random effects are included in the model, whereas the treatment appears to not work under the linear effects model.

## Question 2

Read in data

```
df = read_delim(file = "./data/skin.txt", delim = " ", col_names = c("id", "center", "age", "skin", "gender", "exposure", "y", "tr", "year", "current_age"))
df = mutate(tr = as_factor(tr),
            year = as.numeric(year),
            y = as.numeric(y),
            age = as.numeric(age),
            gender = as_factor(gender),
            skin = as_factor(skin),
            exposure = as.numeric(exposure),
            current_age = age + year - 1)

df
```

```
## # A tibble: 7,081 x 10
##   id      center    age skin  gender exposure    y tr    year current_age
##   <chr>   <chr>   <dbl> <fct> <fct>   <dbl> <dbl> <fct> <dbl>      <dbl>
## 1 "      10~ "    ~  51 "    ~ "    ~      4  0 "    ~      1      51
## 2 "      10~ "    ~  51 "    ~ "    ~      4  1 "    ~      2      52
## 3 "      10~ "    ~  51 "    ~ "    ~      4  1 "    ~      3      53
## 4 "      10~ "    ~  51 "    ~ "    ~      4  1 "    ~      4      54
## 5 "      10~ "    ~  51 "    ~ "    ~      4  0 "    ~      5      55
## 6 "      10~ "    ~  68 "    ~ "    ~      2  0 "    ~      1      68
## 7 "      10~ "    ~  68 "    ~ "    ~      2  0 "    ~      2      69
## 8 "      10~ "    ~  68 "    ~ "    ~      2  0 "    ~      3      70
## 9 "      10~ "    ~  68 "    ~ "    ~      2  0 "    ~      4      71
## 10 "     10~ "    ~  68 "    ~ "    ~      2  0 "    ~      5      72
## # ... with 7,071 more rows
```

### Part 1

```
lme2 = glmer(y ~ tr + year + (1|id), data = df, family = poisson(link = "log"))
```

### Part 2

```
summary(lme2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: poisson ( log )
## Formula: y ~ tr + year + (1 | id)
## Data: df
##
##      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 <0.0000000000000002 ***
## tr          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) tr1
## tr          1 -0.538
## year         -0.504  0.010
```

The fixed effects coefficients show that there is a log rate ratio of 0.1731 for a typical individual comparing someone in the treatment group with someone not in the treatment group with identical propensity for the outcome holding other covariates constant.

There is a log rate ratio increase in the number of skin cancers of 0.0183 for a typical individual for a one unit increase in year for a typical individual in the treatment group.

The intercept of -2.46 is the log rate of new growths of cancer for a typical individual at year = 0 in the non-treatment group.

## Part 3

Based on these results, this study is not evidence that beta carotene is beneficial to skin health. Using both a GEE model and a random effects model, treatment with beta carotene could not be significantly associated with a change in the number of new cancer growths in patients in a population based inference or an individual based inference.

## Part 4

```
lme3 = glmer(y ~ tr + year + skin + age + exposure + (1|id), data = df, family = poisson(link = "log"))
```

```
summary(lme3)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##   Family: poisson ( log )
## Formula: y ~ tr + year + skin + age + exposure + (1 | id)
##   Data: df
##
```

```

##      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  6.4931
##
## Random effects:
##   Groups Name      Variance Std.Dev.
##   id       (Intercept) 1.375    1.173
## Number of obs: 7081, groups: id, 1683
##
## Fixed effects:
##              Estimate Std. Error z value      Pr(>|z|)
## (Intercept) -3.889942   0.316416 -12.294 < 0.0000000000000002 ***
## tr          1      0.125419   0.087064   1.441      0.149716
## year                0.020454   0.017870   1.145      0.252375
## skin          0    -0.328851   0.087853  -3.743      0.000182 ***
## age                0.018417   0.004632   3.976      0.0000701 ***
## exposure        0.188583   0.010675  17.666 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) tr1    year    skin0    age
## tr          1      -0.147
## year                -0.159  0.011
## skin          0    -0.136  0.003 -0.016
## age                -0.937  0.003  0.003 -0.006
## exposure        -0.134 -0.018  0.021  0.106 -0.044
## convergence code: 0
## Model failed to converge with max|grad| = 0.00416052 (tol = 0.002, component 1)
## Model is nearly unidentifiable: very large eigenvalue
## - Rescale variables?

```

## Part 5

After adjusting for skin type, age, and number of previous skin cancers recorded, there is still no evidence of a significant association between the treatment and outcome of interest.

## Part 6

This is similar to the results in HW2 since it could not be shown that there was an association between beta carotene treatment and skin cancer growth either in population based inference nor individual based inference.