P8157 HW3

Anna Ma

2022-10-30

Question 1

Randomized, double-blind, parallel-group, multicenter study comparing two oral treatments (denoted A and B) for toe-nail infection, 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 *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).

```
toenail_df = read_delim(file = "toenail.txt", delim = " ", col_names = c("id", "response", "treatment",
```

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.

- Setup: Since the response variable is binary such that $Y_{ij} = [0, 1]$, we can assume that $y_{ij} \sim binomial(n, p_{ij})$, and that the mean response model is $log(\frac{\mu_{ij}}{1-\mu ij})$
- Random Effects Model with random intercept:
  - $y_{ij} = X_{ij}\beta + b_i + \epsilon_{ij}$
  - $b_i \sim N(0, \sigma_b^2)$
  - $\epsilon_i \sim N(0, \sigma^2)$
  - ϵ_{ij} and b_i are independent
- Mean:
  - Population average: $logit(E[Y_{ij}]) = X_{ij}\beta$
  - Subject specific mean: $logit(E[Y_{ij}|b_i]) = X_{ij}\beta + b_i$
- Variance and Covariance:
  - $Var[Y_{ij}] = \phi v(\mu_{ij}) = \mu_{ij}(1 - \mu_{ij})$
  - Total Variance $Var[Y_{ij}] = \sigma_b^2 + \sigma^2$
  - Within Subject Variance $Var[Y_{ij}|b_i] = \sigma^2$
  - Covariance: $Cov(Y_{ij}, Y_{ik}) = \sigma_b^2$

The model can be written as:

$$logit(E[Y_{ij}|b_i]) = (\beta_0 + b_i) + \beta_1 Treatment_i + \beta_2 Month_{ij} + +\beta_3 Treatment_i * Month_{ij}$$

```
rme_q1 = glmer(response ~ treatment*month + (1|id), family = "binomial", data = toenail_df)
summary(rme_q1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: response ~ treatment * month + (1 | id)
##    Data: toenail_df
##
##      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.
##  id     (Intercept) 20.76    4.557
## Number of obs: 1908, groups:  id, 294
##
## Fixed effects:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -2.50986    0.76401  -3.285  0.00102 **
## treatment1       -0.30483    0.68709  -0.444  0.65729
## month            -0.39973    0.04679  -8.543  < 2e-16 ***
## treatment1: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) trtmn1 month
## treatment1  -0.374
## month        0.130  0.220
## trtmnt1:mnt  0.181 -0.265 -0.541
```

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

For fixed effects:

```
coef_q1 = summary(rme_q1)
coef_q1$coefficients  %>% knitr::kable(digits = 3)
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.510	0.764	-3.285	0.001
treatment1	-0.305	0.687	-0.444	0.657
month	-0.400	0.047	-8.543	0.000
treatment1:month	-0.137	0.069	-1.973	0.048

- β_0: the log odds of the onycholysis outcome for an average/typical individual in treatment B group at baseline is -2.51.
- β_1 : the log odds ratio of the onycholysis outcome for patients in treatment A versus patients in treatment group B at baseline is -0.305.
- β_2 : the log odds ratio of the onycholysis outcome for every unit increase in month for a patient in treatment group B is -0.4
- β_3 : the difference in the log odds ratio of the onycholysis outcome for a individual in treatment A group versus a individual treatment B group with one unit increase of month is -0.137.
- $\beta_2 + \beta_3$: the log odds ratio of the onycholysis outcome for every unit increase in month for a patient in treatment group A is -0.537

For random effect:

```
coef_q1$varcor %>% knitr::kable(digits = 3)
```

grp	var1	var2	vcov	sdcor
id	(Intercept)	NA	20.762	4.557

- Random effect parameter: The variance of the random effects σ_b^2 is 20.76, this suggest that there is variation in the random effects between individuals.
- 95% Confidence Interval: $\hat{\beta}_0 \pm 1.96 * \sigma_b = -2.51 \pm 1.96 * 4.56 = [-11.45, 6.43]$. Therefore, 95% of the subjects have the intercept in the range [-11.45, 6.43].

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.

From the model,we see that the odds ratio of the outcome for patients in treatment A versus patients in treatment group B at baseline is 0.737. And since the estimated coefficient for this interaction is -0.137, which is negative, the difference between the log odds ratio of the onycholysis outcome for a individual in treatment A group versus a individual treatment B grows in the negative direction over time, meaning that the effect of treatment increases over time.

However, from the model,we see that with a p-value of 0.65728, the treatment beta value is not significant. So, although the p-value for the interaction term is 0.048 and therefore significant for the response outcome, it's effect is likely to be driven by month instead of the treatment itself.

4. How are the interpretations different from the model in HW2.

- HW 2 uses a GEE model, which interprets the parameters as population average.
- This HW uses a mixed effect model, random effect with random intercept, which interprets the parameters on an individual-specific level.

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.A total of 1805 subjects were randomized to either placebo or 50 mg of beta carotene per day for 5 years.

The main objective of the analyses is to compare the effects of beta carotene on skin cancer rates.

- 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.

```
skin_df = read.table(file = "skin.txt", header = FALSE,col.names = c("id", "center", "age", "skin", "ge
```

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

- Setup: Since the response variable is a count of the number of new skin cancers per year, we can assume that $y_{ij} \sim poisson(\lambda_{ij})$, and that the mean response model is $log(\mu_{ij}) = \eta_{ij}$
- Random Effects Model with random intercept:
  - $y_{ij} = X_{ij}\beta + b_i + \epsilon_{ij}$
  - $b_i \sim N(0, \sigma_b^2)$
  - $\epsilon_i \sim N(0, \sigma^2)$
  - ϵ_{ij} and b_i are independent
- Mean:
  - Population average: $log(E[Y_{ij}]) = X_{ij}\beta$
  - Subject specific mean: $log(E[Y_{ij}|b_i]) = X_{ij}\beta + b_i$
- Variance and Covariance:
  - $V(\mu_{ij}) = \mu_{ij}$
  - Total Variance $Var[Y_{ij}] = \sigma_b^2 + \sigma^2$
  - Within Subject Variance $Var[Y_{ij}|b_i] = \sigma^2$
  - Covariance: $Cov(Y_{ij}, Y_{ik}) = \sigma_b^2$

The model can be written as:

$$log(E[Y_{ij}|b_i]) = (\beta_0 + b_i) + \beta_1 Treatment_i + \beta_2 Year_{ij} + +\beta_3 Treatment_i * Year_{ij}$$

```
rme_q2 = glmer(y ~ treatment*year + (1|id), family = "poisson", data = skin_df)
summary(rme_q2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: y ~ treatment * year + (1 | id)
##    Data: skin_df
##
##      AIC      BIC   logLik deviance df.resid
##   8427.5   8461.8  -4208.8   8417.5     7076
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.6779 -0.3890 -0.2394 -0.2260  6.4430
##
## 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)     -2.4149719  0.1109094 -21.774   <2e-16 ***
## treatment1       0.0798434  0.1383589   0.577    0.564
## year            -0.0002495  0.0261557  -0.010    0.992
## treatment1:year  0.0348951  0.0358915   0.972    0.331
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) trtmn1 year
## treatment1  -0.660
## year        -0.638  0.505
## tretmnt1:yr  0.458 -0.693 -0.728
```

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

For fixed effects:

```
coef_q2 = summary(rme_q2)
coef_q2$coefficients  %>% knitr::kable(digits = 3)
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.415	0.111	-21.774	0.000
treatment1	0.080	0.138	0.577	0.564
year	0.000	0.026	-0.010	0.992
treatment1:year	0.035	0.036	0.972	0.331

- β_0: For an average/typical patients in the placebo group, the log rate of the number of new skin cancers at baseline is -2.415.
- β_1 : The log rate of the number of new skin cancers for a patient in the treatment group versus a patient in the placebo group is 0.08.
- β_2 : The log rate of the number of new skin cancers for one unit increase in year for a patient in the placebo group is 0.
- β_3 : The difference in the log rate of the number of new skin cancers for one unit increase in year between a patient in treatment group versus a patient in the placebo group is 0.035
- $\beta_2 + \beta_3$: The log rate of the number of new skin cancers for one unit increase in year in a patient in the treatment group is 0.035

Random effect:

```
coef_q2$varcor %>% knitr::kable(digits = 3)
```

grp	var1	var2	vcov	sdcor
id	(Intercept)	NA	2.189	1.48

- Random effect parameter: The variance of the random effects σ_b^2 is 2.189, this suggest that there is not a lot variation in the random effects between individuals.
- 95% Confidence Interval: $\hat{\beta}_0 \pm 1.96 * \sigma_b = -2.415 \pm 1.96 * 1.48 = [-5.32, 0.486]$. Therefore, 95% of the

subjects have the intercept in the range [-5.32, 0.486].

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.

From the model,we can see that the p-value for the treatment covariate is 0.564, and that the p-value for the interaction term of treatment and year is 0.331. Both of the p-value is greater than the significance level of 0.05, therefore we fail to reject the null and conclude that the effect of both covariates are not significant for the response outcome.Thus, there's not enough evidence to show that beta carotene have a significant effect on the rate of skin cancers.

```
rme_q2_2 = glmer(y ~ treatment*year + skin + age + exposure + (1 | id), family = "poisson", data = skin
```

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?

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0534699 (tol = 0.002, component 1)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unide
##  - Rescale variables?
```

```
coef_q22 = summary(rme_q2_2)
coef_q22$coefficients  %>% knitr::kable(digits = 3)
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-4.166	0.321	-12.998	0.000
treatment1	0.028	0.130	0.214	0.831
year	0.001	0.026	0.046	0.964
skin	0.329	0.088	3.751	0.000
age	0.018	0.005	3.973	0.000
exposure	0.189	0.011	17.670	0.000
treatment1:year	0.036	0.036	1.016	0.310

From the model, we can see that the p-value for the treatment covariate is 0.831, and that the p-value for the interaction term of treatment and year is 0.310. Both of the p-value is greater than the significance level of 0.05, therefore we fail to reject the null and conclude that the effect of both covariates are not significant for the response outcome.Thus, like the previous model, there's not enough evidence to show that beta carotene have a significant effect on the rate of skin cancers. On the other hand, covariates skin, age, and exposure have significant effect on the outcome response.

5. How are the interpretations different from the model in HW2.

- HW 2 uses a GEE model, which interprets the parameters as population average.
- This HW uses a mixed effect model, random effect with random intercept, which interprets the parameters on an individual-specific level.