Homework 5

Joseph Froelicher

December 24, 2021

Question 1

For a traditional linear mixed model with a normal outcome, the likelihood is easily written down, and there is a closed form solution the the maximization of the likelihood. This simply means that there is a numerical solution to the maximization of the likelihood, because the random effects can be easily integrated out. However, for mixed models with non-normal outcomes, the random effects are not so easily integrated out. The solution to the maximization problem is not nearly as straightforward. The likelihood cannot be written in closed form, so approximation methods (like Gaussian quadrature, as mentioned, or the Laplace method) are used for the likelihood functions. So the complication is that the likelihood itself is defined by an approximation.

Question 2

Part A

Counts of Medication use For medication, because we have count data, we are going to have to build a generalized linear mixed effects model, using a poisson distribution and the log link. We can do this using the `lme()` function in the `nlme` package. In R we are slightly limited by correlation structures, but I would suggest autoregressive structure (AR(1)), specified by `correlation = corAR1(form = counts ~ time | subject)`. You could also use spatial correlation structure, `correlation = corSpatial(form = counts ~ time | subject)`. Lastly, we need to be conscious of overdispersion and may need to adjust the variance by a factor of ϕ.

FEV1 For FEV1, because we have normal data, we can build a linear mixed effects model, using a gaussian distribution. We can do this using the `lme()` function in the `nlme` package. In R we are slightly limited by correlation structures, but I would suggest autoregressive structure (AR(1)), again, specified by `correlation = corAR1(form = counts ~ time | subject)`. You could also use spatial correlation structure, `correlation = corSpatial(form = counts ~ time | subject)`.

Part B

In this case, it sounds like the investigator is interested in the between subject variability. In order to have different between subject variability, we would need to have a **G** matrix for each of the used at least one medication group, and the use no medications group. This allows us to accound for variability between days where medication is used and days where medication is not used.

Counts of Medication use   Again, generalized linear mixed effects model, using the log link on the poisson distribution. This method is going to require a numerical approximation method for calculating the likelihood (random effects cannot be integrated out). This will be a challenging model to interpret, and the numerical integration are drawbacks.

FEV1   Again, linear mixed effects model, with a Gaussian distribution, with separate **G** matrices for use and no use groups.

Part C

Counts of Medication use   For this model, we are going to need a generalized linear mixed effects model. In order to do this we need to be able to specify both the **G** and **R** matrices (I am not actually sure how to do this in R?). Some other thoughts of mine on the approach include, using a GEE GLM model, and specify the two matrices in there, but then we have neither a closed form solution (GLM), nor the marginal interpretation (GEE), or potentially modeling the medication use as a continuous variable and then using a linear mixed model. And lastly we could

FEV1   This is fairly straightforward for FEV1. We can specify a linear mixed effects model with both random slopes and random intercepts. We can also specify the correlation structure to be AR(1). This is exactly what a linear mixed effects model was designed to do.

Question 3

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
data = read.table("../data/toenail-data.txt")
model1 = glmer(
  outcome ~ -1 + factor(treatment)*month + (1 | id),
  family = binomial,
  data = data
)

model1_sum = summary(model1)
model1_sum
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: outcome ~ -1 + factor(treatment) * month + (1 | id)
##    Data: data
##
##      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
```

2

```
## 
## 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|)
## factor(treatment)0      -2.50986    0.76389  -3.286 0.001018 **
## factor(treatment)1      -2.81469    0.81440  -3.456 0.000548 ***
## month                   -0.39973    0.04679  -8.543  < 2e-16 ***
## factor(treatment)1:month -0.13713    0.06949  -1.973 0.048460 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Correlation of Fixed Effects:
##            fct()0 fct()1 month
## fctr(trtm)1  0.623
## month        0.130  0.307
## fctr(trt)1:  0.181 -0.053 -0.541
```

The average individual in the Intraconazole treatment was 75.18% less likely to have onycholysis (toenail separation) per month (interpretation of the group effect, and month effect). The average individual in the Ter treatment was 160.23% more likely to have onycholysis (toenail separation) per month (interpretation of the group effect, month effect, and interaction effect.

```
library(gee)

model2 = gee(
  outcome ~ -1 + factor(treatment)*month,
  family = binomial(link = "logit"),
  id = id,
  data = data,
  corstr= "unstructured"
)
```

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


## running glm to get initial regression estimate


##       factor(treatment)0       factor(treatment)1                   month
##             -0.55662726             -0.55720892            -0.17030779
## factor(treatment)1:month
##             -0.06722162
```

```
model2_sum = summary(model2)
model2_sum
```

```
## 
##   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:     Unstructured
##
## Call:
## gee(formula = outcome ~ -1 + factor(treatment) * month, id = id,
##     data = data, family = binomial(link = "logit"), corstr = "unstructured")
##
## Summary of Residuals:
##         Min           1Q       Median           3Q          Max
## -0.34037660 -0.25236686 -0.12222540 -0.03330277   0.97443190
##
##
## Coefficients:
##                           Estimate Naive S.E.   Naive z Robust S.E.  Robust z
## factor(treatment)0       -0.69921366 0.17026555 -4.106607   0.16700096 -4.186884
## factor(treatment)1       -0.66161643 0.17065471 -3.876930   0.17768736 -3.723486
## month                    -0.14136770 0.02652509 -5.329584   0.02700355 -5.235152
## factor(treatment)1:month -0.08285003 0.04280132 -1.935689   0.04798823 -1.726466
##
## Estimated Scale Parameter:  1.036634
## Number of Iterations:  5
##
## Working Correlation
##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## [1,] 1.0000000 0.8849438 0.6896985 0.4860334 0.2353017 0.1436830 0.1020914
## [2,] 0.8849438 1.0000000 0.7984038 0.5769551 0.2579315 0.2184895 0.1228562
## [3,] 0.6896985 0.7984038 1.0000000 0.7486853 0.2689375 0.1958789 0.1491227
## [4,] 0.4860334 0.5769551 0.7486853 1.0000000 0.3480472 0.2566720 0.1940261
## [5,] 0.2353017 0.2579315 0.2689375 0.3480472 1.0000000 0.4527287 0.3680872
## [6,] 0.1436830 0.2184895 0.1958789 0.2566720 0.4527287 1.0000000 0.5487164
## [7,] 0.1020914 0.1228562 0.1491227 0.1940261 0.3680872 0.5487164 1.0000000
```

The average individual in the Intraconazole treatment was 136.52% more likely to have onycholysis (toenail separation) per month (interpretation of the group effect, and month effect). The average individual in the Ter treatment was 230.47% more likely to have onycholysis (toenail separation) per month (interpretation of the group effect, month effect, and interaction effect.

Question 4

```
/*Note v= 1 if visit>0; otherwise v=0*/
title "RANDOM INTERCEPT ONLY, ADAPTIVE QUADATURE";
proc glimmix data=seizure method=quad(qpoints=25);
  class subj;
  model seize = age v trt trt*v / solution link=log
       dist=poisson offset=logo;
  random int / subject=subj type=un;
run;
```

Part A

There was a significant result for both the intercept and the indicator variable created that shows if the subjects visit was at baseline or after baseline. The intercept should not be interpreted (age is never 0, exprrapolation). That being said, the intecept is interpreted as the rate ratio of seizures for those on treatment over those on placebo (again at age 0).

A one year increase in age resulted in $e^{0.01543} = 1.0155$ (p = 0.35), times increase in the odds of seizure. The odds ratio of seizures for the treatment group over the placebo group was $e^{0.1108} = 1.117$ (p = 0.0189). And the treatment group and visit after baseline group interaction resulted in a $e^{0.1037} = 1.118$ time increase in the odds of seizure (p = 0.1123).

Part B

The intercept should not be interpreted (age is never 0). From the GEE model, using the empirical standard error estimates, each one year increase in age results in a $e^{0.0321} = 1.0326$ time increase in the odds (p = 0.0296, statistically significat). The visit occurring after baseline results in a $e^{0.1108} = 1.117$ increase in the odds (p= 0.3399). Being in the treatment group is associated with a $e^{0.0175} = 1.0176$ increase in the odds compared to the non-treatment group (p = 0.9348). Lastly, Being in the treatment group and the visit occurring after baseline resulted in a $e^{0.1037} = 1.109$ increase in the odds of having a seizure (p = 0.6274).