

lab7

Ben Tankus

5/13/2021

Lab Questions

In the `epilepsy` data, we really need to use a mixed effects model to account for the fact that observations within a subject are almost certainly not independent. Let's suppose, however, that we are a naive statistician and do not realize that we need to account for this within subject correlation. Instead we fit a Poisson GLM and completely ignore the fact that we have non-independent observations (to be clear, this is **not** a correct way to do the analysis, but we are fitting the model to compare the results we get to the results from the more appropriate GLMM we fit above).

Please write 1-2 sentences in response to the following questions.

```
epil_long <- gather(epilepsy, key = Visit, value = Seizures, Y1, Y2, Y3, Y4)
```

```
mod3 <- glmer.nb(Seizures ~ Trt + Age + Base + Visit + (1|ID), data = epil_long)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
## Model failed to converge with max|grad| = 0.0172878 (tol = 0.002, component 1)
```

```
mod.norand <- glm.nb(Seizures ~ Trt + Age + Base + Visit, data = epil_long)  
summary(mod.norand)
```

```
##  
## Call:  
## glm.nb(formula = Seizures ~ Trt + Age + Base + Visit, data = epil_long,  
##       init.theta = 2.428203315, link = log)  
##  
## Deviance Residuals:  
##      Min       1Q   Median       3Q      Max   
## -2.9093  -0.8063  -0.2144   0.2853   3.6267   
##  
## Coefficients:  
##              Estimate Std. Error z value Pr(>|z|)      
## (Intercept)  0.555805   0.283346   1.962   0.0498 *      
## Trtprogabide -0.193221   0.101384  -1.906   0.0567 .      
## Age          0.018153   0.008223   2.208   0.0273 *      
## Base         0.026901   0.001753  15.346 <2e-16 ***   
## VisitY2      0.064836   0.141120   0.459   0.6459        
## VisitY3     -0.004888   0.141815  -0.034   0.9725        
## VisitY4     -0.128060   0.143150  -0.895   0.3710        
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(2.4282) family taken to be 1)
##
##      Null deviance: 556.57  on 235  degrees of freedom
## Residual deviance: 265.11  on 229  degrees of freedom
## AIC: 1328.5
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta:  2.428
##             Std. Err.:  0.327
##
## 2 x log-likelihood:  -1312.506
```

```
summary(mod3)$coefficients
```

```
##              Estimate Std. Error   z value    Pr(>|z|)
## (Intercept)  0.57827415 0.413550444  1.3983159 1.620182e-01
## Trtprogabide -0.26163096 0.154373419 -1.6947928 9.011474e-02
## Age          0.01418787 0.012568467  1.1288462 2.589627e-01
## Base         0.02721098 0.002806655  9.6951641 3.161291e-22
## VisitY2      0.02246290 0.102895683  0.2183075 8.271896e-01
## VisitY3     -0.08997084 0.103807666 -0.8667071 3.861025e-01
## VisitY4     -0.13735905 0.104775375 -1.3109860 1.898625e-01
```

```
summary(mod.norand)$coefficients
```

```
##              Estimate Std. Error   z value    Pr(>|z|)
## (Intercept)  0.555805104 0.283346123  1.96157653 4.981181e-02
## Trtprogabide -0.193220763 0.101383701 -1.90583655 5.667141e-02
## Age          0.018153029 0.008222955  2.20760411 2.727188e-02
## Base         0.026901326 0.001752940 15.34640789 3.743487e-53
## VisitY2      0.064835509 0.141119604  0.45943659 6.459207e-01
## VisitY3     -0.004887919 0.141814890 -0.03446689 9.725048e-01
## VisitY4     -0.128059910 0.143149524 -0.89458844 3.710071e-01
```

1.) How do the coefficient estimates for the fixed effects compare in the two models? Why do you think the coefficient estimates are approximately the same or very different?

The coefficients for the fixed effects model are similar to the mixed effects models, while the std. errors are fairly different. The estimates are similar because adding a random effect will only impact the standard error.

2.) How do the standard errors for the coefficients compare in the two models? Why do you think the standard errors are approximately the same or very different?

The standard errors in the fixed effects model are smaller as it assumes the patients are independent, which introduces less uncertainty in the model.

3.) How do the p-values compare for the two models?

Following the standard error changes, the pvalues for the fixed effects model are not as statistically significant as the mixed model.

4.) Based on your answers from (1) to (3), what is the danger in assuming independence between all the observations, when, in fact, some of the observations are not independent?

There is a danger of type I error in incorrectly assuming observations are independent, rejecting a null hypothesis that it true.

Let's next look at the first 40 fitted values for the incorrectly fitted model without a random effect for subject and the GLMM with a random effect for subject.

```
cbind(fitted(mod.norand), fitted(mod3))[1:40, ]
```

```
##      [,1]      [,2]
## 1  4.114279  3.690443
## 2  4.040266  3.672372
## 3  3.225341  2.926119
## 4  4.155882  3.492867
## 5 15.342999 14.583237
## 6  6.101769  5.787391
## 7  4.226461  3.331193
## 8 15.136382 22.574001
## 9  6.335683  5.747624
## 10 3.792795  7.090316
## 11 13.574365 16.415722
## 12 6.548405  7.452118
## 13 4.295417  4.152304
## 14 10.372622 10.790652
## 15 29.026210 16.886525
## 16 10.727925  5.310506
## 17 4.703532  2.402212
## 18 60.618261 35.059546
## 19 5.057771  4.114183
## 20 4.371233  4.233891
## 21 4.075767  3.695934
## 22 3.251546  3.488480
## 23 4.923524  3.716738
## 24 5.829131  7.251840
## 25 13.196728 32.823423
## 26 4.590726  2.318799
## 27 3.221110  2.775595
## 28 9.203042 12.856515
## 29 15.391873 11.481482
## 30 7.140135  7.045624
## 31 3.444470  2.243968
## 32 3.241994  3.184119
## 33 3.321658  4.374844
## 34 4.237141  3.129299
## 35 5.703707 15.760604
## 36 3.953333  4.683581
## 37 3.153871  2.655932
```

```
## 38 12.528847 7.232704
## 39 6.455325 7.080023
## 40 2.884001 1.594328
```

5.) How do the fitted values from the two models compare? Are the fitted values from one model consistently higher or lower than the other model?

For the most part, the fitted values for the mixed model is lower than the fitted models from the fixed model.

6.) Notice fitted values 37 through 40. These values correspond to the 10th subject in the data set (subject ID 112). Can you justify why these fitted values are much higher in the GLMM? In your justification, can you include a plot of some sort that highlights this particular subject?

Subject ID 112 is the subject who was the outlier in the progabide data set, and I would discount them from the analysis. These fitted values using the mixed model are likely much higher as the model accounts for this ID as dependent, and much higher than the others. In the below visual you can see these outlier points in the bottom right of the right-most plot.

```
epil_long$resid <- residuals(mod3)
epil_long$fits <- exp(predict(mod3))
ggplot(epil_long, aes(Seizures, fits)) + geom_point() + facet_wrap(~Trt) + geom_abline()
```

