ST 518 - Homework 7

Nora Quick

R Question:

1. Recall the data in ex2119 in the Sleuth3 library. These data record results from 10 different studies in which the relationship between breast cancer and whether or not a woman had breast fed her child(ren) was examined. Earlier in the course, you looked at these data, and treated study as a fixed effect. Another way to consider these data is by taking study to be a random effect.

a. Fit an appropriate GLMM to the breast cancer data, using study as a random effect. Report the estimated random effect variance and the fixed effect estimate for the Lactate variable, as well as its standard error.

```
?ex2119
bc <- ex2119

glm3 <- glmer(cbind(Cancer,NoCancer) ~ Lactate + (1|Study), data = bc, family = binomial)
(init <- getME(glm3, name=c("theta", "fixef")))
```

```
## $theta
## Study.(Intercept)
##          1.199213
##
## $fixef
## (Intercept)  Lactateyes
##  -0.5753359  -0.1095810
```

```
glm3 <- glmer(cbind(Cancer,NoCancer) ~ Lactate + (1|Study), data = bc, start = init, family = binomial)
summary(glm3)
```

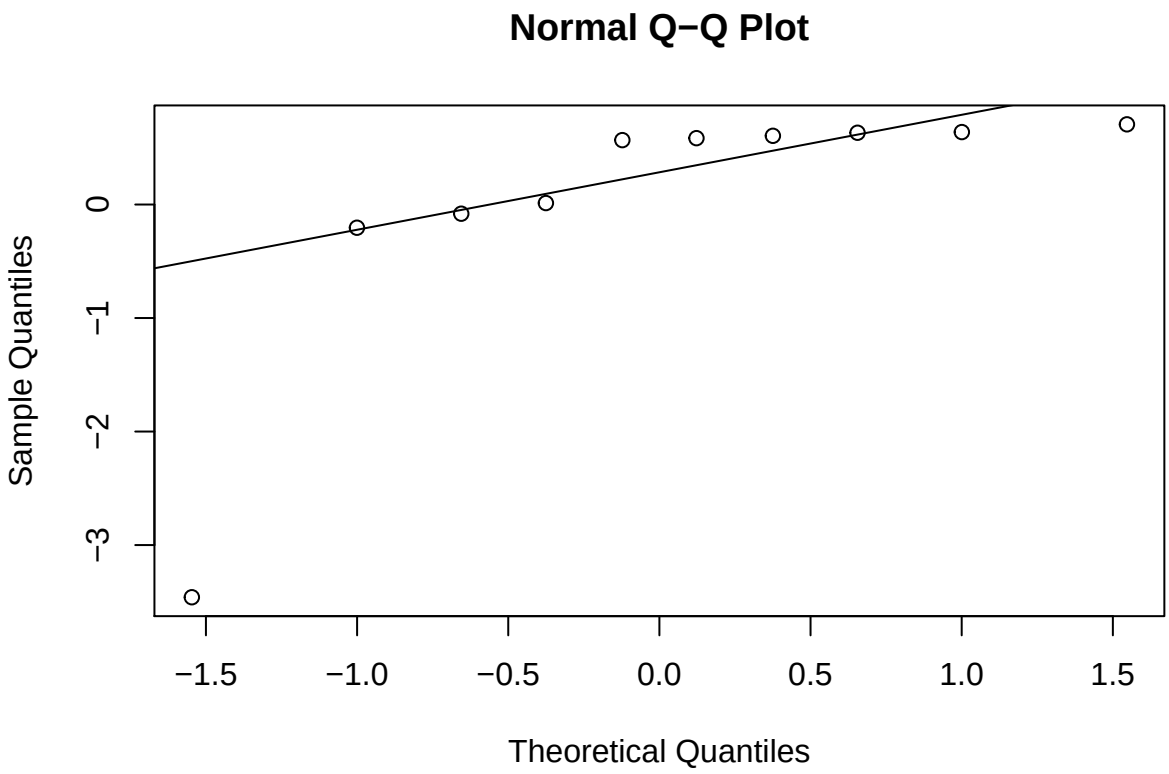
```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: cbind(Cancer, NoCancer) ~ Lactate + (1 | Study)
##    Data: bc
##
##      AIC      BIC   logLik deviance df.resid
##    248.5    251.5   -121.2    242.5       17
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -1.69254 -0.60223  0.01219  0.44225  1.44836
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  Study  (Intercept) 1.438    1.199
## Number of obs: 20, groups:  Study, 10
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.57534    0.37980  -1.515     0.13
## Lactateyes  -0.10958    0.02303  -4.758 1.95e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr)
## Lactateyes -0.036
```

The random effects variance is 1.438 and the standard deviation 1.199.The fixed effect value for lactate (yes) is -0.11 and the correlation fixed effects for lactate (yes) is -0.036.

b. Create a normal probability plot of the random effects. Do you notice anything unusual?

```
qqnorm(unlist((ranef(glm3)$Study)))
qqline(unlist((ranef(glm3)$Study)))
```



c. Write a few sentences about how you might proceed with an analysis from here. There are several different approaches you could take—just decide on one and describe it briefly.

The above plot looks a little strange so there must be an explanation. To find an explanation I would fit a new model with the outliers and then without the outliers to see how the model/plot change. If there is a difference between the two models we choose the best model.

Conceptual Questions:

2. Please answer true or false to the following. Explain/justify your answers.

a. One reason for using a mixed effects model is to account for the similarities among observations that are clustered together in groups.

True, there could be similarities/trends that only one group experiences outside the norm of the other groups. For example, from the lectures, one tank of fish could have has a similarity in bacteria in the tank causing them to have more tumors than the other tanks. We must account for that.

b. If a GLMM does not converge in R, we should just go back to using a GLM approach.

False, we need to account for the non-convergance. We need to re-run the algorithm the values from the model we fit as the starting point.

c. If an estimated random effect variance is small, that's an indication that we don't need that random effect in the model.

False, it means there could be overdispersion within the model.

3. Consider the following R output from fitting a GLMM to the Donner party data ('Donner' in the 'vcdExtra' package) using 'family' as a random effect. Write a short paragraph that describes the results from fitting this model.

Using family as the random effect we can see that it produces a relatively good AIC and BIC indicating that the model is okay for the data. The summary indicates that family is the most significant variable which could be correct. It could make sense based on older parents having older children and possibly grandchildren while younger families will have the same trend. Age and gender (male) are still significant in this model and if we recall back to the last time we used this data those two were very important factors that we wanted to see consistant here.