

# BI5302 Mixed effects models exercise

## Exercise: LMM - RIKZ data

0. In this exercise we will use the RIKZ data set (in the `RIKZRichness.txt` file). The aim of the research was to explain the variation in species richness (`R`) using the variable Normal Amsterdam Peil (`NAP`) using data collected from various beaches in the Netherlands. All measurements were taken in June 2002. Please note that the full analysis of these data is much more involved than the one suggested here. The less self-explanatory variables are:

- `R`: Species richness (number of species). This is the response variable
- `angle1`: angle of individual stations on a beach
- `angle2`: angle of the beach
- `exposure`: index summarizing multiple variables: wave action, length of the surf zone, slope, grain size and depth of anaerobic layer
- `sorting1`: index of particle structure at individual sampling station
- `NAP`: height of the sampling station relative to the average sea level

1. Load the `nlme` package to make the `lme()` function available. Also load the `lattice` package. Import the data file `RIKZRichness.txt` into R. How many observations are in the dataset? How many variables are in the dataset which could be relevant as predictors of species richness?

2. Convert the variable `Beach` to a factor and store it as a new variable called `Fbeach` in the same dataframe. Log-transform the species richness (`R`) variable and store as `logR`. Why do you think you are advised to do this?

3. How many beaches are there? How many observations for each beach?

4. Plot the relationship between log-richness (`logR`) and `NAP` conditional on each `Beach` (Hint: use the `coplot` or `xyplot` functions). Describe this relationship. Is the relationship the same for each beach?
5. Fit a simple linear model (using the `lm` function) with log transformed `R` as the response variable and `NAP` as the predictor. Produce the usual model validation plots of the residuals using the `par(mfrow= c(2, 2))` and the `plot` functions. Identify any issues highlighted by these plots. What inference does this model suggest? In addition, extract the residuals (use the `resid` function) and plot the residuals against beach. Can you see a problem?
6. Fit a new linear model as above but with `NAP` and `FBeach` as predictors. Extract the residuals and re-plot the residuals for each beach. Does this look better? How many more parameters have you estimated in this model?
7. Fit a linear mixed effects model using the `lme` function and assign it to an object called `rikz.lme`. As before use log-richness (`logR`) as the response variable and `NAP` as the predictor, but including `FBeach` as a random effect this time.
8. Extract the normalized residuals using the `residuals` function and the argument `type="normalized"`. Extract the fitted values using the `fitted` function. Plot the residuals vs the fitted values and include a horizontal line at residual value= 0. In addition, plot the normalized residuals for each Beach. How does this plot compare with the equivalent plots from the linear models? Is the model an improvement?
9. Produce a summary of the model output using the `summary` function. What is the within beach (residual) variation? What is the between beach variation? Remember, the output from the `lme` function gives these values in standard deviation units, not as variances, so you will need to calculate the variances manually. Compare the parameter estimate for `NAP` and the standard error of the parameter estimate to those obtained from the linear models.
10. (Optional) convert your `RIKZ` dataframe to a grouped data object using the command: `rikz<- groupedData(logR~NAP|Fbeach, data= rikz)`. This will allow you to use some convenience functions that come with the `nlme` package. Investigate what the following lines of code do: `plot(augPred(rikz.lme), aspect="xy", grid=T)`, as well as `plot(rikz.lme, logR~ fitted(), id=0.05, adj=-0.3)` and `qqnorm(rikz.lme, ~resid(.),abline= c(0,1))`. Do you think that fitting a random intercept model adequately explains the relationship between Richness and `NAP` for each Beach?
11. Perhaps an improvement to the model would be to allow both the intercepts and the `NAP` slopes to vary randomly with each beach (as opposed to an intercepts only model).

You can fit this model using the `lme` function and specify the random effects as `random= ~ NAP|Fbeach`. Fit the model and assign it to a new variable called `rikz.lme2`. The `random= ~ NAP|Fbeach` formula assumes that there is variation in the intercept as well as the effect of NAP between beaches. **When a continuous predictor is in the random formula it also needs to be in the fixed formula:** The effect in formula of the fixed part measures the mean coefficient at the population level (average across all beaches), and the effect in the random part formula measures variation around that mean coefficient value between beaches.

12. Plot the fitted lines using

```
plot(augPred(rikz.lme2), aspect="xy", grid=T)
```

and the residuals versus fitted with

```
plot(rikz.lme2, form=resid(., type="p")~ fitted(.)|Fbeach, abline=0, lty=2).
```

Normality of random intercepts:

```
par(mfrow= c(1, 2)),
```

```
hist(unlist(ranef(rikz.lme2)$'(Intercept)'), xlab= "Random Intercept", main=
"")
```

and normality of random slopes:

```
hist(unlist(ranef(rikz.lme2)$'NAP'), xlab= "Random NAP effect", main= "").
```

Compare these to previous model. Is there an improvement?

13. Produce a summary of the random effects and fixed effects using the `summary` function. Compare these to the random intercept only model.

14. Use the `anova` function to compare between the two mixed effects models. Which model is preferred by which method? What is the null hypothesis you are testing? Why might you need to treat this test with caution?

15. Using the model `rikz.lme` structure as a basis, use `anova()` to test if there is a significant effect of “temperature”, “grainsize” or “exposure”. Check the summary of the best model. What happened with the beach effect? What could that mean? Should we leave the model as it is?

16. What model do you retain as your final model? What conclusions do you draw from it?

End of the Mixed Effects Models exercise.