Advancing in R

Module 8: Mixed Models

Supplementary exercises

In this exercise we will evaluate how the rearing temperature (Temp) affects the wet mass at eclosion (Wet.Mass) of male dung flies that belong to a quantitative genetic breeding design. Each fly comes from one of eight source populations (Population) in the UK, and many of the sampled flies are related, sharing the same lineage on the maternal (Maternal.Fam) or paternal (Paternal.Fam) side. We want to absorb any pseudoreplication that may arise because of relatedness via lineage or population of origin, and we also want to assess if the effect of Temp depends on population (i.e., perhaps if local adaptation to high temperatures has occurred more in some populations than in others). Consequently, in the course of this exercise, we will discover how to specify "crossed" fixed and random effects. We will also exploit some newly available graphical options for mixed models fit using lme4.

1. Open a new project in a new folder, and start a new script. Save it with a .R suffix to make sure that RStudio interprets commands from it correctly.
2. Read the data from the file "KatieFlies4R.csv" (in the course materials) into a new object.
3. Check that the file has loaded correctly, and examine the data structure using str(). You may wish to recode the variables specifying the maternal and paternal family lineages as factors to avoid problems with coding later on.
4. Check the data variables for any noticeable errors. Examine all the variables, and make a note of any concerns about their distributions. You will notice that the file contains information on both males and females, but we are only interested in males for today, so you can create a new object that filters out the females.
5. Plot the response variable of interest (Wet.Mass) as a function of rearing temperature, but do so in a way that allows you to gauge whether the relationship depends on the source population. Can you predict coefficients for a fixed model that includes Temp as its only predictor?
6. Build a simple linear model that regresses Wet.Mass on Temp. Examine the diagnostics and model summary, and compare your coefficient guesses to the values in the summary. Note the df for this model: do you think this value is a reasonable representation of the real sample size?
7. Rather than jump straight into a mixed model, try first building a fixed model that now fits a fixed effect for Population, as well as the interaction between Population and Temp. How many coefficients will this model have? Examine the model diagnostics and then verify your guess by checking the model summary.
8. Do you think the interaction is worth keeping? Use an F-test (by calling anova()) to decide. If necessary, continue simplifying until you have a minimal adequate model. Do you think that this model acceptably deals with the pseudoreplication issues? Why or why not?
9. Let's incrementally increase the complexity of the model by adding random effects, but let's first do this while keeping Population as a fixed factor. Try the following syntax:
```
Mmixed.mod1<-lmer(Wet.Mass~Temp*Population+(1|Maternal.Fam)
+(1|Paternal.Fam),data=MFLIES)
```
10. Try to simplify the model. How do your conclusions for the resulting minimal adequate model compare to those for the last fixed model?
11. Although the previous model arguably deals with all of the pseudoreplication, it remains unsatisfying in one respect: it fits coefficients for every population, when we might prefer to treat population as a random effect. After all, we did not sample those eight locations because we specifically wanted to predict what flies from those location were like, but rather because we wanted to assess population genetic differences in general.

If we can model population as a random effect, we may be better able to generalize our results across other contexts. Try fitting Population as a random effect instead of a fixed effect now (along with random effects for Maternal.Fam and Paternal.Fam). Examine the model again.

12. The previous model is missing a crucial piece: it does not allow the effect of Temp to vary by population. To fix this, we can slightly modify the code for the random effects specification for Population as follows: (Temp|Population). Unlike the previous situation, in which we coded a 1 before the vertical line, now we are telling lme4 that the effect of Temp should be allowed to vary by Population.
13. Examine the fixed effects for both of the previous models, and note the change. Try using a likelihood ratio test to compare the models. Which is preferred? Is this consistent with the conclusions of contrasts between earlier models that included an interaction or did not?
14. Now to illustrate the effects of your final model: until recently, this was a real struggle, but the latest versions of lme4 provide some powerful additions to the predict() function to help us. We will need to provide predict with newdata as usual, but the new data will not only need to include all of the fixed predictors, but also the random effects as well. One could argued about the best combination of values to use, but in the code below I create a very long data frame that includes all possible combinations for all maternal lineages, paternal lineages, populations, and a sequence of ten temperature values:

```
NEWTEMP<-
expand.grid(Temp=seq(18,22,length=10),Population=levels(MFLIES$P
opulation), Maternal.Fam=levels(MFLIES$Maternal.Fam),
Paternal.Fam=levels(MFLIES$Paternal.Fam))
```

15. I can feed this newdata object to predict, but because I'm most interested in how the effect of temperature varies by population (and not super interested in what happens across maternal and paternal families), I can add an argument that tells predict to focus only on the effects across a subset of the residual effects structure:

```
PREDMASS<-predict(Mmixed.mod4,newdata=NEWTEMP,
re.form=~(Temp|Population))
```

16. Finally, you may want to add the lines illustrating these predictions to a plot showing the data across your populations, either using graphics or ggplot2. Spend a bit of time thinking about what this plot implies about the biology, and how it supports the conclusions you drew from your models.
17. Make sure you have adequately annotated your script, and then save it and your figure.