

Stats with Sparrows - 13

Julia Schroeder

Repeatability with linear mixed models, and beyond

I have promised you in the last hand out that there is an easier method to calculate repeatability - and that's with linear mixed models. These also work much better with unbalanced datasets - no need to figure out n naught ect. So let's have a look at these - we'll be using Zac's fly dataset for this. We need a new package - called lme4.

```
a<-read.table("Wylde_single.mounted.txt", header=T)
head(a)
```

```
library(lme4)
```

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

We'll investigate the repeatability of Femur length:

```
lmm1<-lmer(Femur_length~1+(1|ID), data=a)
```

Before we look at the summary statistics, we will take a closer look at the code:

The first bit makes sense - lmer is the function (instead of lm for a linear model). Then we have the response variable left of the tilde (~). Then there's a 1 - that's odd. You need to know here that if you do not want to fit a fixed effect, just estimate the intercept and nothing else - you write 1. It's something you need to know - that's how this is coded. I could have just as well put sex there or any other fixed effect - but because I want to get the repeatability without any other influence - I do not code a fixed effect. Then, we see something new - (1|ID). That is how in lmer we specify random effects - those that we want the model to use to partition the variance into. So here, we say - partition the variance into variance explained among ID, and residuals (we don't need to say we also want residual variance). Let's have a look at the summary statistics:

```
summary(lmm1)

## Linear mixed model fit by REML ['lmerMod']
## Formula: Femur_length ~ 1 + (1 | ID)
## Data: a
##
## REML criterion at convergence: -247.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.09822 -0.38745 -0.00387  0.36983  3.08564
##
## Random effects:
```

```
## Groups   Name            Variance Std.Dev.
## ID       (Intercept) 1.2570822 1.12120
## Residual                0.0003399 0.01844
## Number of obs: 360, groups: ID, 180
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  4.64978    0.08357   55.64
```

Let's examine the output. First, we're reminded of the model structure - only intercept fitted (1), and then ID as a random effect. The data is a. Then we get a REML convergence criterion - we'll ignore this for now. Then there are the residuals - and some information about their spread.

Next are the random effects - this is new. But it's fairly easy to interpret - we get the variance explained by among ID groups, and the residual variance. We will use these data later to calculate the repeatability. For some reason that I still fail to understand, lmer also provides us with the standard deviation - we could just as well calculate this from the variance so it's a bit of a duplicate report.

We then get information about the data structure - and we can see there are 180 individuals, each measured twice give us a total sample size of 360. So this dataset is really balanced - and we could actually have used ANOVA anyways.

Next, we get information about the fixed effects - this should be familiar to you. Because we did only fit the intercept (1), this parameter estimate is the mean of the response variable.

Now let's calculate the repeatability. Remember - it's the among group variance explained, divided by the total variance. And the total variance is the among group variance plus the residual (or within-group) variance. :

```
Repeatability<-1.257/(1.257+0.0003)
Repeatability
## [1] 0.9997614
```

Excercises:

In the fly dataset, the repeatability is about measurement error - observers would measure the same thing twice to see how accurate they measure. However, we can also use repeatability to learn more about biology. The repeatability of a trait in individuals who are captured at multiple points during their lives, tells us something about the phenotypic flexibility of these traits. For instance, if we'd capture a sparrow multiple times during their life, we would expect their body mass to change quite a bit, but a trait that doesn't change across time, like tarsus length, shouldn't change much. However, in both cases, we'd still expect some consistency within individuals, and larger differences between individuals (as we've seen in the last handout). (1) Use the linear mixed model method to calculate repeatability of tarsus, wing and mass in house sparrows individuals, and think about biological reasons for why the estimates differ. (2) Random effects in linear mixed models

are not only useful for estimating the variance components. In addition to getting that information, modelling a factor that is repeated in the data also accounts for something that is called “pseudo-replication”. The problem with having a repeated-measures dataset like the sparrows is that we often enter the same individual more than once into a model. This can be problematic for similar reasons as the ANOVA gets off with unbalanced datasets - think about it - imagine a dataset of ten individuals, but where one individual is measured 100 times, and the others each only twice. The one individual will inform the model overly much, and that would not give the correct result. If, however, we account for the variance within individuals by modelling ID as random effect, we account for that bias, and the resulting fixed effects will be reliable. So, that’s another reason to know your data structure - if there’s pseudo-replication in your dataset in a grouping factor (e.g. by plots, individuals, observers, years ..) it’s a good idea to model that as random effect. Your task now is to play with this information - use the sparrow dataset and re-run models that you’ve ran before, but this time, model ID as random effect. Compare the results between “simple” linear models without the random effect, and those with. (3) Continue your exploration from (2), but this time, add a second random effect - use also year as a factor on the random part of the model.