Mixed Effect ANOVAs

Here we work with data from a large, repeated-measures experiment using whole wetlands as units in a 3-way factorial design, where treatments were:

- either intensive land use or seminatural lands,
- cattle were excluded with fence (or not), and
- prescribed fire applied or not (Boughton et al. 2015).

2006 was a pasture-only year. Fences were installed in winter 2006-2007, and then fire was applied in winter 2007-2008 and periodically since. So we expect changes through time.

The goal here is to evaluate the interactive effects of these common ranch practices on plant diversity (here $D = e^{H'}$ - see Jost (2006) for more details on how D is calculated). The plantdata06-15.csv data set lists species codes for each wetland (called ponds in the data), from 2006 (when only pasture treatments had been applied) through 2015. We want to calculate diversity (D). We can do that using the `vegan` package:

1. Import and attach the **`plantdata0615.txt`** file. Below I assume you called it "data."
2. And load these packages to use (if not already installed, first do that):

```
library(dplyr) # to make selecting subsets easy
library(lme4) # to run glm and glmm
library(bbmle) # for AICs
library(MuMIn) # for pseudo-R-squared values of mixed effect models
library(vegan) # for lots of ecology-related calculations
library(MASS) # for glm.nb
library(lmerTest) # for follow-up analyses of lmer output
```

3. Now we calculate D, using the diversity command in `vegan`. If you read Help or Package tabs, you will see that it expects a simple data matrix, where unlabelled rows (sites) are matched with columns containing species' counts. So we must be a little clever:

```
ncol(data) # calculate the number of columns in data = 203
# we know from looking at data that non-species info is in columns 1-6. So we
calculate diversity on the other columns.
div <- diversity(data[7:203], index="shannon")

# div represents Shannon diversity. We want that as e raised to that power.
D <- exp(div) # this represents number of species expected in another sample
```

4. And because we want to re-connect the info in columns 1-6 with D, we now bind D to the info (leaving all the species aside) and then make factors:

```
info <- data[1:6] # makes a data frame called info from cols 1-6 in data
Ddata <- cbind(info, D) # sticks together info and D in a new dataframe
Ddata$fyear <- factor(Ddata$year) # to make year a factor in Ddata
Ddata$fblock <- factor(Ddata$block) # to make block a factor in Ddata
Ddata$ffence <- factor(Ddata$fence) # to make fence a factor in Ddata
Ddata$fburn <- factor(Ddata$burn) # to make burn a factor in Ddata
```

5. And because we will want to use initial (year 2006) conditions as a covariate (i.e., wetlands already varied) in analyses, we make a column:

```
yr2006 <- filter(Ddata, fyear == "2006")
D06 <- yr2006$D # this matches length of any one year
```

6. Now let's run a simple ANOVA using **`glm`**, where year2006 D values are a covariate and blocks were handled as *fixed effects*. We only do this to compare to blocks as random effects later. We also use glm to test distribution assumptions, and for simplicity we focus on just 2015 (seven years after all treatments had been applied).

```
fmodelgau <- glm(D ~ pasture*ffence*fburn + D06 + fblock, data=yr2015,
family=gaussian)
fmodelgam <- glm(D ~ pasture*ffence*fburn + D06 + fblock, data=yr2015,
family=Gamma)
fmodelp <- glm(D ~ pasture*ffence*fburn + D06 + fblock, data=yr2015,
family=poisson)
fmodeln <- glm.nb(D ~ pasture*ffence*fburn + D06 + fblock, data=yr2015)
```

7. Which distribution assumption most plausibly fits data? And what pops out as a result?

```
AICctab(fmodelgau,fmodelgam,fmodelp,fmodeln, base=T,delta=T,weights=T)
summary(YOU FILL IN THE MODEL)
```

8. And how much variation did that model "catch"? Try this:

```
r.squaredGLMM(YOU FILL IN THE MODEL) # from MuMIn; fixed & fixed+random “R^2”
```

9. Now we handle blocks as a ***random effect****, rather than a fixed effect*:

```
rmodelgau <- lmer(D ~ pasture*ffence*fburn + D06 + (1|fblock), data=yr2015) #
which runs a gaussian model
rmodelgam <- glmer(D ~ pasture*ffence*fburn + D06 + (1|fblock), data=yr2015,
family=Gamma)
rmodelp <- glmer(D ~ pasture*ffence*fburn + D06 + (1|fblock), data=yr2015,
family=poisson)
rmodeln <- glmer.nb(D ~ pasture*ffence*fburn + D06 + (1|fblock), data=yr2015)
```

Did treating blocks as random effects improve the "story"?

10. The above could be done for each year. What about an overall effect, **across all the years**? **For that repeated measures analysis, time can be treated as a random effect.** So now edit code above to include years as a second random factor, and use the whole data set instead:

```
D06rep <- rep(D06, 9) # to make a D06 for all-years analyses (9 years)
```

And re-run your choice of models above but using the following model terms

```
... pasture*ffence*fburn + D06rep + (1|fblock) + (1|fyear), data=Ddata
```

What are the overall experimental results? How important are the random effects?

11. And how much variation did the model "catch"? Use `r.squaredGLMM` again.

    And how are the residuals distributed for your fav model?

```
plot(YOU FILL IN THE MODEL)
```

Also try this command for more information (e.g., data for graphs), via the lmerTest package:

```
anova(YOU FILL IN THE MODEL) # to get F ratios, akin to classic ANOVAs
```

12. Now let's consider what would happen if we mistakenly considered each year as an independent replicate (i.e., temporal pseudoreplication), blindly used a plain ol' lm (family="gaussian"), and didn't think of blocks.

```
wrongmodel <- lm(D ~ pasture*ffence*fburn, data=Ddata)
summary(wrongmodel)
```

Compare coefficients (effect sizes) and their std. errors among the "correct" and wrong models:

- What about residual variance and R^2 (leftover, unexplained variation)?
- Also compare models for AICcs – which most parsimoniously explains the data, even if it was a bit more complicated to construct?
- Would our paper on effects of the experiment have been representative of data?

Bottom Lines: The model that best captures variance best represents effects. Mixed effect models can represent a lot of [unplanned, unessential] variation that fixed effect models cannot, and so better tease out the story. ***Expect to use mixed effect models in your research!***

References

- Boughton EH, et al. 2015. Interactive effects of pasture management intensity, release from grazing, and prescribed fire on forty subtropical wetland plant assemblages. J. Applied Ecology 53:159-170.
- Jost, L. 2006. Entropy and diversity. Oikos 113:363-375.