

Virginia Anne Nichols

From: Maarit Mäenpää
Sent: 7. maj 2024 12:28
To: Virginia Anne Nichols
Cc: Bo Melander
Subject: Resources for the GLMM's + code templates
Attachments: Brooks et al. 2017.pdf

Hi Gina,

Sorry – slightly long email, but this is to give you a few resources and examples to get started with. If you apply these on the different parts of your data, we can later meet again to discuss the next steps, and to troubleshoot, should you wish to do so. Let me know how it goes – I am happy to keep helping with this!

Model formula

For the model formula, this sounds like the current best candidate. Of course, add or remove the fixed factors that may be needed for the analysis – should early/late establishment also be included?

```
response ~ Tillage * CC * Straw * Year + (1|Block) +  
(1|Subplot:CC:Tillage:Straw) + (1|CC:Tillage:Straw) + (1|Tillage:Straw) +  
(1|Straw)
```

With the glm, you can fit the model with lme4:

```
model <- glmer(response ~ Tillage * CC * Straw * Year + (1|Block) +  
(1|Subplot:CC:Tillage:Straw) + (1|CC:Tillage:Straw) + (1|Tillage:Straw) +  
(1|Straw), family=XXX(link="YYY"), data=DATA)
```

And with glmmTMB (often more flexible). See attached for a paper describing the package – there are some good examples.

```
model <- glmmTMB(response ~ Tillage * CC * Straw * Year + (1|Block) +  
(1|Subplot:CC:Tillage:Straw) + (1|CC:Tillage:Straw) + (1|Tillage:Straw) +  
(1|Straw), family=XXX(link="YYY"), data=DATA)
```

Sidenote: You don't always need all interactions: In general I recommend you only add the interactions that you have a hypothesis or prediction for to the model to begin with. Here, you had illustrated the data to show them all, so it is probably necessary, but in the future, remember you can fit based on expectations.

Probability distributions

COUNTS: For the family (=probability distribution), you can use Poisson, if you don't have overdispersion. Poisson distribution has strict assumptions for shape: the mean of the distribution should be the same as the variance. When variance exceeds the mean, you have overdispersion. To check, find your residual deviance and residual df in summary() output: deviance/df should be ~ 1, otherwise you have over- (or under-) dispersion, and Poisson is not a good fit. In that case, you can switch to a model using negative binomial distribution instead. Preferred link-function for both: log

PROPORTIONS: Binomial or beta-family distributions. If you have differing sample sizes that created the proportions, you need to also apply weights=CASES into the model formula – but this doesn't seem to be the case here. Preferred link-function, in most cases: logit

For more on the different family options, specifically in glmmTMB, see page 28: <https://cran.r-project.org/web/packages/glmmTMB/glmmTMB.pdf>

4-way proportion model

For the 4-way proportions for cover, you could actually analyse this a little differently to what we talked about as well. You could pivot your data into long-format that has a column for the proportions, and another column for the Cover category (Soil, Weeds, Volunteers and CC). If you then analyse the proportions with a model like this:

```
model <- glmmTMB(proportion ~ CoverCategory * Tillage * CC * Straw * Year +  
(1|Block) + (1|CC:Tillage:Straw) + (1|Tillage:Straw) + (1|Straw),  
family=binomial(link="logit"), data=DATA)
```

You get an estimate of how different the proportions of each cover category are, and you are checking whether that is dependent on tillage, CC, straw or year effects.

Checking assumptions

Always check model assumptions! When checking the residual plots for glm's that don't assume normality, or especially for non-continuous data, it is easier to check the model residuals using package DHARMA. DHARMA simulates residuals from the model structure, and tests the real data against that, so the model assumption checking is flexible to all kinds of models. You can plot the assumption check using:

```
plot(simulateResiduals(mod))
```

Interpretation is very similar to residual plots from linear models, but if you want to know more in detail, see here: <https://cran.r-project.org/web/packages/DHARMA/vignettes/DHARMA.html>

Hope that is useful!

Best,
Maarit

Maarit Mäenpää
Statistician

Dept. of Agroecology
Aarhus University
Blichers Allé 20, Postboks 50
DK-8830 Tjele

Phone:

E-mail: m.maenpaa@agro.au.dk

Web: agro.au.dk



"Please note that this email may contain personal data. You must ensure that this data cannot be accessed by anyone else without good reason, and that it is deleted immediately when it is no longer required in relation to the purpose for which it was sent".