

Generalized linear mixed models - Solutions

Data

For this laboratory we will use the Portal database, already presented in Lab 2, which contains long-term monitoring data for several rodent species at a study site in Arizona.

Ernest, M., Brown, J., Valone, T. and White, E.P. (2018) *Portal Project Teaching Database*.
https://figshare.com/articles/Portal_Project_Teaching_Database/1314459.

This database consists of three data tables:

- portal_surveys.csv contains information on each captured individual.

```
surveys <- read.csv("../donnees/portal_surveys.csv")
str(surveys)
```

```
## 'data.frame': 35549 obs. of 9 variables:
## $ record_id : int 1 2 3 4 5 6 7 8 9 10 ...
## $ month : int 7 7 7 7 7 7 7 7 7 7 ...
## $ day : int 16 16 16 16 16 16 16 16 16 16 ...
## $ year : int 1977 1977 1977 1977 1977 1977 1977 1977 1977 1977 ...
## $ plot_id : int 2 3 2 7 3 1 2 1 1 6 ...
## $ species_id : chr "NL" "NL" "DM" "DM" ...
## $ sex : chr "M" "M" "F" "M" ...
## $ hindfoot_length: int 32 33 37 36 35 14 NA 37 34 20 ...
## $ weight : int NA NA NA NA NA NA NA NA NA NA ...
```

- portal_species.csv gives the genus, species and taxonomic group corresponding to each species ID.

```
species <- read.csv("../donnees/portal_species.csv")
str(species)
```

```
## 'data.frame': 54 obs. of 4 variables:
## $ species_id: chr "AB" "AH" "AS" "BA" ...
## $ genus : chr "Amphispiza" "Ammospermophilus" "Ammodramus" "Baiomys" ...
## $ species : chr "bilineata" "harrisi" "savannarum" "taylori" ...
## $ taxa : chr "Bird" "Rodent" "Bird" "Rodent" ...
```

- portal_plots.csv indicates the type of treatment applied to each plot: “Control” = control, no fence; “Rodent Exclosure” = fence to exclude all rodents; “Long-term Krat Exclosure” or “Short-term Krat Exclosure” = fence with gate to exclude kangaroo rats of the genus *Dipodomys*. “Spectab exclosure” = fence with gate to exclude only the species *Dipodomys spectabilis*.

```
plots <- read.csv("../donnees/portal_plots.csv")
str(plots)
```

```
## 'data.frame': 24 obs. of 2 variables:
## $ plot_id : int 1 2 3 4 5 6 7 8 9 10 ...
## $ plot_type: chr "Spectab exclosure" "Control" "Long-term Krat Exclosure" "Control" ...
```

1. Preparation of data

- From the `surveys` table, only keep individuals observed since 1988 that correspond to rodents (`taxa == "Rodent"` in the `species` table).
- Select the 15 most abundant species from the data frame obtained in (a), then count the number of individuals for each of these species, by year and plot. Include a count of 0 for plots and years where the species is absent.
- Recode the two treatments “Long-term Krat Exclosure” and “Short-term Krat Exclosure” into one treatment, “Krat Exclosure”, and join the `plot` data frame to the data frame obtained in (b).

```
plots$plot_type[grepl("Krat", plots$plot_type)] <- "Krat Exclosure"
```

Note: The function `grepl(pattern, x)` returns `TRUE` or `FALSE` depending on whether or not the variable `x` contains the given text `pattern`.

2. Modeling the abundance of a species

For this part, we limit ourselves to the data of the species *Dipodomys ordii* (DO), Ord’s kangaroo rat.



- Visualize the distribution of the number of individuals of species DO according to the type of treatment. What type of model would be appropriate to model this relationship? (Ignore the effects of other variables at this point.) Estimate the model parameters and check whether the resulting coefficients qualitatively match the expected effects of each treatment.
- Now add to the model in (a) random effects of plot and year. Which of these two grouping variables has the greatest effect on the response? Explain how and why the standard errors of the fixed effects changed between this model and the model in the previous question.
- Check that the model assumptions in (b) are met, especially the dispersion of residuals and the normality of random effects. If overdispersion occurs, estimate the coefficient of dispersion.

3. Modeling several species

Now let’s take the complete dataset produced in Part 1, the 15 most abundant species.

- Use a model without random effects to estimate abundance by species, type of treatment and their interaction. Based on the description of the experiment, why is it important to include the interaction here?
- Fit a model equivalent to (a), except that the species is a random rather than a fixed effect. Name one advantage and one disadvantage of this choice.

Note: If the GLMM is having trouble converging, we can specify the `control` argument to `glmer` to increase the maximum number of iterations or to change the optimizer. In this case, changing the optimizer to `bobyqa` with `control = glmerControl(optimizer = "bobyqa")` should fix the problem.

- c) With the `expand.grid` function, create a data set for predictions that contains all combinations of treatment and species.

```
pred_df <- expand.grid(plot_type = unique(abond$plot_type),  
                      species_id = unique(abond$species_id))
```

Calculate for each combination the predicted abundance according to the models in (a) and (b), using the `predict` function. Visualize the predicted values. Is there shrinkage of the estimates for the mixed model?

Note: By default, the predictions of a GLM(M) are on the scale of the linear predictor. For predictions on the scale of the response, specify the argument `type = "response"`.

- d) Add the random effects of plot and year on the model in (b) and verify the goodness of fit.
- e) Finally, use a strategy similar to (c) to obtain predictions of the mean abundance of each species in the control plots as a function of year.

Note: By default, `predict` takes into account all random effects. To consider only some effects but not others, specify the `re.form` argument to `predict`. For example, `re.form = ~(1 | year)` considers the effect of the year only. To ignore all random effects in predictions, write `re.form = ~0`.