

Modèles linéaires généralisés à effets mixtes

Données

Nous utiliserons pour ce laboratoire la base de données Portal, déjà présentée lors du laboratoire 2, qui contient des données de suivi à long terme de plusieurs espèces de rongeurs sur un site d'étude en 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.

Cette base de données est formée de trois tableaux de données:

- portal_surveys.csv contient les informations sur chaque individu capturé.

```
surveys <- read.csv("../donnees/portal_surveys.csv", stringsAsFactors = FALSE)
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 indique le nom de genre, d'espèce et le groupe taxonomique correspondant à chaque code d'espèce.

```
species <- read.csv("../donnees/portal_species.csv", stringsAsFactors = FALSE)
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 indique le type de traitement appliqué à chaque parcelle: “Control” = témoin; “Rodent Exclosure” = clôture pour exclure tous les rongeurs; “Long-term Krat Exclosure” ou “Short-term Krat Exclosure” = clôture avec porte pour exclure les rats-kangourous du genre *Dipodomys*. “Spectab exclosure” = clôture avec porte pour exclure seulement l'espèce *Dipodomys spectabilis*.

```
plots <- read.csv("../donnees/portal_plots.csv", stringsAsFactors = FALSE)
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. Préparation des données

- À partir du tableau `surveys`, conservez les individus observés depuis 1988 qui correspondent à des rongeurs (`taxa == "Rodent"` dans le tableau `species`).
- Choisissez les 15 espèces les plus abondantes dans le tableau de données obtenu en (a), puis comptez le nombre d'individus pour chacune de ces espèces, par année et par placette. Incluez un compte de 0 pour les placette et années où l'espèce est absente.
- Recodez les deux traitements “Long-term Krat Exclosure” et “Short-term Krat Exclosure” en un seul traitement, “Krat Exclosure”, puis joignez le tableau de données `plots` au tableau obtenu en (b).

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

Note: La fonction `grepl(pattern, x)` retourne `TRUE` ou `FALSE` dépendamment si la variable `x` contient ou non le motif de texte `pattern`.

2. Modéliser l'abondance d'une espèce

Pour cette partie, nous nous limitons aux données de l'espèce *Dipodomys ordii* (DO), le rat-kangourou d'Ord.



- Visualisez la distribution du nombre d'individus de l'espèce DO en fonction du type de traitement. Quel type de modèle serait approprié pour modéliser cette relation? (Ignorez les effets des autres variables pour l'instant.) Estimez les paramètres du modèle et vérifiez si les coefficients obtenus correspondent qualitativement aux effets prévus de chaque traitement.
- Ajoutez maintenant au modèle en (a) des effets aléatoires de la parcelle et de l'année. Laquelle de ces deux variables de groupement a-t-elle le plus grand effet sur la réponse? Expliquez comment et pourquoi les erreurs-types des effets fixes changent entre ce modèle et celui de la question précédente.
- Vérifiez si les suppositions du modèle en (b) sont respectées, en particulierité la dispersion des résidus et la normalité des effets aléatoires. S'il y a surdispersion, estimez le coefficient de dispersion.

3. Modéliser plusieurs espèces

Prenons maintenant le jeu de données complet produit dans la partie 1, soit les 15 espèces les plus abondantes.

- Utilisez un modèle sans effets aléatoires pour estimer l'abondance selon l'espèce, le type de traitement et leur interaction. Selon la description de l'expérience, pourquoi est-il important d'inclure l'interaction ici?
- Ajustez un modèle équivalent à celui en (a), sauf que l'espèce est un effet aléatoire plutôt que fixe. Nommez un avantage et un désavantage de ce choix.

Note: Si le GLMM a du mal à converger, nous pouvons spécifier l'argument `control` de `glmer` pour augmenter le nombre maximal d'itérations ou pour changer d'optimisateur. Dans ce cas-ci, changer l'optimisateur à `bobyqa` avec `control = glmerControl(optimizer = "bobyqa")` devrait régler le problème.

- c) Avec la fonction `expand.grid`, créez un jeu de données pour les prédictions qui contient toutes les combinaisons de traitement et d'espèce.

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

Calculez pour chaque combinaison l'abondance prévue selon chacun des modèles en (a) et (b), à l'aide de la fonction `predict`. Visualisez les valeurs attendues. Y a-t-il une contraction des estimés dans le cas du modèle mixte?

Note: Par défaut, les prédictions d'un GLM(M) sont sur l'échelle du prédicteur linéaire. Pour des prédictions sur l'échelle de la réponse, spécifiez l'argument `type = "response"`.

- d) Ajoutez les effets aléatoires de la placette et de l'année au modèle en (b) et vérifiez l'ajustement du modèle.
- e) Finalement, utilisez une stratégie semblable à (c) pour obtenir des prédictions de l'abondance moyenne de chaque espèce dans les placettes témoin en fonction de l'année.

Note: Par défaut, `predict` tient compte de tous les effets aléatoires. Pour seulement considérer certains effets mais pas les autres, il faut spécifier l'argument `re.form` de `predict`. Par exemple, `re.form = ~(1 | year)` considère l'effet de l'année seulement. Pour ignorer tous les effets aléatoires dans les prédictions, écrivez `re.form = ~0`.