Méthodologie

Présentation des données

Les données que nous avons utilisées nous proviennent de ...

Motivation

Les modèles linéaires généralisés à effets mixtes (GLMM) combinent :

- Les caractéristiques des modèles linéaires généralisés (GLM) pour modéliser des variables non-normalement distribuées.
- Les propriétés des modèles à effets mixtes pour gérer des données groupées ou hiérarchiques.

Modèles Linéaires Généralisés

Un GLM relie le prédicteur linéaire η à la moyenne μ de la réponse à travers une fonction de lien g :

$$g(\mu) = \eta = \beta_0 + \sum_{i=1}^{m} \beta_i x_i$$

Les distributions possibles incluent :

- **Normale** : Régression linéaire classique, avec lien identité.
- **Binomiale** : Régression logistique pour données binaires, avec lien logit.
- **Poisson** : Régression de Poisson pour données de comptage, avec lien logarithmique.

Régression Logistique

Modélise une réponse binaire ($y \sim B(n, p)$), où p est la probabilité de succès :

$$P(y|n,p) = \binom{n}{y} p^y (1-p)^{n-y}$$

La probabilité p est reliée au prédicteur par la fonction logistique :

$$p = \frac{1}{1+e^{-\eta}} \quad \text{où} \quad \eta = \beta_0 + \sum_{i=1}^{m} \beta_i x_i.$$

Le log-vraisemblance est exprimé comme :

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i \log p_i + (1-y_i) \log (1-p_i) \right]$$

Régression de Poisson

Utilisée pour modéliser des données de comptage ($y \sim Pois(\lambda)$), où λ est la moyenne et la variance :

$$P(y|\lambda) = \frac{\lambda^y}{y!} e^{-\lambda}$$

Le lien logarithmique assure $\lambda > 0$:

$$\log \lambda = \beta_0 + \sum_{i=1}^{m} \beta_i x_i$$

L'espérance est $E[y] = \lambda$.

Modèles Linéaires Mixtes

Ces modèles ajoutent des termes d'effets aléatoires $\mathbf{Zu}$ au prédicteur linéaire :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \boldsymbol{\varepsilon},$$

avec :

- $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$, les effets aléatoires.
- $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{R})$, les résidus.

La matrice de covariance totale est :

$$\mathrm{Var}(\mathbf{y}) = \mathbf{ZGZ}^T + \mathbf{R}.$$

Les paramètres sont estimés par maximum de vraisemblance (ML) ou par vraisemblance restreinte (REML).

Modèles Linéaires Généralisés à Effets Mixtes (GLMM)

Un GLMM étend les GLM en intégrant des effets aléatoires :

$$g(\mu) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu},$$

où :

- $g(\cdot)$ est la fonction de lien.
- $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$ est le vecteur d'effets aléatoires.

Les paramètres sont estimés via des méthodes comme :

- Approximations Laplaciennes.
- Quadrature gaussienne adaptative.
- Méthodes MCMC (chaînes de Markov Monte Carlo).

Prédictions et Simulations

Les GLMM permettent deux types de prédictions :

- **Conditionnelles** : Basées sur les effets aléatoires spécifiques ($\mathbf{u}$).
- **Marginales** : En intégrant sur les effets aléatoires.

Les simulations utilisent des approches paramétriques pour évaluer la variabilité et tester les hypothèses. Une approche courante est le bootstrap paramétrique :

1. Générer des données simulées basées sur les paramètres estimés.
2. Réajuster le modèle pour chaque jeu de données simulé.
3. Analyser la distribution des estimations obtenues.