Université de Montpellier
Département de Mathématiques Appliquées

Outstanding 2 : Modèle mixte

Atelier Projet — HAX916X

Réalisé par :
ATTOUMANI Ibrahim
DIALLO Ousmane

Année Universitaire 2025 – 2026

Table des matières

1. Introduction

Dans ce cours, nous introduirons la notion d'effets aléatoires, mais, avant de définir formellement le modèle mixte, il est essentiel de comprendre le contexte et les situations dans lesquelles les effets aléatoires apparaissent.

Nous analysons la relation entre les notes des étudiants et les heures d'étude, en tenant compte de l'effet de la classe.

On note:

- `N`: Note
- `H`: Heures d'étude
- `C`: Classes (`Effet aléatoire`)

Le modèle s'écrit de la façon suivante:

$$N \sim H + C \quad (i)$$

Ici, les heures d'étude (H) ne sont pas significatives, puisque les notes des étudiants dépendent principalement de la classe, pas des heures d'étude. Ainsi, le modèle (ii) ci-dessous:

$$N \sim C \quad (ii)$$

est aussi performant que le modèle (i).

À présent, nous allons modéliser les notes, la classe en fonction des heures d'études:

$$H \sim N + C \quad (iii)$$

Les notes sont significatives, étant donné que les heures d'étude des étudiants sont influencées par les notes.

Le modèle où l'on considère seulement l'effet de la classe:

$$H \sim C \quad (iv)$$

est moins performant que le modèle (iii) car les notes améliorent la prédiction des heures d'étude.

En guise de conclusion, les notes influencent les heures d'étude, mais les heures d'étude n'ont pas un impact significatif sur les notes. Les modèles mixtes révèlent ces dynamiques en tenant compte des variations entre les classes.

2. Modèle mixte

Un **modèle mixte** est un modèle statistique combinant à la fois des facteurs à **effets fixes** (comme dans une **ANOVA**) et des facteurs à **effets aléatoires**.

Définition 1 : modèle mixte à un facteur

Considérons l'effet d'un facteur A sur une variable Y, en supposant que les niveaux de A sont aléatoires. On note $i = 1, \dots, n$ les niveaux de A et $k = 1, \dots, n_i$ les répétitions. Le modèle s'écrit alors :

$$Y_{ik} = \mu + \alpha_i + \epsilon_{ik}$$

avec :

- μ : moyenne générale
- $\alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)$: effet aléatoire du i-ème niveau du facteur A
- $\epsilon_{ik} \sim \mathcal{N}(0, \sigma^2)$: erreur résiduelle indépendante
- α_i et ϵ_{ik} sont indépendants

Définition 2 : modèle mixte à deux facteurs

Si l'on considère deux facteurs A et B, tous deux aléatoires, le modèle mixte s'écrit :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

où :

- $\alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)$: effet aléatoire du facteur A
- $\beta_j \sim \mathcal{N}(0, \sigma_\beta^2)$: effet aléatoire du facteur B
- $(\alpha\beta)_{ij} \sim \mathcal{N}(0, \sigma_{\alpha\beta}^2)$: interaction aléatoire
- $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$: erreur résiduelle
- Tous les effets aléatoires et ϵ_{ijk} sont indépendants

2.1. Application directe : modèle mixte

Ici, nous allons illustrer l'utilisation d'un **modèle mixte** à partir du jeu de données **sleepstudy** du package lme4.

Ces données contiennent des mesures de **temps de réaction** de sujets sur plusieurs jours de privation de sommeil, avec :

- **Reaction** : temps de réaction en millisecondes
- **Days** : nombre de jours de privation de sommeil
- **Subject** : identifiant du sujet (facteur aléatoire)

```r
# chargement des donnees
data(sleepstudy)
```

```r
# Ajuster un modele mixte
mod <- lmer(Reaction ~ Days + (1 | Subject), data = sleepstudy)
summary(mod)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Reaction ~ Days + (1 | Subject)
##    Data: sleepstudy
##
## REML criterion at convergence: 1786.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2257 -0.5529  0.0109  0.5188  4.2506
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Subject  (Intercept) 1378.2   37.12
##  Residual              960.5   30.99
## Number of obs: 180, groups:  Subject, 18
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 251.4051     9.7467   25.79
## Days         10.4673     0.8042   13.02
##
## Correlation of Fixed Effects:
##      (Intr)
## Days -0.371
```

La sortie du modèle mixte ajusté avec la fonction `lmer` de R fournit une analyse détaillée des effets fixes et aléatoires dans les données `sleepstudy`. Ce modèle permet d'expliquer la variabilité de la réaction (`Reaction`) selon le nombre de jours (`Days`), en prenant en compte l'effet aléatoire des sujets (`Subject`).

Le modèle s'écrit :

$$\text{Reaction}_{ij} = \beta_0 + \beta_1 \cdot \text{Days}_{ij} + \text{Subject}_{0i} + \epsilon_{ij}$$

avec :

- $\beta_0 = 251.41$: intercept moyen (réaction au jour 0)
- $\beta_1 = 10.47$: hausse moyenne du temps de réaction par jour de privation
- $\text{Subject}_{0i} \sim \mathcal{N}(0, 1378.2)$: effet aléatoire spécifique au sujet (intercept)
- $\epsilon_{ij} \sim \mathcal{N}(0, 960.5)$: résidu

Chaque jour de privation de sommeil augmente le temps de réaction en moyenne de 10.47 unités ($Pr(>|t|) < 0.001$), en tenant compte d'une variation importante entre individus ($\sigma_{\text{sujet}} = 37.12$). La variation résiduelle intra-individuelle est de $\sigma = 30.99$.

Remarque : Calcul de la p-value à partir de la t-value

On peut calculer la p-value à partir de la t-value en utilisant la fonction de répartition cumulative de la distribution de Student:

$$\mathbb{P}(T_{df} \leq t) = \int_{-\infty}^{t} f(x)dx \quad \text{(où f est la densité d'une Student à df degrés de libertés).}$$

On peut le calculer sous R de la façon suivante:

$$2 * pt(-abs(t), df)$$

où:
- t est la t-value - df est le degré de liberté $(n-k)$, avec n le nombre d'individus et k le nombre paramètres éstimés