S8_B3_Effets fixes_aléatoires

Table of contents

---

1 Contexte : pourquoi parler d'effets fixes et aléatoires ?

On travaille avec des données corrélées :

- patients vus plusieurs fois (données longitudinales),
- patients inclus dans différents centres,
- élèves dans des classes, classes dans des écoles, etc.

Schéma classique :

- i indexe les individus (patients) à l'intérieur d'un groupe,
- j indexe les groupes (centres, services, écoles).

On observe une réponse Y_{ij} (par exemple un score, une mesure biologique) et des covariables X_{ij} (âge, sexe, traitement, temps, etc.).
Dans un modèle linéaire simple (sans structure de corrélation explicite) :

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \varepsilon_{ij}.$$

Problème : les observations ne sont pas indépendantes (mêmes patients, mêmes centres, etc.).

Idée des modèles linéaires mixtes (MLM) :

Combiner des effets fixes (covariables dont on veut estimer l'effet) et des effets aléatoires (variables qui modélisent la structure de corrélation entre les observations).

□

2 Effets fixes : covariables dont on veut estimer l'effet

2.A Définition intuitive

Un effet fixe, c'est :

- une covariable pour laquelle on veut estimer un effet moyen dans la population,
- avec un coefficient unique valable pour tous les individus (et pour tous les centres) dans le modèle.

Exemples typiques :

- traitement (A vs B),
- âge, sexe,
- dose d'un médicament,
- temps (en jours depuis l'inclusion).

Dans un modèle mixte, on écrit par exemple :

$$Y_{ij} = \beta_0 + \beta_1, \text{traitement} ij + \beta_2, \text{âge} ij + \cdots + \text{(effets aléatoires)} + \varepsilon_{ij}.$$

Les β_k sont des effets fixes :

- β_1 = effet moyen du traitement dans l'ensemble de la population,
- β_2 = effet moyen de l'âge, etc.

On interprète β_1 comme dans une régression classique :

- si Y_{ij} est continu : différence moyenne de Y associée au traitement, après ajustement,
- si on est dans un modèle logistique ou Poisson : log-odds ratio, log-rate ratio, etc.

2.B Ce que signifie « covariables dont on veut estimer l'effet »

Quand on dit :

Effets fixes : covariables dont on veut estimer l'effet,

on veut dire :

- ce sont les variables pour lesquelles l'effet lui-même est l'objet du résultat scientifique,
- on va rapporter ces effets dans un tableau de résultats (estimateur, intervalle de confiance, p-value),
- on suppose que cet effet est le même dans tous les centres / tous les patients, à l'intérieur du modèle.

Exemple :

- on veut savoir si le nouveau traitement réduit la mortalité, en moyenne,
- on met le traitement en effet fixe,
- on interprète $\hat{\beta}_{\text{traitement}}$ comme l'effet moyen du traitement.

□

3 Effets aléatoires : modéliser la structure de corrélation

3.A Idée de base

Les effets aléatoires servent à modéliser :

- le fait que des observations appartenant au même groupe se ressemblent plus que des observations de groupes différents,
- la variabilité entre groupes (centres, patients, écoles, etc.),
- la corrélation entre observations d'un même groupe.

On écrit par exemple, pour des patients i dans des centres j :

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + u_j + \varepsilon_{ij},$$

avec :

- u_j = effet aléatoire de centre (spécifique au centre j),
- souvent on suppose $u_j \sim \mathcal{N}(0, \tau^2)$,
- $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ = erreur résiduelle individuelle.

Ici :

- β_0 et β_1 sont des effets fixes,
- u_j est un effet aléatoire,
- u_j introduit de la corrélation entre les Y_{ij} du même centre.

3.B « Variables qui modélisent la structure de corrélation »

La phrase :

Effets aléatoires : variables qui modélisent la structure de corrélation entre les observations se traduit par : - on ajoute des termes aléatoires (comme u_j) qui sont partagés par plusieurs observations, - deux observations qui partagent le même u_j sont corrélées, - la manière dont on spécifie ces effets aléatoires (intercepts, pentes aléatoires, structure) détermine la forme de la corrélation.

3.B.1 Exemple: intercept aléatoire de centre $(1 \mid \textbf{centre})$

Modèle :

$$Y_{ij} = \beta_0 + u_j + \beta_1 X_{ij} + \varepsilon_{ij}.$$

- toutes les observations d'un même centre partagent le même u_j,
- cela crée une corrélation intra-centre,
- $\tau^2 = \text{Var}(u_j)$ mesure la variabilité inter-centre.

3.B.2 Exemple: Intercept aléatoire de patient $(1 \mid \textbf{patient})$ dans des données longitudinales

Modèle :

$$Y_{ij} = \beta_0 + b_i + \beta_1 t_{ij} + \varepsilon_{ij},$$

où t_{ij} est le temps et b_i un effet aléatoire patient.

- toutes les mesures du même patient partagent le même b_i,
- cela modélise le fait que chaque patient a un niveau moyen propre,
- $\text{Var}(b_i)$ donne la variabilité inter-patient.

3.B.3 Exemple: Pente aléatoire $(\textbf{temps} \mid \textbf{patient})$

Modèle :

$$Y_{ij} = \beta_0 + b_{0i} + (\beta_1 + b_{1i}) t_{ij} + \varepsilon_{ij}.$$

- chaque patient a son intercept propre b_{0i} et sa pente propre b_{1i},
- cela permet de modéliser des trajectoires individuelles différentes dans le temps.

□

4 Résumé conceptuel : ce qui distingue effets fixes et aléatoires

On peut résumer ainsi :

- Effets fixes
  - On estime des coefficients uniques (par exemple β_1 pour le traitement).
  - On s'intéresse directement à la valeur de ces coefficients.
  - On les interprète comme des effets moyens (différence moyenne, log-odds ratio moyen, etc.).
- Effets aléatoires
  - On modélise des variables latentes comme u_j, b_i, propres aux centres ou aux patients.
  - On ne s'intéresse pas à la valeur spécifique de chaque u_j en tant que résultat principal, mais à :
  - la variance de ces effets (par exemple τ^2),
  - la structure de corrélation qu'ils induisent.
- Ils permettent de dire : "Les centres/patients diffèrent entre eux selon une distribution (par exemple normale) avec telle variance".

Les modèles linéaires mixtes :

- combinent des effets fixes (pour les covariables dont on veut l'effet estimé)
- et des effets aléatoires (pour représenter la structure de corrélation et la variabilité entre groupes).

□

5 Exemple simple en R

On simule des données avec :

- $J = 8$ centres,
- $n_j = 40$ patients par centre,
- un effet fixe x (par exemple une covariable continue),
- un effet aléatoire de centre u_j,
- une erreur individuelle ε_{ij}.

```
J <- 8
n_per_center <- 40
center <- factor(rep(1:J, each = n_per_center))

# Covariable individuelle
x <- rnorm(J * n_per_center, mean = 0, sd = 1)

# Paramètres "vrais"
beta0 <- 2      # effet fixe intercept
beta1 <- 1.5    # effet fixe pour x
tau   <- 1      # écart-type inter-centre
sigma <- 2      # écart-type inter-individuel
```

```
# Effet aléatoire de centre
u <- rnorm(J, mean = 0, sd = tau)
u_center <- u[center]

# Bruit individuel
eps <- rnorm(J * n_per_center, mean = 0, sd = sigma)

# Réponse
y <- beta0 + u_center + beta1 * x + eps

dat <- data.frame(
    y = y,
    x = x,
    center = center
)

head(dat)
```

```
            y           x center
1  4.06775203 -0.56047565      1
2  1.90674422 -0.23017749      1
3  6.61061002  1.55870831      1
4 -0.03367521  0.07050839      1
5  0.98815299  0.12928774      1
6  7.84846318  1.71506499      1
```

On ajuste alors un modèle mixte :

- effet fixe : x,
- effet aléatoire : intercept aléatoire pour le centre.

```
mod <- lmer(y ~ x + (1 | center), data = dat)
summary(mod)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ x + (1 | center)
   Data: dat

REML criterion at convergence: 1363.6

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.85395 -0.64960  0.02815  0.68295  2.71681

Random effects:
 Groups   Name        Variance Std.Dev.
 center   (Intercept) 0.5729   0.7569
```

```
 Residual             3.9426   1.9856
Number of obs: 320, groups:  center, 8

Fixed effects:
            Estimate Std. Error t value
(Intercept)   2.1368     0.2897   7.375
x             1.5949     0.1165  13.692

Correlation of Fixed Effects:
  (Intr)
x -0.012
```

Dans la sortie :

- partie Fixed effects :
  - (Intercept) : estimation de β_0,
  - x : estimation de β_1, effet fixe de la covariable.
- partie Random effects :
  - center : variance estimée de u_j (approximation de τ^2),
  - Residual : variance résiduelle (approximation de σ^2).

On peut extraire les deux types d'effets :

```r
# Effets fixes
fixef(mod)
```

```
(Intercept)           x
   2.136808    1.594914
```

```r
# Effets aléatoires de centre (BLUPs)
ranef(mod)$center
```

```
  (Intercept)
1  0.07338514
2  0.64048205
3 -1.13642609
4  0.81385579
5 -0.59847350
6  0.64985586
7 -0.51545981
8  0.07278056
```

Interprétation :

- fixef(mod) donne les estimations des effets fixes (ce qu'on met dans le tableau de résultats).
- ranef(mod)$center donne, pour chaque centre, l'estimation de u_j (effet aléatoire centre) :

  - ces valeurs traduisent le fait que certains centres sont au-dessus ou en dessous de la moyenne,
  - mais ce n'est pas leur valeur exacte isolée qui est le résultat principal,
- ce qui nous intéresse surtout, c'est la variance de ces effets (la variabilité inter-centre) et la corrélation qu'ils induisent.

□

6 En une phrase

- **Effets fixes** : covariables dont on veut estimer et interpréter l'effet moyen (β), commun à toute la population.
- **Effets aléatoires** : composantes aléatoires (comme u_j, b_i) ajoutées pour représenter la variabilité entre groupes et la corrélation entre observations, via leurs variances (τ^2, etc.).

Les modèles linéaires mixtes sont précisément construits pour combiner les deux, afin de : - obtenir des effets fixes cohérents, - et modéliser correctement la structure de dépendance des données.