S8_B4_Intercept_variances_effets_aleatoires

Table of contents

1 Contexte et objectif

On veut expliquer :

1. ce qu'est un intercept (et une pente) dans une régression,
2. ce que fait un modèle linéaire mixte :
    - deux composantes de variance : entre groupes et à l'intérieur des groupes,
3. ce qu'est une matrice de variance–covariance (sans jargon inutile),
4. la différence entre effets aléatoires emboîtés et effets aléatoires croisés.

□

2 Rappel : droite de régression et « intercept »

2.A Modèle linéaire simple

On commence par un modèle de base avec une seule covariable X :

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i.$$

- Y_i : résultat pour l'individu i (par exemple : tension artérielle).
- X_i : covariable (par exemple : âge).
- β_0 et β_1 : paramètres du modèle.
- ε_i : part non expliquée (résidu) pour l'individu i.

Graphiquement, c'est une droite sur un nuage de points (X_i, Y_i).

2.B Qu'est-ce que l'« intercept » ?

Dans la formule :

- β_0 est appelé intercept.

Interprétation :

- si on prolonge la droite jusqu'à $X = 0$, la valeur de Y lue sur l'axe vertical est β_0 ;
- donc l'intercept est la valeur moyenne de Y quand $X = 0$ (dans le modèle).

En français : c'est le « point de départ » de la droite.

Représentation sur R :

```r
set.seed(123)
n <- 100
X <- rnorm(n, mean = 50, sd = 10)
beta0 <- 100
beta1 <- 2
epsilon <- rnorm(n, mean = 0, sd = 15)
Y <- beta0 + beta1 * X + epsilon

## 1) Ajuster les deux modèles
mod_avec_int <- lm(Y ~ X)       # avec intercept
mod_sans_int <- lm(Y ~ X - 1)   # sans intercept

coef(mod_avec_int)
```

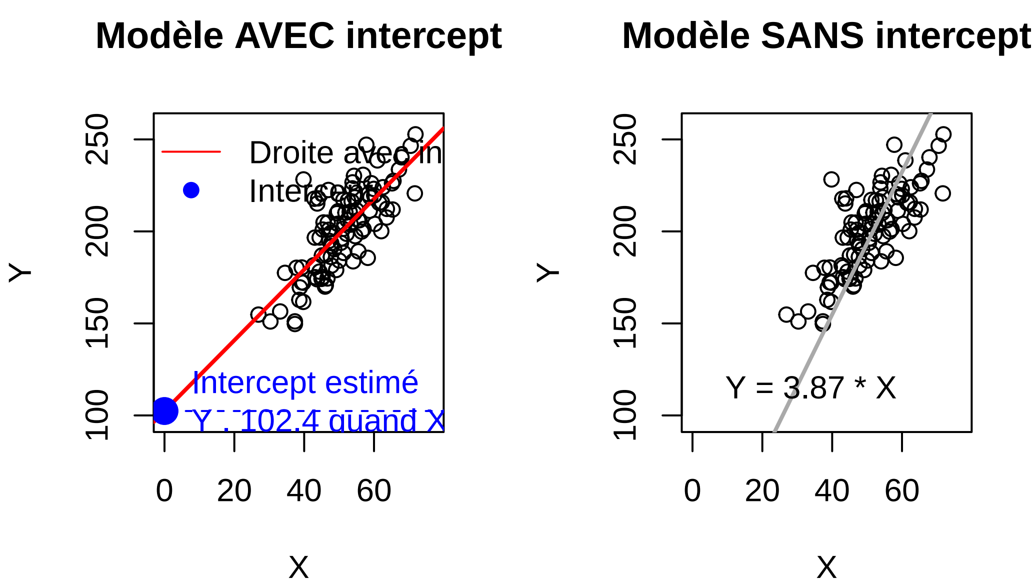
```
(Intercept)           X 
 102.393325    1.921293 
```

```r
coef(mod_sans_int)
```

```
      X
3.87073
```

```r
# Intercept estimé du modèle avec intercept
beta0_hat <- coef(mod_avec_int)[1]
beta1_hat <- coef(mod_avec_int)[2]

## 2) Limites des axes : on inclut les données ET l'intercept
xlim <- c(0, max(X) + 5)
ylim <- c(min(c(Y, beta0_hat)) - 5, max(Y) + 5)

par(mfrow = c(1, 2))

## --- PLOT 1 : Modèle AVEC intercept (on montre clairement l'intercept) --- ##
plot(
    X, Y,
    xlim = xlim, ylim = ylim,
    main = "Modèle AVEC intercept",
    xlab = "X", ylab = "Y"
)

# Droite de régression ajustée (avec intercept)
abline(mod_avec_int, col = "red", lwd = 2)

# Point de l'intercept estimé : (X = 0, Y = beta0_hat)
points(0, beta0_hat, pch = 19, col = "blue", cex = 1.8)

# Ligne horizontale au niveau de l'intercept
abline(h = beta0_hat, col = "blue", lty = 2)

# Petit texte explicatif à côté du point
text(
    x = 0,
    y = beta0_hat,
    labels = paste0("Intercept estimé\nY   ", round(beta0_hat, 1), " quand X = 0"),
    pos = 4, offset = 0.7, col = "blue"
)

legend(
    "topleft",
    legend = c("Droite avec intercept", "Intercept"),
    col    = c("red", "blue"),
    lty    = c(1, NA),
    pch    = c(NA, 19),
    bty    = "n"
)
```

```r
## --- PLOT 2 : Modèle SANS intercept (comparaison) --- ##
plot(
    X, Y,
    xlim = xlim, ylim = ylim,
    main = "Modèle SANS intercept",
    xlab = "X", ylab = "Y"
)

# Droite ajustée SANS intercept (forcée à passer par (0,0))
abline(mod_sans_int, col = "darkgray", lwd = 2)

text(
  x = min(xlim) + diff(xlim) * 0.05,
  y = min(ylim) + diff(ylim) * 0.1,
  labels = paste0("Y = ", round(coef(mod_sans_int)[1], 2), " * X"),
    pos = 4
)
```




```r
par(mfrow = c(1, 1))
```

2.C Qu'est-ce que la pente ?

- β_1 est la pente de la droite.

Interprétation :

- si X augmente de 1 unité, la valeur moyenne de Y augmente (ou diminue) en moyenne de β_1 unités.

C'est l'effet moyen de X sur Y.

□

3 Ajouter des groupes : modèle linéaire mixte simple

Maintenant, les individus sont regroupés :

- par exemple, des patients dans des centres,
- ou des mesures répétées dans des patients.

On indexe :

- i = le groupe (centre),
- j = l'observation dans le groupe i (patient ou visite).

On propose un modèle linéaire mixte simple (intercept aléatoire de groupe) :

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

- Y_{ij} : réponse pour l'observation j dans le groupe i.
- X_{ij} : covariable (âge, temps, etc.).
- β_0, β_1 : effets fixes (comme dans la régression simple).
- u_i : effet aléatoire de groupe i.
- ε_{ij} : erreur résiduelle pour l'observation j du groupe i.

Hypothèses sur les termes aléatoires :

- $u_i \sim \mathcal{N}(0, \sigma_u^2)$: les groupes ont des niveaux différents, mais ces niveaux suivent une loi normale autour de 0.
- $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$: bruit individuel.
- u_i est indépendant des ε_{ij}.

Intuition :

- β_0 : niveau moyen global.
- u_i : décalage spécifique du groupe i par rapport au niveau moyen (certains centres sont « un peu plus hauts », d'autres « un peu plus bas »).
- $\beta_1 X_{ij}$: effet de la covariable.
- ε_{ij} : bruit individuel autour de tout ça.

□

4 Les deux variances : entre groupes et à l'intérieur des groupes

On veut expliciter les 2 variances importantes :

1. variance entre groupes : σ_u^2,
2. variance à l'intérieur des groupes (résiduelle) : σ^2.

4.A Variance d'une observation Y_{ij}

On part du modèle :

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

On considère X_{ij} comme donné (fixe). La partie aléatoire est $u_i + \varepsilon_{ij}$.
On cherche :

$$\text{Var}(Y_{ij} \mid X_{ij}) = \text{Var}(u_i + \varepsilon_{ij}).$$

Comme u_i et ε_{ij} sont indépendants :

$$\text{Var}(u_i + \varepsilon_{ij}) = \text{Var}(u_i) + \text{Var}(\varepsilon_{ij}) = \sigma_u^2 + \sigma^2.$$

Donc :

- σ_u^2 est la part de la variance venant des différences entre groupes,
- σ^2 est la part de la variance venant des différences entre individus à l'intérieur d'un groupe.

La variance totale conditionnelle est :

$$\text{Var}(Y_{ij} \mid X_{ij}) = \sigma_u^2 + \sigma^2.$$

Variance totale conditionelle = variance entre groupes + variance à l'intérieur des groupes.

4.B Corrélation entre deux observations d'un même groupe

Prenons deux observations différentes j et k dans le même groupe i :

- $Y_{ij} = \beta_0 + \beta_1 X_{ij} + u_i + \varepsilon_{ij}$,
- $Y_{ik} = \beta_0 + \beta_1 X_{ik} + u_i + \varepsilon_{ik}$.

On s'intéresse à :

$$\text{Cov}(Y_{ij}, Y_{ik} \mid X_{ij}, X_{ik}).$$

Les termes déterministes (β_0, $\beta_1 X_{ij}$, $\beta_1 X_{ik}$) n'apportent pas de variance parce que ce sont des constantes.
Il reste :

$$\text{Cov}(u_i + \varepsilon_{ij}, u_i + \varepsilon_{ik}).$$

En développant et en utilisant les indépendances :

- $\text{Cov}(u_i, u_i) = \text{Var}(u_i) = \sigma_u^2$,
- les autres covariances (entre u_i et ε, entre deux ε différents) sont nulles.

Donc :

$$\text{Cov}(Y_{ij}, Y_{ik} \mid X) = \sigma_u^2.$$

La corrélation entre deux observations du même groupe est alors :

$$\text{Corr}(Y_{ij}, Y_{ik} \mid X) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma^2}.$$

Cette quantité est le coefficient de corrélation intraclasse (ICC).

□

5 Matrice de variance–covariance : expliquer simplement

Jusqu'ici, chaque groupe i avait un seul effet aléatoire : u_i (un « niveau » propre au groupe). Maintenant, imaginons qu'on autorise :

- un niveau propre au groupe,
- et une pente propre au groupe pour une variable (par exemple le temps).

5.A Exemple : intercept et pente aléatoires

On écrit :

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

- t_{ij} : temps (ou autre covariable).
- b_{0i} : décalage du groupe i sur l'axe vertical (niveau propre).
- b_{1i} : décalage de la pente du groupe i (vitesse d'évolution propre).

On suppose :

- b_{0i} et b_{1i} sont aléatoires,
- ils ont une certaine variance chacun,
- et ils peuvent être corrélés (par exemple, les groupes qui partent plus haut vont aussi plus vite).

On peut résumer ça dans un petit tableau 2×2 qu'on appelle matrice de variance–covariance :

$$\Sigma_b = \begin{pmatrix} \text{Var}(b_{0i}) & \text{Cov}(b_{0i}, b_{1i}) \\ \text{Cov}(b_{0i}, b_{1i}) & \text{Var}(b_{1i}) \end{pmatrix}.$$

Concrètement :

- Var(b_{0i}) : à quel point les niveaux des groupes sont dispersés,
- Var(b_{1i}) : à quel point les pentes des groupes sont dispersées,
- Cov(b_{0i}, b_{1i}) :
  - positive : les groupes qui partent plus haut ont aussi des pentes plus grandes,
  - négative : les groupes qui partent plus haut ont des pentes plus faibles.

La « matrice de variance–covariance », c'est juste une façon compacte de stocker :

- plusieurs variances,
- et les covariances entre les effets aléatoires.

□

6 Effets aléatoires emboîtés vs effets aléatoires croisés

6.A Effets aléatoires emboîtés (hierarchiques)

On dit que des effets sont emboîtés (ou imbriqués) quand chaque unité de niveau inférieur appartient à un seul niveau supérieur.

Exemple typique :

- des patients dans des centres,
- chaque patient est suivi dans un seul centre.

Notation :

- i : centre,
- j : patient dans le centre i,
- k : mesure répétée dans le patient j du centre i.

On pourrait écrire un modèle :

$$Y_{ijk} = \beta_0 + u_i + v_{ij} + \varepsilon_{ijk}.$$

- u_i : effet aléatoire de centre (tous les patients du centre i partagent ce u_i),
- v_{ij} : effet aléatoire de patient emboîté dans le centre i,
- ε_{ijk} : bruit résiduel.

Structure :

- patient emboîté dans centre,
- on note parfois « patient dans centre »,
- en R : (1 | centre/patient) ou (1 | centre) + (1 | centre:patient).

Idée :

- on a une première source de variabilité entre centres,
- puis, à l'intérieur d'un centre donné, une variabilité entre patients.

6.B Effets aléatoires croisés

Effets croisés : une unité de niveau inférieur peut être associée à plusieurs modalités de deux facteurs.

Exemples :

1. Patients et médecins :

```
-   un patient peut être vu par plusieurs médecins,

-   un médecin voit plusieurs patients.

-   Patients et médecins sont « croisés », pas hiérarchiques.
```

2. Élèves × enseignants dans un système où les élèves ont plusieurs profs et les profs ont plusieurs classes.

Modèle croisé simple :

- i : patient,
- j : médecin,
- k : consultation.

On pourrait écrire :

$$Y_{ijk} = \beta_0 + u_i + v_j + \varepsilon_{ijk}.$$

- u_i : effet aléatoire du patient i,
- v_j : effet aléatoire du médecin j,
- chaque observation Y_{ijk} reçoit un bout de variabilité propre au patient et un bout propre au médecin.

En R, on écrirait par exemple :

```
lmer(Y ~ 1 + (1 | patient) + (1 | medecin), data = ...)
```

Différence clé :

- emboîté : chaque unité de niveau inférieur n'appartient qu'à un seul niveau supérieur (patient dans un centre).

- croisé : les deux facteurs se combinent librement (un patient peut être vu par plusieurs médecins, un médecin peut voir plusieurs patients).

□

7 Petit exemple R pour illustrer les deux variances

On simule des données avec :

- 10 centres,
- 40 patients par centre,
- un intercept aléatoire de centre,
- une variance résiduelle.

```r
J <- 10
n_per_center <- 40

centre <- factor(rep(1:J, each = n_per_center))
x <- rnorm(J * n_per_center, mean = 0, sd = 1)

beta0 <- 2
beta1 <- 1.5
sigma_u <- 1    # écart-type inter-centre
sigma_e <- 2    # écart-type intra-centre

u <- rnorm(J, mean = 0, sd = sigma_u)
u_centre <- u[centre]

eps <- rnorm(J * n_per_center, mean = 0, sd = sigma_e)

y <- beta0 + beta1 * x + u_centre + eps

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

head(dat)
```

```
          y          x centre
1  5.578167  2.1988103      1
2  2.607432  1.3124130      1
3  6.051474 -0.2651451      1
4  3.856798  0.5431941      1
5  4.602392 -0.4143399      1
6 -2.843757 -0.4762469      1
```

On ajuste un modèle mixte :

```r
mod <- lmer(y ~ x + (1 | centre), data = dat, REML = TRUE)
summary(mod)
```

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: y ~ x + (1 | centre)
   Data: dat

REML criterion at convergence: 1731.1

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.69721 -0.65353  0.03108  0.62633  2.71704

Random effects:
 Groups   Name        Variance Std.Dev.
 centre   (Intercept) 0.3296   0.5741
 Residual             4.2620   2.0645
Number of obs: 400, groups:  centre, 10

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept)   1.7175     0.2089   9.0058   8.222 1.77e-05 ***
x             1.3779     0.1060 390.9037  13.004  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
  (Intr)
x -0.019
```

Dans la sortie :

- partie Random effects :
  - centre : estimation de σ_u^2 (variance entre centres),
  - Residual : estimation de σ^2 (variance à l'intérieur des centres).

On peut extraire les variances :

```r
vc <- as.data.frame(VarCorr(mod))
vc
```

```
       grp        var1 var2      vcov     sdcor
1   centre (Intercept) <NA> 0.3296151 0.5741211
2 Residual        <NA> <NA> 4.2619585 2.0644511
```

Et l'ICC :

```
sigma_u2_hat <- vc$vcov[vc$grp == "centre"]
sigma2_hat   <- vc$vcov[vc$grp == "Residual"]

icc_hat <- sigma_u2_hat / (sigma_u2_hat + sigma2_hat)
icc_hat
```

```
[1] 0.07178695
```

Cela correspond à :

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma^2}.$$

□

8 Résumé

1. Intercept :

```
-   paramètre $\beta_0$,

-   valeur moyenne de $Y$ quand $X = 0$,

-   point où la droite coupe l'axe vertical.
```

2. Modèle linéaire mixte simple :

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

- Y_{ij} : observation j dans le groupe i,
- X_{ij} : covariable,
- β_0, β_1 : effets fixes,
- u_i : effet aléatoire de groupe, variance σ_u^2 (entre groupes),
- ε_{ij} : bruit résiduel, variance σ^2 (dans les groupes).

3. Deux variances :

    - variance entre groupes : σ_u^2,
    - variance à l'intérieur des groupes : σ^2,
    - variance totale : $\sigma_u^2 + \sigma^2$,

- ICC : $\rho = \sigma_u^2/(\sigma_u^2 + \sigma^2)$.

4. Matrice variance–covariance :
   - sert à stocker les variances et covariances de plusieurs effets aléatoires (par exemple niveau et pente d'un même groupe),
   - c'est juste un petit tableau 2×2 avec Var et Cov.
5. Effets aléatoires :
   - emboîtés : patients dans centres (chaque patient dans un seul centre),
   - croisés : patient et médecin (un patient peut voir plusieurs médecins, un médecin voit plusieurs patients).