

# Modèles linéaires mixtes

Sandrine Le Squin

Février 2020

## Exemple

Comment évolue le nombre de régime au jeune âge d'une famille de palmier à huile ?

72 palmiers.

On dispose du nombre de régime produit annuellement pour chaque palmier à  $t=0, 1, 2, 3$  et 4 ans après l'entrée en production (= jeune âge)

# Modèle mixte pour données longitudinales

- ▶ En utilisant le package ggplot2, représenter le nombre de régime (bn) en fonction du temps (time), avec un graphique par palmier. Quel type de modélisation vous semble le plus adapté?
- ▶ Faire la régression linéaire du nombre de régime en fonction du temps. Représentez les résidus en fonction de chaque palmier. Que constatez vous ?
- ▶ Avec la fonction lmer du package lme4, estimez les paramètres des modèles suivants :

$$Y_{ij} = \beta_1 + \beta_2 x_i + b_j + \epsilon_{ij}, \quad i = 0, \dots, 4 \text{ et } j = 1, \dots, 72$$

$$\text{avec } b_j \sim N(0, \sigma_b), \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

$$Y_{ij} = \beta_1 + \beta_2 x_i + a_j x_i + \epsilon_{ij}, \quad i = 0, \dots, 4 \text{ et } j = 1, \dots, 72$$

$$\text{avec } a_j \sim N(0, \sigma_a), \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

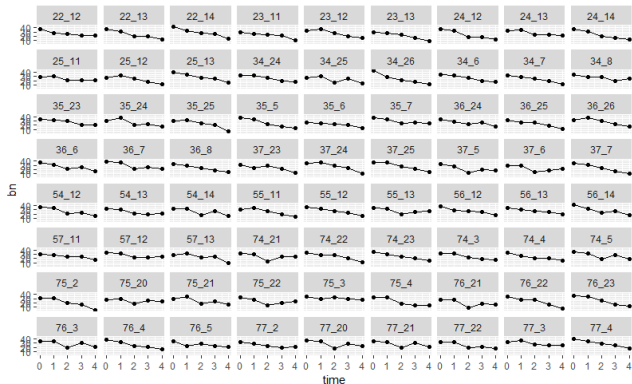
$$Y_{ij} = \beta_1 + \beta_2 x_i + b_j + a_j x_i + \epsilon_{ij}, \quad i = 0, \dots, 4 \text{ et } j = 1, \dots, 72$$

$$\text{avec } u_j = \begin{bmatrix} b_j \\ a_j \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_b^2 & \rho_{ab} \\ \rho_{ab} & \sigma_a^2 \end{bmatrix}\right)$$

# Modèle mixte pour données longitudinales

Nombre de régime en fonction du temps, par palmier

```
ggplot(long.data, aes(time, bn))+geom_point()+geom_line()  
+ facet_wrap(~palmId)
```

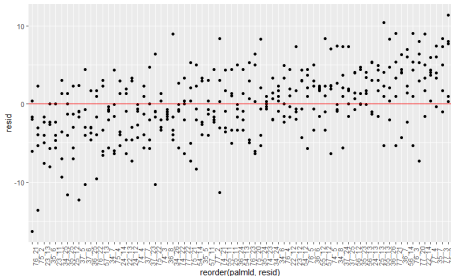


# Modèle mixte pour données longitudinales

## Régression linéaire du nombre de régime en fonction du temps

$$Y_{ij} = \beta_1 + \beta_2 x_i + \epsilon_{ij}, i = 0, \dots, 4 \text{ et } j = 1, \dots, 72$$

```
> mod <- lm(bn~time, data = long.data)
> res <- resid(mod)
> ggplot(long.data, aes(reorder(palmId, res), res))+geom_point()
  geom_hline(yintercept=0, colour=2)
```



# Modèle mixte pour données longitudinales

## Modèle avec intercept aléatoire

$$Y_{ij} = \beta_1 + \beta_2 x_i + b_j + \epsilon_{ij}, \quad i = 0, \dots, 4 \text{ et } j = 1, \dots, 72$$

```
> mod1 <- lmer(bn ~ time + (1|palmId), data = long.data)
> summary(mod1)
...
Random effects:
Groups      Name          Variance Std.Dev.
palmId      (Intercept)    2.08     1.442
Residual                    16.33     4.042
Number of obs: 360, groups:  palmId, 72

Fixed effects:
Estimate Std. Error      df    t value    Pr(>|t|)
(Intercept)  34.0472    0.4062  255.9830    83.82    <2e-16 ***
time         -4.3597    0.1506  287.0000   -28.95    <2e-16 ***
```

# Modèle mixte pour données longitudinales

## Modèle avec pente aléatoire

$$Y_{ij} = \beta_1 + \beta_2 x_i + a_j x_i + \epsilon_{ij}, \quad i = 0, \dots, 4 \text{ et } j = 1, \dots, 72$$

```
> mod2 <- lmer(bn ~ time + (0+time|palmId), data = long.data)
> summary(mod2)
...
Random effects:
Groups      Name Variance Std.Dev.
palmId      time  0.4458   0.6677
Residual                15.7446   3.9679
Number of obs: 360, groups:  palmId, 72

Fixed effects:
Estimate Std. Error      df    t value Pr(>|t|)
(Intercept)  34.0472    0.3622  287.0000   94.00   <2e-16 ***
time         -4.3597    0.1675  238.5689  -26.03   <2e-16 ***
```

# Modèle mixte pour données longitudinales

## Modèle avec pente et intercept aléatoire (corrélation)

$$Y_{ij} = \beta_1 + \beta_2 x_i + b_j + a_j x_i + \epsilon_{ij}, i = 0, \dots, 4 \text{ et } j = 1, \dots, 72$$

```
> mod3 <- lmer(bn ~ time + (time|palmId), data = long.data)
> summary(mod3)
```

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
palmId	(Intercept)	1.113	1.0550	
	time	0.424	0.6512	-0.19
Residual		15.286	3.9097	
Number of obs: 360, groups: palmId, 72				

Fixed effects:

Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	34.0472	0.3779	71.0000	90.09	<2e-16 ***
time	-4.3597	0.1647	71.0001	-26.47	<2e-16 ***



# Modèle mixte pour données longitudinales

## Modèle avec pente et intercept aléatoire

$$Y_{ij} = \beta_1 + \beta_2 x_i + b_j + a_j x_i + \epsilon_{ij}, \quad i = 0, \dots, 4 \text{ et } j = 1, \dots, 72$$

Est ce que  $b$  et  $a$  sont indépendant?

```
> mod4<-lmer(bn~time+(1|palmId)+(0+time|palmId),  
data=long.data)
```

```
summary(mod4)
```

Random effects:

Groups	Name	Variance	Std.Dev.
palmId	(Intercept)	0.8131	0.9017
palmId.1	time	0.3687	0.6072
Residual		15.4016	3.9245

# Modèle mixte pour données longitudinales

## Comparaison des deux modèles

```
anova(mod3, mod4) # refitting model(s) with ML
```

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
mod4	5	2065.56	2084.99	-1027.78	2055.56			
mod3	6	2067.54	2090.86	-1027.77	2055.54	0.02	1	0.9023

⇒ On conserve le modèle avec intercept aléatoire et pente aléatoire indépendante

# Modèle mixte pour données longitudinales

## Test de la pente

```
anova(mod4, mod2) # refitting model(s) with ML
```

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
mod2	4	2064.24	2079.78	-1028.12	2056.24			
mod4	5	2065.56	2084.99	-1027.78	2055.56	0.68	1	0.4085

## Test de l'intercept

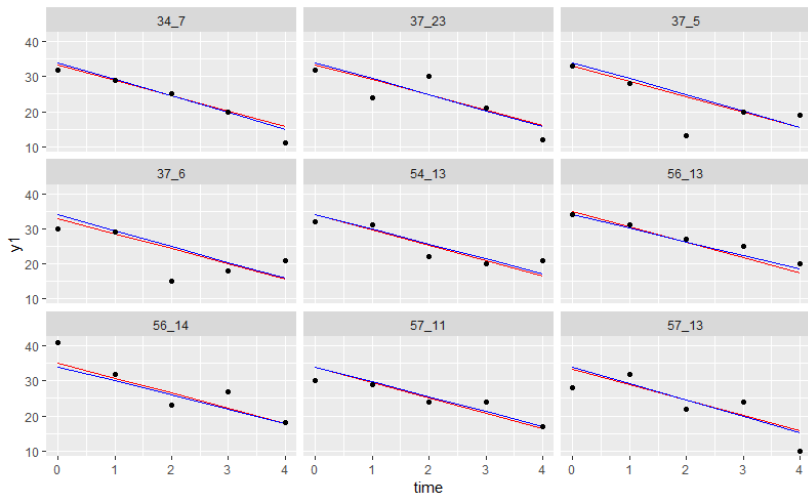
```
anova(mod4, mod2) # refitting model(s) with ML
```

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
mod1	4	2068.68	2084.23	-1030.34	2060.68			
mod4	5	2065.56	2084.99	-1027.78	2055.56	5.13	1	0.0236

# Modèle mixte pour données longitudinales

```
y1 <- predict(mod1); # conditionnelle.  
y2 <- predict(mod2); # conditionnelle  
y3 <- predict(mod3); # conditionnelle  
# ajouter re.form=NA pour pred marginale (voir long_data.R)  
  
long.data$y1 <- y1;  
long.data$y2 <- y2;  
long.data$y3 <- y3  
  
## subset de 9 individus  
sub <- subset(long.data, palmId %in% c("57_11","37_5","56_13",  
"34_7","54_13","56_14","57_13","37_6","37_23"))  
sub <- droplevels(sub)  
  
p <- ggplot(sub, aes(time, y1)) + geom_line(colour = "red")  
p <- p + geom_line(aes(time, y2), col = "blue")  
p + geom_point(data=sub, aes(time, bn))+facet_wrap(~palmId)
```

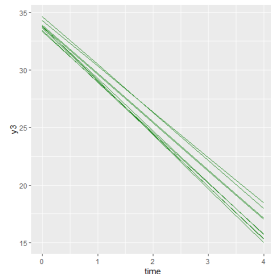
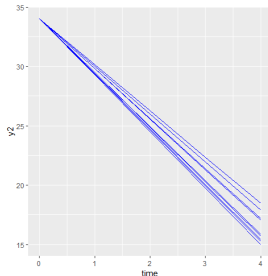
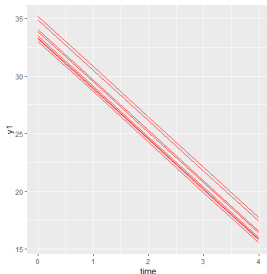
# Modèle mixte pour données longitudinales



# Modèle mixte pour données longitudinales

## Autre représentations possibles

```
# intercept aleatoire  
ggplot(sub,aes(time,y1,group=palmId))+geom_line(colour="red")  
# pente aleatoire  
ggplot(sub,aes(time,y2,group=palmId))+geom_line(colour="blue")  
# intercept + pente aleatoire  
ggplot(sub,aes(time,y3,group=palmId))+  
geom_line(colour="forestgreen")
```



# Modèle mixte pour données longitudinales

Note sur l'écriture matricielle des modèles:

Partie fixe des modèles mixtes (communs aux 4 modèles)

Rappel :  $Y = X\beta + Zu + \epsilon$

$$Y_{360 \times 1} = \begin{bmatrix} y_{0,1} \\ y_{1,1} \\ y_{2,1} \\ y_{3,1} \\ y_{4,1} \\ \dots \\ y_{4,72} \end{bmatrix}, X_{360 \times 2} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ \dots & \dots \\ 1 & 4 \end{bmatrix}, \beta = [\beta_1, \beta_2]'$$

$y_{0,1}$ : Nombre de régimes au temps 0 du palmier 1 etc.

# Modèle mixte pour données longitudinales

Note sur l'écriture matricielle des modèles:

Modèle avec intercept aléatoire

$$Z_{360 \times 72} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix} \text{ et } u_{1 \times 72} = [b_1, \dots, b_{72}]'$$



# Modèle mixte pour données longitudinales

Note sur l'écriture matricielle des modèles:

Modèle avec pente aléatoire

$$Z_{360 \times 72} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 2 & 0 & \dots & 0 \\ 3 & 0 & \dots & 0 \\ 4 & 0 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & 4 \end{bmatrix} \text{ et } u_{1 \times 72} = [a_1, \dots, a_{72}]'$$

# Modèle mixte pour données longitudinales

Note sur l'écriture matricielle des modèles:

Modèle avec intercept et pente aléatoire

$$Z = \begin{bmatrix} Z_1 & 0 & \dots & 0 \\ 0 & Z_2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & Z_{72} \end{bmatrix} \text{ et } u = [u_1, \dots, u_{72}]' \text{ avec,}$$

$$\forall j = 1, \dots, 72, Z_j = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \text{ et } u_j = \begin{bmatrix} b_j \\ a_j \end{bmatrix}$$

## Exemple

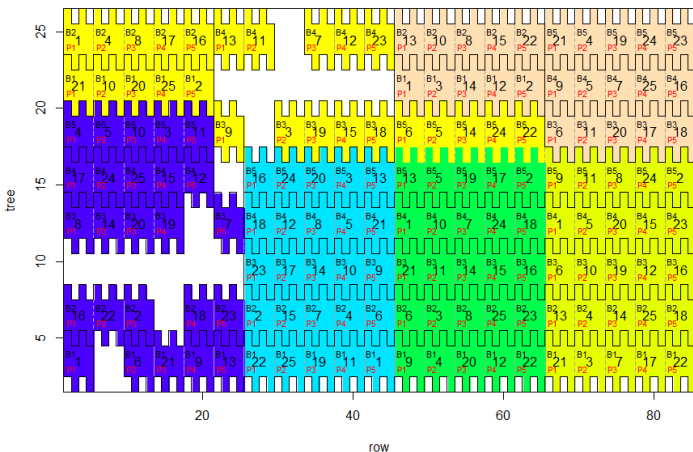
On cherche à comparer le rendement en huile à 4 ans de 25 familles de palmiers. Dispositif expérimental :

- ▶ 25 familles de palmiers à huile
- ▶ 6 répétitions
- ▶ 5 blocs par répétition (blocs car hétérogénéité du terrain)
- ▶ 5 parcelles élémentaires par blocs. Taille des parcelles = 12 palmiers

Répartition aléatoire des 25 traitements dans les 5 blocs de taille 5, répétés 6 fois, de telle sorte que chaque paire de traitement soit présente une fois dans chaque bloc.

## Modèle mixte avec effets aléatoires emboîtés

## Représentation graphique de l'essai



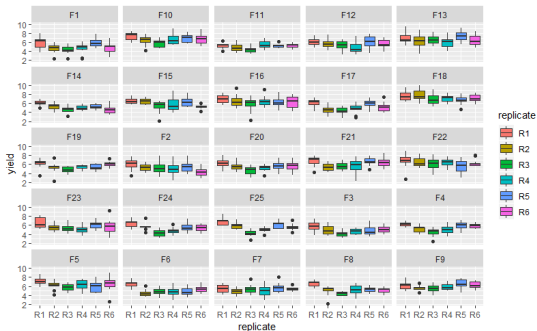
# Modèle mixte avec effets aléatoires emboîtés

- ▶ Avec ggplot2, représenter graphiquement le rendement en huile de chaque famille (facet\_wrap) en fonction de la répétition.
- ▶ Dans un premier temps, nous ignorons l'effet bloc. Quel(s) effet(s) aléatoire(s) pourraient être intégrés dans le modèle ?
- ▶ Représentez graphiquement l'interaction replicate\*family
- ▶ Avec la fonction lmer du package lme4, estimez les paramètres des modèles suivants :
  - $Y_{ijk} = \mu + \beta_i + b_j + \epsilon_{ijk}, i = 1, \dots, 25, j = 1, \dots, 6, k = 1, \dots, 12$   
 $b_j \sim N(0, \sigma_b^2), \epsilon_{ijk} \sim N(0, \sigma^2)$
  - $Y_{ijk} = \mu + \beta_i + b_j + b_{ij} + \epsilon_{ijk}, i = 1, \dots, 25, j = 1, \dots, 6, k = 1, \dots, 12$   
 $b_j \sim N(0, \sigma_1^2), b_{ij} \sim N(0, \sigma_2^2), \epsilon_{ijk} \sim N(0, \sigma^2)$
- ▶ Le facteur bloc permet-il de mieux prendre en compte l'hétérogénéité du terrain ?
- ▶ Y a-t-il un effet famille ?

# Modèle mixte avec effets aléatoires emboîtés

## Rendement en huile en fonction de la famille et de la répétition

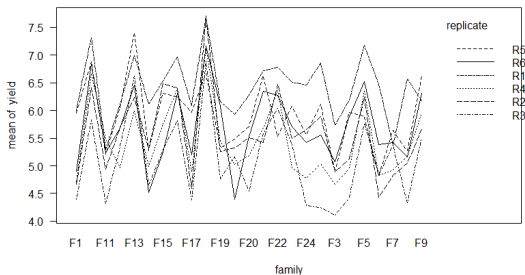
```
> ldesign <- read.csv2("lattice.design.csv")  
> ggplot(ldesign, aes(replicate, yield)) +  
  geom_boxplot(aes(fill = replicate)) + facet_wrap(~family)
```



# Modèle mixte avec effets aléatoires emboîtés

## Plot de l'interaction replicate x family

```
>attach( ldesign )  
>interaction.plot(family, replicate, yield, las = 1)  
>detach()
```



# Modèle mixte avec effets aléatoires emboîtés

## Modèle avec répétition en aléatoire

$$Y_{ijk} = \mu + \beta_i + b_j + \epsilon_{ijk}, \quad i = 1, \dots, 25, \quad j = 1, \dots, 6, \quad k = 1, \dots, 12$$

```
> mod <- lmer(yield ~ family + (1|replicate), data = ldesign)
> summary(mod)
```

Random effects:

Groups	Name	Variance	Std.Dev.
replicate	(Intercept)	0.1927	0.4390
Residual		0.9996	0.9998

Number of obs: 1778, groups: replicate, 6



# Modèle mixte avec effets aléatoires emboîtés

## Modèle avec répétition + famille dans répétition en aléatoire

$$Y_{ijk} = \mu + \beta_i + b_j + b_{ij} + \epsilon_{ijk}, \quad i = 1, \dots, 25, \quad j = 1, \dots, 6, \quad k = 1, \dots, 12$$

```
>mod2 <- lmer(yield ~ family + (1|replicate/family),  
  data = ldesign)  
>summary(mod2)  
Random effects:  
Groups          Name          Variance Std.Dev.  
family:replicate (Intercept) 0.06127  0.2475  
replicate        (Intercept) 0.19002  0.4359  
Residual                    0.94969  0.9745  
Number of obs: 1778, groups: family:replicate, 150;  
replicate, 6
```

# Modèle mixte avec effets aléatoires emboîtés

## Comparaison des modèles

`anova(mod, mod2)`

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
mod	27	5098.13	5246.17	-2522.06	5044.13			
mod2	28	5089.05	5242.59	-2516.53	5033.05	11.07	1	0.0009

# Modèle mixte avec effets aléatoires emboîtés

## Ajout de l'effet bloc

L'effet famille est emboîté dans bloc lui même emboîté dans répétition

```
> mod3 <- lmer(yield ~ family + (1|replicate/block/family),  
  data = ldesign)  
> summary(mod3)  
Random effects:  
Groups                                Name                Variance Std.Dev.  
family:(block:replicate) (Intercept) 0.053566 0.2314  
block:replicate          (Intercept) 0.009235 0.0961  
replicate                (Intercept) 0.188516 0.4342  
Residual                 0.949688 0.9745
```

# Modèle mixte avec effets aléatoires emboîtés

## Comparaison des modèles

```
anova(mod3, mod2)
```

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
mod2	28	5089.05	5242.59	-2516.53	5033.05			
mod3	29	5090.08	5249.09	-2516.04	5032.08	0.98	1	0.3227

# Modèle mixte avec effets aléatoires emboîtés

## Comparaison des modèles

```
anova(mod3, mod2)
```

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
mod2	28	5089.05	5242.59	-2516.53	5033.05			
mod3	29	5090.08	5249.09	-2516.04	5032.08	0.98	1	0.3227

# Modèle mixte avec effets aléatoires emboîtés

## Test de l'effet famille

- Y a t'il un effet famille ?

→ pseudo-test F, basé sur l'approximation de Satterthwaite. Disponible dans le package `lmerTest`

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
family	354.81	14.78	24.00	118.38	15.57	0.0000

```
> library(lmerTest)
> mod2 <- lmer(yield ~ family + (1|replicate/family),
data = ldesign)
> summary(mod2)
> anova(mod2)
```

# Modèle mixte avec effets aléatoires emboîtés

Note sur l'écriture matricielle des modèles:

Partie fixe

Rappel :  $Y = X\beta + Zu + \epsilon$

$$Y_{1778 \times 1} = \begin{bmatrix} y_{1,1,1} \\ y_{1,1,2} \\ \dots \\ y_{25,6,12} \end{bmatrix}, X_{1778 \times 26} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix}$$

$$\beta = [\mu, \beta_1, \dots, \beta_{26}]'$$

Normalement,  $25 \times 6 \times 12 = 1800$  observations.

Mais quelques palmiers morts, il y a 1778 observations (voir `nrow(ldeggin)`)

# Modèle mixte avec effets aléatoires emboîtés

Note sur l'écriture matricielle des modèles:

Modèle avec replicate en aléatoire

$$Z_{1778 \times 6} = \begin{bmatrix} Z_1 & 0 & \dots & 0 \\ 0 & Z_2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & Z_6 \end{bmatrix} \text{ et } u = [b_1, \dots, b_6]' \text{ avec,}$$

$$\forall j = 1, \dots, 6, Z_j = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}, Z_j \text{ de dimension } n_j \times 1$$

Théoriquement,  $n_j = 25 \times 12 = 300$  mais palmiers morts, donc variable selon la répétition



# Modèle mixte avec effets aléatoires emboîtés

Note sur l'écriture matricielle des modèles:

Modèle avec replicate et famille dans replicate en aléatoire

$$Z_{1778 \times 156} = \left[ \begin{array}{cccc|ccc} Z_{R1} & 0 & \dots & 0 & Z_{F1R1} & \dots & 0 \\ 0 & Z_{R2} & \dots & \dots & \dots & \dots & 0 \\ & & \ddots & & & \ddots & \\ 0 & 0 & \dots & Z_{R6} & 0 & \dots & Z_{F25R6} \end{array} \right]$$

$$u = [b_1, \dots, b_6, b_{1,1}, \dots, b_{25,6}]'$$

$$Z_{Rj} = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}, \quad Z_{FiRj} = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix} \quad i = 1, \dots, 25, j = 1, \dots, 6$$

$Z_{Rj}$  de dimension  $n_j \times 1$ ,  $n_j \simeq 300$ ,  $Z_{FiRj}$  de dimension  $n_{ij} \times 1$ ,  $n_{ij} \simeq 12$

Théoriquement,  $n_j = 25 \times 12 = 300$  et  $n_{ij} = 12$  mais quelques palmiers morts

# Modèle mixte avec effets aléatoires emboîtés

Ajout d'effet(s) aléatoire(s) et dépendance des observations

→  $Cov(Y_{ijk}, Y_{i'j'k'}), \forall i = 1, \dots, 25, j = 1, \dots, 6, k = 1, \dots, 12$  ?

Exemple du modèle avec effets aléatoires 1|replicate/family

$$Cov(Y_{ijk}, Y_{i'j'k'}) = \begin{cases} \sigma_1^2 + \sigma_2^2 + \sigma^2 & \text{si } i = i', j = j', k = k' \\ \sigma_1^2 + \sigma_2^2 & \text{si } i = i', j = j', k \neq k' \\ \sigma_1^2 & \text{si } i \neq i', j = j', k \neq k' \\ 0 & \text{sinon} \end{cases}$$

→ Des palmiers appartenant à la même famille et plantés dans la même répétition ne sont pas indépendants

→ Des palmiers appartenant à la même répétition ne sont pas indépendants

## Note sur les contrastes dans R

Par défaut dans R, c'est la famille F1 qui est la modalité de référence.

```
options()$contrasts  
contrasts(ldesign$family)
```

Pour modifier les contrastes, par exemple  $\sum \beta_i = 0$ :

```
options(contrasts = c("contr.sum", "contr.poly"))  
contrasts(ldesign$family)
```

## Exemple

Étude de l'interaction famille x environnement.

Est ce que certaines familles sont plus résistantes à la sécheresse ?

↪ Observations de plusieurs familles testées dans 3 environnements:

- ▶ 20 familles de palmiers à huile, représentative de l'ensemble du matériel végétal
- ▶ 3 sites contrastés vis à vis des conditions environnementales (sécheresse, ensoleillement)
- ▶ Données "unbalanced": certaines combinaison site x familles n'ont pas été observées

# Modèle mixte avec interaction

## Dispositif expérimental de l'essai

	Site1	Site2	Site3
GE01	125	0	100
GE02	125	123	100
GE03	125	113	99
GE04	0	118	100
GE05	0	124	100
GE06	125	116	100
GE07	125	120	99
GE08	125	106	0
GE09	125	74	99
GE10	0	122	100
GE11	0	114	99
GE12	125	123	99
GE13	125	123	0
GE14	125	123	0
GE15	0	0	99
GE16	125	122	100
GE17	0	114	99
GE18	0	123	100
GE19	125	0	100
GE20	125	0	100

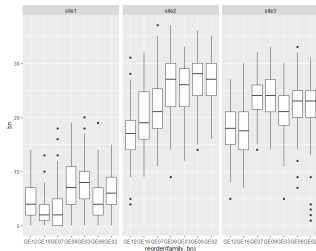
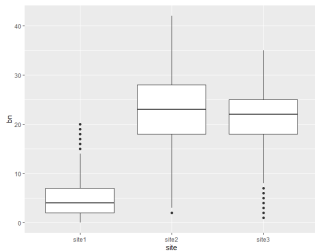
# Modèle mixte avec interaction

- ▶ Avec ggplot2, représenter graphiquement
  - le nombre de régime en fonction de chaque site
  - le nombre de régime en fonction de chaque site et par famille, pour les 7 familles en communs
- ▶ Quel(s) effet(s) aléatoire(s) pourraient être intégrés dans le modèle ?
- ▶ Le modèle est le suivant :
$$Y_{ijk} = \mu + \beta_i + u_{ij} + \epsilon_{ijk}, \quad i = 1, \dots, 3, \quad j = 1, \dots, 20, \quad k = 1, \dots, k_j$$
avec  $u_{ij} = b_j + b_{ij}$  où  $b_j$  effet aléatoire de la famille (iid) et  $b_{ij}$  l'effet aléatoire de l'interaction
- ▶ Quelles hypothèses sont possibles sur la matrice de var-cov de  $u_{ij}$ ?
- ▶ Estimer les paramètres des modèles avec la fonction lme du package nlme

# Modèle mixte avec interaction

Représentation graphique du nombre de régimes par site, puis site et famille

```
> ge <- read.csv2("ge.csv")  
> ggplot(ge, aes(site, bn)) + geom_boxplot()  
> sub <- droplevels(subset(ge[ge$family %in% c("GE02", "GE03",  
"GE06", "GE07", "GE09", "GE12", "GE16"),]))  
> ggplot(ge, aes(family, bn))+geom_boxplot()+facet_wrap(~site)
```



# Modèle mixte avec interaction

Hypothèses possibles sur la matrice de variance-covariance de  $u$   
 $\forall i, i' = 1, \dots, 3$  et  $j = 1, \dots, 20$

Modèle	$V(u_{ij})$	$Cov(u_{ij}, u_{i'j})$	Nb paramètres
D (Diagonal)	$\sigma_{u_j}^2$	0	3
CS (Compound Symmetry)	$\sigma_u^2$	$\sigma_u$	2
HCS (Heterogeneous Compound Symmetry)	$\sigma_{u_j}^2$	$\sigma_u$	3 + 1
UNS (Unstructured)	$\sigma_{u_j}^2$	$\sigma_{u_{jj'}}$	$3 \cdot (3 + 1) / 2$

## ► Diagonal Model

Une variance génétique par site + absence de covariance génétique par paire de site

## ► Compound Symmetry

Une seule variance génétique + une seule covariance génétique par paire de site

## ► Heterogeneous Compound Symmetry

Une variance génétique par site + une seule covariance génétique par paire de site

## ► Unstructured

Une variance génétique par site + une covariance génétique par paire de site



# Modèle mixte avec interaction

## Diagonal Model

Une variance par site, absence de covariance

- ▶ Variance génétique site 1:  $\hat{\sigma}_{u_1}^2 = 2.82$
- ▶ Variance génétique site 2:  $\hat{\sigma}_{u_2}^2 = 19.28$
- ▶ Variance génétique site 3:  $\hat{\sigma}_{u_3}^2 = 5.65$
- ▶ Variance résiduelle:  $\hat{\sigma}^2 = 21.33$
- ▶ AIC: 30707.23

```
> mod2 <- lme(bn ~ site, data = ge,  
random = list(family = pdDiag(~ site - 1)))  
> VarCorr(mod2)  
> summary(mod2)$AIC
```

# Modèle mixte avec interaction

## Compound Symmetry Model

Une seule variance et une seule covariance

- ▶ Variance génétique du site  $i$ :  $\hat{\sigma}_{u_i}^2 = 9.78, \forall i = 1, \dots, 3$
- ▶ Covariance génétique entre 2 site  $i$  et  $i'$ :  $\hat{\sigma}_{u_{ii'}} = 6.1, \forall i, i' = 1, \dots, 3$
- ▶ Variance résiduelle:  $\hat{\sigma}^2 = 21.33$
- ▶ AIC: 30705.97

**Remarque** : Équivaut au modèle lmer avec 1|family/site

```
> mod3 <- lme(bn ~ site, data = ge,  
  random = list(family = pdCompSymm(~ site -1)))  
> VarCorr(mod3)  
> summary(mod3)$AIC
```

# Modèle mixte avec interaction

## Heterogeneous Compound Symmetry Model

Une variance par site, une seule covariance

- ▶ Variance génétique site 1:  $\hat{\sigma}_{u_1} = 0.7$
- ▶ Variance génétique site 2:  $\hat{\sigma}_{u_2} = 10.27$
- ▶ Variancee génétique site 3:  $\hat{\sigma}_{u_3} = 3.09$
- ▶ Covariance génétique entre 2 site  $i$  et  $i'$  :  $\hat{\sigma}_{u'_{ii'}} = 3.01, \forall i, i' = 1, \dots, 3$
- ▶ Variance résiduelle:  $\hat{\sigma}^2 = 21.33$
- ▶ AIC: 30701.5

```
> mod4 <- lme(bn ~ site, data = ge,  
  random = list(family = pdBlocked(list(pdIdent(~ 1),  
                                         pdDiag(~ site - 1)))))  
> VarCorr(mod4)  
> summary(mod4)$AIC
```

# Modèle mixte avec interaction

## Unstructured Model

Une variance par site, une covariance par site

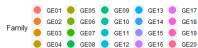
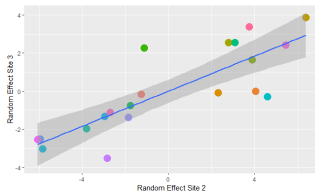
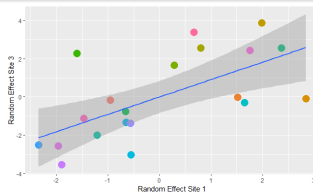
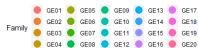
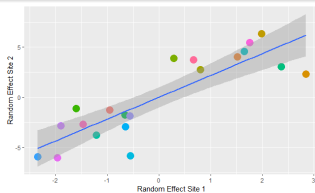
- ▶ Variance génétique site 1:  $\hat{\sigma}_{u_1} = 3.19$
- ▶ Variance génétique site 2:  $\hat{\sigma}_{u_2} = 17.8$
- ▶ Variancee génétique site 3:  $\hat{\sigma}_{u_3} = 5.65$
- ▶ Covariance génétique site 1,2 :  $\hat{\sigma}_{u_{12}} = 5.73$
- ▶ Covariance génétique site 1,3 :  $\hat{\sigma}_{u_{13}} = 3.22$
- ▶ Covariance génétique site 2,3 :  $\hat{\sigma}_{u_{23}} = 7.62$
- ▶ Variance résiduelle:  $\hat{\sigma}^2 = 21.33$
- ▶ AIC: 30695.47

```
> mod5 <- lme(bn ~ site, data = ge, random = ~site - 1|family)
> VarCorr(mod5)
> summary(mod5)$AIC
```

# Modèle mixte avec interaction

## Conclusion

Meilleur modèle: Unstructured model



- ▶ <http://bbolker.github.io/mixedmodels-misc/glmmFAQ.html>
- ▶ Bates D, Maechler M, Bolker B, Walker S (2014) *Fitting Linear Mixed-Effects Models Using lme4*  
<https://cran.r-project.org/web/packages/lme4/vignettes/lmer.pdf>
- ▶ Bates D (2010) *lme4: Mixed-effect modeling with R*  
[http://webcom.upmf-grenoble.fr/LIP/Perso/DMuller/M2R/R\\_et\\_Mixed/documents/Bates-book.pdf](http://webcom.upmf-grenoble.fr/LIP/Perso/DMuller/M2R/R_et_Mixed/documents/Bates-book.pdf)
- ▶ <http://lme4.r-forge.r-project.org/>  
voir par exemple présentation sur modèle mixte pour données longitudinales [lme4.r-forge.r-project.org/slides/2011-03-16-Amsterdam/2Longitudinal.pdf](http://lme4.r-forge.r-project.org/slides/2011-03-16-Amsterdam/2Longitudinal.pdf)
- ▶ Pinheiro J, Bates D, R-core (2013). *Mixed-Effects Models in S and S-PLUS*  
(package nlme et pas lme4, mais des exemples simples dans le premier chapitre)