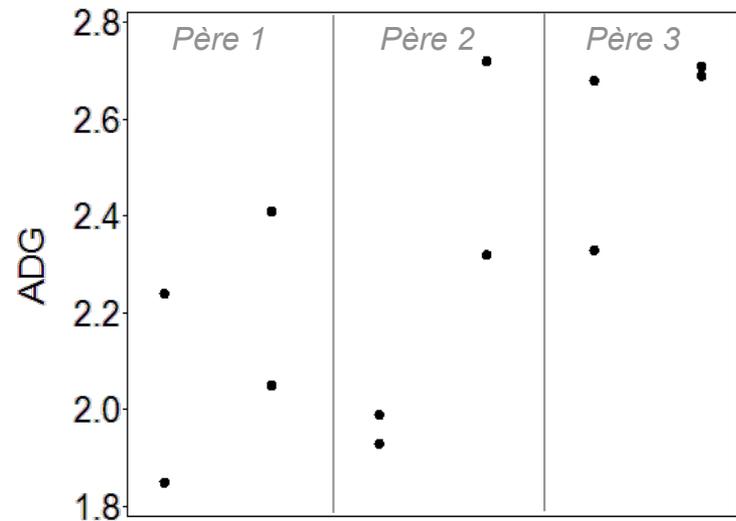
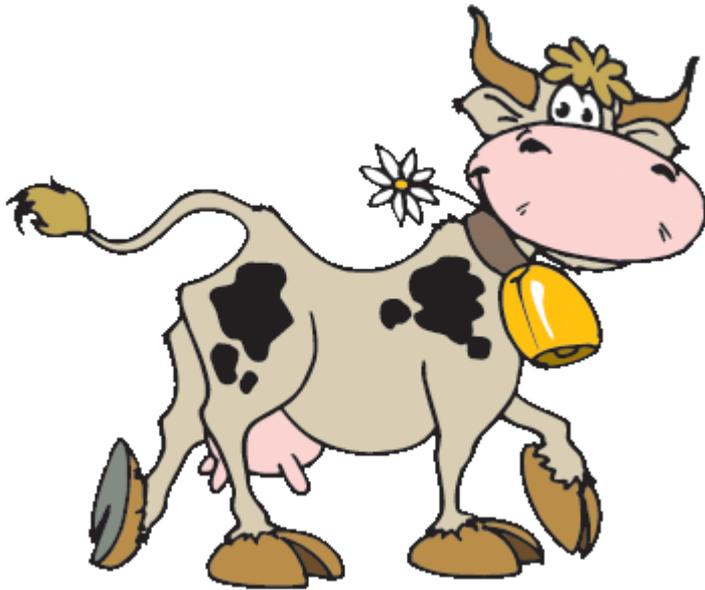


CH 02. Analyse de la variance à deux critères de classification hiérarchisés



Facteurs croisés et facteurs hiérarchisés

Les modèles à deux facteurs considérés jusqu'ici combinent les facteurs de manière croisée. Dans l'exemple des variétés et des engrais, chaque parcelle se voit affecter (aléatoirement) une des combinaisons variété \times engrais.

Variété	Type d'engrais	
	A	B
1	✓	✓
2	✓	✓
3	✓	✓

Une situation différente peut être rencontrée lorsque les niveaux du second facteur diffèrent selon le niveau du premier facteur auquel il sont associés. Ce sera souvent le cas lorsque le second facteur est aléatoire.

Facteurs hiérarchisés

Exemple de facteurs hiérarchisés:

En amélioration animale, on étudie la valeur de pères (A) en les croisant avec plusieurs mères (B) et on observe les descendance. Pour des raisons pratiques, on ne peut réaliser tous les croisements avec les mêmes mères. Ici, les deux facteurs sont aléatoires.

Facteur A	Facteur B					
	B1	B2	B3	B4	B5	B6
A1	✓	✓				
A2			✓	✓		
A3					✓	✓

La notation indicée des facteurs hiérarchisés

Dans l'exemple qui suit (tableau de droite), il y a bien 6 tirages du facteur B. Le modèle notera cependant $j=1...2$. Selon cette notation, $j=1$ ne représente pas le même niveau de B selon que $i=1$, $i=2$ ou $i=3$. Pour éviter toute confusion, on notera $j(i)$ au lieu de ij

Deux facteurs croisés

Facteur A	Facteur B	
	B1	B2
A1	11	12
A2	21	22
A3	31	32

Deux facteurs hiérarchisés

Facteur A	Facteur B					
	B1	B2	B3	B4	B5	B6
A1	1(1)	2(1)	/	/	/	/
A2	/	/	1(2)	2(2)	/	/
A3	/	/	/	/	1(3)	2(3)

Comme les niveaux de B ($j = 1, \dots$) ne sont pas les mêmes, il n'y a pas de raison de considérer un effet d'interaction. De même, estimer un effet du niveau j de B moyenné sur les différents niveaux de A n'a juste pas de sens.

Objectif et plan

Objectif

Expliquer une variable quantitative Y à partir de deux facteurs catégoriels hiérarchisés.

On considèrera le modèle suivant:

$$Y_{ijk} = \mu + a_i + b_{j(i)} + \varepsilon_{ijk}$$

Plan du chapitre

- Facteurs croisés et hiérarchisés
- Exemples
- Les hypothèses du modèle d'ANOVA 2 aléatoire hiérarchisée
- Le modèle d'ANOVA 2 hiérarchisée sous la forme d'une modèle linéaire
- Tests d'hypothèses et décomposition de la variance
- Réécriture du modèle
- Estimation par la méthode REML
- Inférence sur les moyennes

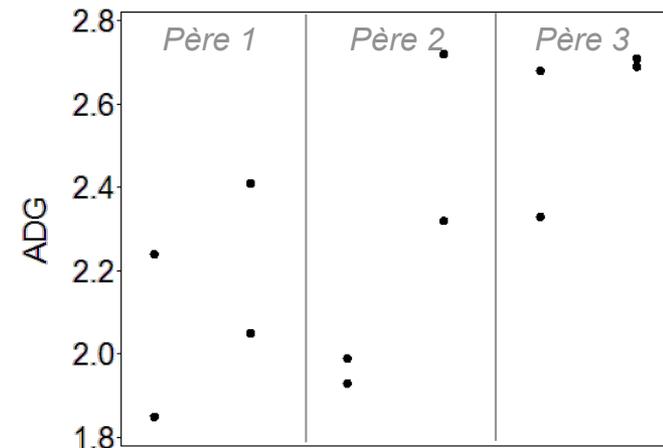
Exemple de motivation

Contexte

Amélioration animale: trois taureaux reproducteurs sont croisés chacun avec deux mères. On observe la croissance de deux jeunes par descendance.

Données

	Pere	Mere	Croissance
1	1	1	2.24
2	1	1	1.85
3	1	2	2.05
4	1	2	2.41
5	2	1	1.99
6	2	1	1.93
7	2	2	2.72
8	2	2	2.32
9	3	1	2.33
10	3	1	2.68
11	3	2	2.69
12	3	2	2.71



Objectifs/Questions

- Estimer la moyenne de la croissance dans la population d'élevage dont sont tirés les pères et les mères
- Quelle est l'importance de la variabilité entre pères, entre mères et dans la descendance?
- Estimer la valeur des taureaux reproducteurs

Le modèle de l'ANOVA 2 hiérarchisée (aléatoire)

On suppose que les observations suivent le modèle probabiliste suivant :

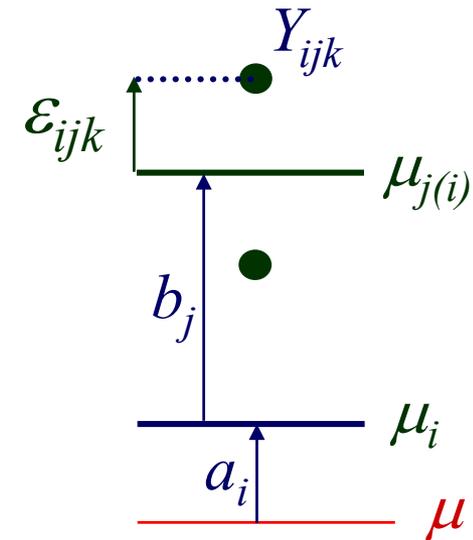
$$Y_{ijk} = \mu + a_i + b_{j(i)} + \varepsilon_{ijk} \quad i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n_{ij}$$

$$a_i \sim N(0, \sigma_a^2)$$

$$b_{j(i)} \sim N(0, \sigma_b^2)$$

$$\varepsilon_{ijk} \sim N(0, \sigma^2)$$

- $\mu_{j(i)}$ Résultat moyen attendu dans la cellule $j(i)$
- ε_{ijk} Déviation entre le descendant k et la moyenne μ_{ij} de ses frères
- μ Moyenne générale
- a_i Effet du $i^{\text{ème}}$ père
- $b_{j(i)}$ Effet de la $j^{\text{ème}}$ mère (combinée au $i^{\text{ème}}$ père)



! Il n'y a pas de terme d'interaction dans ce modèle

! On fait l'hypothèse que les $b_{j(i)}$ suivent la même distribution (quel que soit le père)

Ecriture du modèle sous la forme GLM

$$Y_{ijk} = \mu + a_i + b_{j(i)} + \varepsilon_{ijk} \quad i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n_{ij}$$

$$a_i \sim i N(0, \sigma_a^2)$$

$$b_{j(i)} \sim i N(0, \sigma_b^2)$$

$$\varepsilon_{ijk} \sim i N(0, \sigma^2)$$

$$\mathbf{Y} = \begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ \vdots \\ Y_{311} \\ Y_{312} \\ Y_{321} \\ Y_{322} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ a_1 \\ a_2 \\ a_3 \\ b_{1(1)} \\ b_{2(1)} \\ b_{1(2)} \\ b_{2(2)} \\ b_{1(3)} \\ b_{2(3)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \varepsilon_{122} \\ \vdots \\ \varepsilon_{311} \\ \varepsilon_{312} \\ \varepsilon_{321} \\ \varepsilon_{322} \end{bmatrix} = \mathbf{X}_{GLM} \boldsymbol{\beta}_{GLM} + \boldsymbol{\varepsilon}$$

Écriture du modèle sous la forme GLM

Pour rappel, R réalise le codage de la matrice X d'une manière différente et l'interprétation des paramètres doit être adaptée.

$$\mathbf{Y} = \begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ \vdots \\ Y_{311} \\ Y_{312} \\ Y_{321} \\ Y_{322} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} \mu \\ a_1 \\ a_2 \\ a_3 \\ b_{1(1)} \\ b_{2(1)} \\ b_{1(2)} \\ b_{2(2)} \\ b_{1(3)} \\ b_{2(3)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \varepsilon_{122} \\ \vdots \\ \varepsilon_{311} \\ \varepsilon_{312} \\ \varepsilon_{321} \\ \varepsilon_{322} \end{bmatrix} = \mathbf{X}_{GLM} \boldsymbol{\beta}_{GLM} + \boldsymbol{\varepsilon}$$

Décomposition de la variance - Espérance des carrés moyens

Décomposition de la variance

$$SST = SSA + SSB(A) + SSE$$

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y})^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_i - \bar{Y})^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{ij})^2$$

$N-1$ $a-1$ $ab-a$ $N-ab$

a_i $b_{j(i)}$ ε_{ijk}

Espérance des carrés moyens

$$E(MSA) = \sigma^2 + nb\sigma_a^2 + n\sigma_b^2 \qquad E(MSB(A)) = \sigma^2 + n\sigma_b^2 \qquad E(MSE) = \sigma^2$$

Estimateurs des moindres carrés (+ basés sur les E(CM) pour les variances)

$$\hat{\mu} = \bar{Y} = \frac{1}{N} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y})^2$$

$$\hat{\sigma}^2 = S^2 = \frac{1}{N - ab} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{ij})^2$$

$$\hat{\sigma}_a^2 = \frac{MSA - MSB(A)}{nb}$$

$$\hat{\sigma}_b^2 = \frac{MSB(A) - MSE}{n}$$

Conséquences du modèle ANOVA2 hiérarchisé aléatoire

Conséquences de la définition du modèle :

$$E(Y_{ijk}) = \mu \qquad \sigma_{Y_{ijk}}^2 = \sigma_Y^2 = \sigma_a^2 + \sigma_b^2 + \sigma^2$$

Les observations liées au même niveau i de a sont corrélées :

$$\sigma\{Y_{ijk}, Y_{ij'k'}\} = \sigma\{a_i + b_{j(i)}, a_i + b_{j'(i)}\} = \sigma_a^2 + \sigma_b^2$$

partagent a_i et $b_{j(i)}$

$$\sigma\{Y_{ijk}, Y_{i'j'k'}\} = \sigma\{a_i, a_{i'}\} = \sigma_a^2$$

partagent a_i , pas $b_{j(i)}$

$$\sigma\{Y_{ijk}, Y_{i'j'k'}\} = 0$$

aucun effet aléatoire commun

Matrice variance-covariance de Y

$$\text{cov}(Y_{ijk}, Y_{ijk}) : \sigma_Y^2$$

$$\text{cov}(Y_{ijk}, Y_{ij'k'}) : \text{même père, mères } \neq$$

$$\text{cov}(Y_{ijk}, Y_{ijk'}) : \text{même père, même mère}$$

$$\text{cov}(Y_{ijk}, Y_{i'j'k'}) : \text{pères et mères } \neq$$

$ijk \rightarrow$	111	112	121	122	211	212	221	222	311	312	321	322	ijk ↓
$V(Y) =$	$\sigma_a^2 + \sigma_b^2 + \sigma^2$	$\sigma_a^2 + \sigma_b^2$	σ_a^2	σ_a^2	0	0	0	0	111
	$\sigma_a^2 + \sigma_b^2$	$\sigma_a^2 + \sigma_b^2 + \sigma^2$	σ_a^2	σ_a^2	⋮			⋮	⋮			⋮	112
	σ_a^2	σ_a^2	$\sigma_a^2 + \sigma_b^2 + \sigma^2$	$\sigma_a^2 + \sigma_b^2$	⋮			⋮	⋮			⋮	121
	σ_a^2	σ_a^2	$\sigma_a^2 + \sigma_b^2$	$\sigma_a^2 + \sigma_b^2 + \sigma^2$	0	0	0	0	122
	0	0	$\sigma_a^2 + \sigma_b^2 + \sigma^2$	$\sigma_a^2 + \sigma_b^2$	σ_a^2	σ_a^2	0	0	211
	⋮			⋮	$\sigma_a^2 + \sigma_b^2$	$\sigma_a^2 + \sigma_b^2 + \sigma^2$	σ_a^2	σ_a^2	⋮			⋮	212
	⋮			⋮	σ_a^2	σ_a^2	$\sigma_a^2 + \sigma_b^2 + \sigma^2$	$\sigma_a^2 + \sigma_b^2$	⋮			⋮	221
	0	0	σ_a^2	σ_a^2	$\sigma_a^2 + \sigma_b^2$	$\sigma_a^2 + \sigma_b^2 + \sigma^2$	0	0	222
	0	0	0	0	$\sigma_a^2 + \sigma_b^2 + \sigma^2$	$\sigma_a^2 + \sigma_b^2$	σ_a^2	σ_a^2	311
	⋮			⋮	⋮			⋮	$\sigma_a^2 + \sigma_b^2$	$\sigma_a^2 + \sigma_b^2 + \sigma^2$	σ_a^2	σ_a^2	312
	⋮			⋮	⋮			⋮	σ_a^2	σ_a^2	$\sigma_a^2 + \sigma_b^2 + \sigma^2$	$\sigma_a^2 + \sigma_b^2$	321
	0	0	0	0	σ_a^2	σ_a^2	$\sigma_a^2 + \sigma_b^2$	$\sigma_a^2 + \sigma_b^2 + \sigma^2$	322

Programme R - package lm

```
dataset <- data.frame("Sire" = kronecker(c(1, 2, 3), c(1, 1, 1, 1)),
                     "Dam" = kronecker(c(1, 2, 1, 2, 1, 2), c(1, 1)),
                     "ADG" = c(2.24, 1.85, 2.05, 2.41, 1.99, 1.93,
                                2.72, 2.32, 2.33, 2.68, 2.69, 2.71))
dataset$Sire <- as.factor(dataset$Sire)
dataset$Dam <- as.factor(dataset$Dam)
```

```
mod_lm <- stats::lm(ADG ~ Sire/Dam, dataset)
summary(mod_lm)
```

```
Call:
stats::lm(formula = ADG ~ Sire/Dam, data = dataset)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.2000	-0.1762	0.0000	0.1762	0.2000

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.0450	0.1539	13.291	1.12e-05	***
Sire2	-0.0850	0.2176	-0.391	0.7096	
Sire3	0.4600	0.2176	2.114	0.0789	.
Sire1:Dam2	0.1850	0.2176	0.850	0.4278	
Sire2:Dam2	0.5600	0.2176	2.574	0.0421	*
Sire3:Dam2	0.1950	0.2176	0.896	0.4047	

Programme R - package lm

```
stats::anova(mod_lm)
```

Analysis of Variance Table

MSA

Response: ADG

	Df	Sum Sq	Mean Sq	F	Value	Pr(>F)
Sire	2	0.47752	0.23876	5.0424	0.0519	.
Sire:Dam	3	0.38585	0.12862	2.7163	0.1375	
Residuals	6	0.28410	0.04735			

MSB(A)

MSE

On peut alors retrouver les estimations des paramètres de variance en utilisant les estimateurs des moindres carrés (cf dia 10).

$$\hat{\sigma}^2 = MSE = 0.047$$

$$\hat{\sigma}_a^2 = \frac{MSA - MSB(A)}{nb} = 0.028$$

$$\hat{\sigma}_b^2 = \frac{MSB(A) - MSE}{n} = 0.041$$

Programme R - package lm

```
stats::anova(mod_lm)
```

Analysis of Variance Table

Response: ADG

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Sire	2	0.47752	0.23876	5.0424	0.0519 .
Sire:Dam	3	0.38585	0.12862	2.7163	0.1375
Residuals	6	0.28410	0.04735		

Considérant les espérances des carrés moyens (dia 10), on peut construire des rapports de CM pour tester les hypothèses suivantes :

$$H_0: \sigma_a^2 = 0 \quad H_1: \sigma_a^2 > 0 \quad MSA / MSB(A) \sim F(a-1, a(b-1)) \quad \text{sous } H_0$$

$$H_0: \sigma_b^2 = 0 \quad H_1: \sigma_b^2 > 0 \quad MSB(A) / MSE \sim F(a(b-1), N-ab) \quad \text{sous } H_0$$

Ces rapports **ne sont pas** ceux retournés par la fonction `anova(mod_lm)` !!!

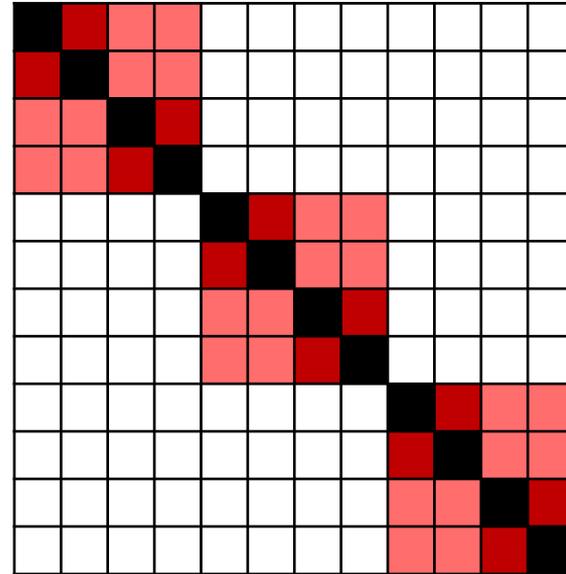
Limites de l'approche GLM

L'approche GLM basée sur la méthode d'estimation des moindres carrés ordinaires

1. Suppose que la matrice variance-covariance des Y_{ij} est $\sigma^2 \mathbf{I}$
2. Ne fournit pas les estimations correctes si les données sont non balancées
3. N'est pas nécessairement la meilleure méthode d'estimation
4. N'estime pas directement les paramètres de variances
5. Obtient dans certains cas une composante de variance négative
6. Ne fournit pas directement l'estimateur de μ

Structure de la matrice Variance Covariance des Y

$$\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$$



$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V}_3 \end{bmatrix}$$

$$\mathbf{V}_1 = \mathbf{V}_2 = \mathbf{V}_3 = \begin{bmatrix} \sigma_a^2 + \sigma_b^2 + \sigma^2 & \sigma_a^2 + \sigma_b^2 & \sigma_a^2 & \sigma_a^2 \\ \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 + \sigma^2 & \sigma_a^2 & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2 & \sigma_a^2 + \sigma_b^2 + \sigma^2 & \sigma_a^2 + \sigma_b^2 \\ \sigma_a^2 & \sigma_a^2 & \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 + \sigma^2 \end{bmatrix}$$

Modélisation R, package lme4

```
require(lme4)

mod_lmer <- lme4::lmer(ADG ~ 1 + (1|Sire) + (1|Dam:Sire), dataset)
summary(mod_lmer)
```

fixe

aléatoire

```
Linear mixed model fit by REML ['lmerMod']
Formula: ADG ~ 1 + (1 | Sire) + (1 | Dam:Sire)
Data: dataset
```

REML criterion at convergence: 6.4

Scaled residuals:

Min	1Q	Median	3Q	Max
-1.22504	-0.69738	-0.02222	0.62825	1.31385

Random effects:

Groups	Name	Variance	Std.Dev.
Dam:Sire	(Intercept)	0.04063	0.2016
Sire	(Intercept)	0.02754	0.1659
Residual		0.04735	0.2176

Number of obs: 12, groups: Dam:Sire, 6; Sire, 3

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	2.3267	0.1411	16.5

Modélisation R, package lme4

```
getME(mod_lmer, "z")
```

```
      1:1 1:2 1:3 2:1 2:2 2:3 1 2 3
1      1  .  .  .  .  .  1 . .
2      1  .  .  .  .  .  1 . .
3      .  .  .  1  .  .  1 . .
4      .  .  .  1  .  .  1 . .
5      .  1  .  .  .  .  . 1 .
6      .  1  .  .  .  .  . 1 .
7      .  .  .  .  1  .  . 1 .
8      .  .  .  .  1  .  . 1 .
9      .  .  1  .  .  .  . . 1
10     .  .  1  .  .  .  . . 1
11     .  .  .  .  .  1  . . 1
12     .  .  .  .  .  1  . . 1
```

```
(getME(mod_lmer, "Lambda") * sigma(mod_lmer)) ** 2
```

```
[1,] 0.04063333 . . . . . . . .
[2,] . 0.04063333 . . . . . . .
[3,] . . 0.04063333 . . . . . .
[4,] . . . 0.04063333 . . . . .
[5,] . . . . 0.04063333 . . . .
[6,] . . . . . 0.04063333 . . .
[7,] . . . . . . 0.02753542 . .
[8,] . . . . . . . 0.02753542 .
[9,] . . . . . . . . 0.02753542
```

Inference sur la moyenne de la population

L'estimateur non biaisé de la moyenne de la population est : $\hat{\mu} = \bar{Y}$

La variance de cet estimateur est :

$$\text{var}_{\infty}(\hat{\beta}) = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}$$

$$\sigma_{\bar{Y}}^2 = \frac{\sigma_a^2}{a} + \frac{\sigma_b^2}{ab} + \frac{\sigma^2}{N} = \frac{nb\sigma_a^2 + n\sigma_b^2 + \sigma^2}{N}$$

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	2.3267	0.1411	16.5

On peut ainsi montrer que : $\frac{\bar{Y} - \mu}{\hat{\sigma}_{\bar{Y}}} \sim t(a - 1)$

Et on obtient l'intervalle de confiance pour μ : $\bar{Y} \pm t(1 - \alpha/2; a - 1) \sqrt{\frac{MSA}{N}}$

```
require(emmeans)
emmeans(mod_1mer, ~ 1)

  1      emmean   SE df lower.CL upper.CL
overall 2.33 0.141  2    1.72    2.93

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
```

Les hypothèses du modèle d'ANOVA2 aléatoire

Les hypothèses principales du modèle ANOVA2 aléatoire hiérarchisé sont que les effets des pères sont égaux, ou que les effets des mères sont égaux :

$$H_0: \sigma_a^2 = 0 \quad H_1: \sigma_a^2 > 0$$

$$H_0: \sigma_b^2 = 0 \quad H_1: \sigma_b^2 > 0$$

Ces hypothèses peuvent être testées en partant des espérances des CM:

$$MSA / MSB(A) \sim F(a-1, a(b-1)) \quad \text{sous } H_0$$

$$MSB(A) / MSE \sim F(a(b-1), N-ab) \quad \text{sous } H_0$$

Cette méthode est toutefois délicate à généraliser pour des modèles plus complexes ou lorsque les données sont non balancées.

Les hypothèses du modèle d'ANOVA2 aléatoire

Une manière alternative de tester ces hypothèses est d'utiliser le test du rapport de vraisemblance de modèles emboîtés basé sur la statistique suivante :

$$LRT = -2 \log_e \left[\frac{L(R)}{L(F)} \right] = -2 [\log_e L(R) - \log_e L(F)] \sim \chi_{(df)}^2 \text{ sous } H_0$$

où $L(R)$ et $L(F)$ sont les vraisemblances des modèles restreint et complet (respectivement) et df le nombre de paramètres soustraits du modèle complet (ici $df = 1$ puisqu'on enlève σ_a^2).

On rejette l'hypothèse H_0 lorsque $LRT > \chi_{(0.95;df)}^2$

Pour les hypothèses de l'ANOVA2, on considère les modèles restreints suivants:

$$H_0: \sigma_a^2 = 0 \quad H_1: \sigma_a^2 > 0$$

$$Y_{ijk} = \mu + b_{j(i)} + \varepsilon_{ijk}$$

$$H_0: \sigma_b^2 = 0 \quad H_1: \sigma_b^2 > 0$$

$$Y_{ijk} = \mu + a_i + \varepsilon_{ijk}$$

Modélisation R, package lme4

```
require(lmerTest)
```

```
lmerTest::ranova(mod_lmer)
```

ANOVA-like table for random-effects: Single term deletions

Model:

```
ADG ~ (1 | Sire) + (1 | Dam:Sire)
```

	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
<none>	4	-3.1915	14.383			
(1 Sire)	3	-3.3093	12.619	0.23559	1	0.6274
(1 Dam:Sire)	3	-3.7285	13.457	1.07390	1	0.3001

Prédiction des effets aléatoires

Dans certains cas, on peut être amené à vouloir prédire l'effet d'un père ou d'une mère sur le caractère étudié. Comme l'espérance de \mathbf{u} est nulle, on utilise l'espérance conditionnelle aux valeurs \mathbf{y} observées.

$$E[\mathbf{u}|\mathbf{y}] = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

$$\hat{\mathbf{u}}|\mathbf{y} = \hat{\mathbf{G}}\mathbf{Z}'\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

```
ranef(mod_lmer, condVar = TRUE)
```

```
$`Dam:Sire`  
  (Intercept)  
1:1 -0.122833330  
1:2 -0.206417400  
1:3  0.032280407  
2:1 -0.005940756  
2:2  0.147419580  
2:3  0.155491499
```

```
$Sire  
  (Intercept)  
1 -0.08726453  
2 -0.03998023  
3  0.12724476
```

```
coef(mod_lmer)
```

```
$`Dam:Sire`  
  (Intercept)  
1:1  2.203833  
1:2  2.120249  
1:3  2.358947  
2:1  2.320726  
2:2  2.474086  
2:3  2.482158
```

```
$Sire  
  (Intercept)  
1  2.239402  
2  2.286686  
3  2.453911
```