# L'ANALYSE DE VARIANCE A DEUX FACTEURS

## Données expérimentales

| $Y_{ijk}$ | **Variété 1** | **Variété 2** | **Variété 3** |
|---|---|---|---|
| **Sol 1** | 6 10 11 | 13 15 | 14 22 |
| **Sol 2** | 12 19 15 18 | 31 | 18 9 12 |

| $\widehat{\theta_{ij}}$ | **1** | **2** | **3** | $\widehat{\theta_{i.}}$ |
|---|---|---|---|---|
| **1** | 9 | 14 | 18 | 13.67 |
| **2** | 16 | 31 | 13 | 20.0 |
| $\widehat{\theta_{.j}}$ | 12.5 | 22.5 | 15.5 | $\widehat{\theta_{..}} =$ 16.83 |

**Rappel:**
On a conclu à l'existence de $p$ différences significatives entre les couples (*sol, variété*).
On veut aller plus loin et tester l'incidence respective des deux facteurs *sol* et *variété* indépendamment.

On estime des moyennes liées aux deux facteurs de variation *sol* et *variété*:

$$\widehat{\theta_{i.}} = \text{moyenne des } \widehat{\theta_{ij}} \text{ pour chaque ligne } i$$
$$\widehat{\theta_{.j}} = \text{moyenne des } \widehat{\theta_{ij}} \text{ pour chaque colonne } j$$

**Rappel:**
$\widehat{\theta_{ij}}$ = moyenne de la case $(i,j)$
$\widehat{\theta_{..}}$ = moyenne des $\widehat{\theta_{ij}}$

# Estimation des effets

| $\widehat{\theta_{ij}}$ | **1** | **2** | **3** | $\widehat{\theta_{i.}}$ | $\widehat{\alpha_i}$ $\widehat{\theta_{i.}} - \widehat{\theta_{..}}$ |
|---|---|---|---|---|---|
| **1** | 9 | 14 | 18 | 13.67 | -3.17 |
| **2** | 16 | 31 | 13 | 20.0 | 3.17 |
| $\widehat{\theta_{.j}}$ | 12.5 | 22.5 | 15.5 | $\widehat{\theta_{..}} =$ 16.83 | |
| $\widehat{\beta_j}$ $\widehat{\theta_{.j}} - \widehat{\theta_{..}}$ | -4.33 | 5.67 | -1.33 | | |

| $\widehat{\gamma_{ij}}$ | **1** | **2** | **3** |
|---|---|---|---|
| **1** | -0.33 | -5.33 | 5.67 |
| **2** | 0.33 | 5.33 | -5.67 |

On utilise ces moyennes pour décomposer les $\widehat{\theta_{ij}}$ en:

- effet du sol $(\widehat{\alpha_i})$ : $\widehat{\theta_{i\cdot}} - \widehat{\theta_{\cdot\cdot}}$
- effet de la variété $\left(\widehat{\beta_j}\right)$ : $\widehat{\theta_{\cdot j}} - \widehat{\theta_{\cdot\cdot}}$
- reste $(\widehat{\gamma_{ij}})$ : $\widehat{\theta_{ij}} - \left(\widehat{\theta_{i\cdot}} - \widehat{\theta_{\cdot\cdot}}\right) - \left(\widehat{\theta_{\cdot j}} - \widehat{\theta_{\cdot\cdot}}\right) - \widehat{\theta_{\cdot\cdot}}$

## ECRITURE EQUIVALENTE DU MODELE GENERAL

$$Y_{ijk} = \theta_{ij} + \varepsilon_{ijk}$$

$$\downarrow$$

$$\underline{Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}}$$

**Modèle complet à deux facteurs**

Du modèle des couples (***sol, variété***),

on passe

au modèle complet à deux facteurs (équivalent au modèle en $\boldsymbol{\theta}_{ij}$).

Ainsi définis, les effets du modèle vérifient les égalités suivantes :

$$\sum_i \alpha_i = 0$$

$$\sum_j \beta_j = 0$$

$$\sum_j \gamma_{ij} = 0 \qquad \text{pour tout } i$$

$$\sum_i \gamma_{ij} = 0 \qquad \text{pour tout } j$$

Ces égalités sont appelées:

**contraintes**
ou
**conditions supplémentaires**

(On pourra reprendre le second transparent 8.1, pour vérifier ces conditions).

# ESTIMATION DES EFFETS DU MODELE A DEUX FACTEURS

| Effet | | Estimateur |
|---|---|---|
| Moyenne générale | $\mu$ | $\widehat{\theta_{..}} = \frac{1}{IJ} \sum_i \sum_j \widehat{\theta_{ij}}$ |
| Effet de la modalité $i$ du premier facteur | $\alpha_i$ | $\widehat{\theta_{i.}} - \widehat{\theta_{..}} = \frac{1}{J} \sum_j \widehat{\theta_{ij}} - \widehat{\theta_{..}}$ |
| Effet de la modalité $j$ du second facteur | $\beta_j$ | $\widehat{\theta_{.j}} - \widehat{\theta_{..}} = \frac{1}{I} \sum_i \widehat{\theta_{ij}} - \widehat{\theta_{..}}$ |
| Reste | $\gamma_{ij}$ | $\widehat{\theta_{ij}} - (\widehat{\theta_{i.}} - \widehat{\theta_{..}}) - (\widehat{\theta_{.j}} - \widehat{\theta_{..}}) - \widehat{\theta_{..}}$ ou $\widehat{\theta_{ij}} - \widehat{\theta_{i.}} - \widehat{\theta_{.j}} + \widehat{\theta_{..}}$ |

Ces estimateurs sont ceux du modèle à deux facteurs précédemment définis.
Ils sont valables dans tous les cas (équilibrés ou déséquilibrés).

On rappellera que $\widehat{\theta_{..}}$ est la **moyenne des $\widehat{\theta_{ij}}$**, et non la moyenne de l'ensemble des données $Y_{ijk}$.

## ECRITURE DU MODELE A DEUX FACTEURS

| **Modèle complet à deux facteurs** | | | |
|---|---|---|---|
| $\theta_{ij} = \mu \ +$ | $\alpha_i \ +$ | $\beta_j \ +$ | $\gamma_{ij}$ |
| | $\downarrow$ | $\downarrow$ | $\downarrow$ |
| | Effet du premier facteur | Effet du second facteur | Reste ou Interaction |

Lorsque les $\gamma_{ij}$ sont tous nuls, on utilise le **sous-modèle** :

$$\theta_{ij} = \mu + \alpha_i + \beta_j$$

Le **reste** (ou interaction) dépend des deux facteurs simultanément.

Le sous-modèle sans interaction est appelé **modèle additif** car les effets de chacun des deux facteurs sont additionnés.

Un tel modèle suppose que les données vérifient:

$$\begin{array}{l} \theta_{11} - \theta_{12} = \theta_{21} - \theta_{22} \\ \theta_{12} - \theta_{13} = \theta_{22} - \theta_{23} \\ ect \ldots \end{array}$$

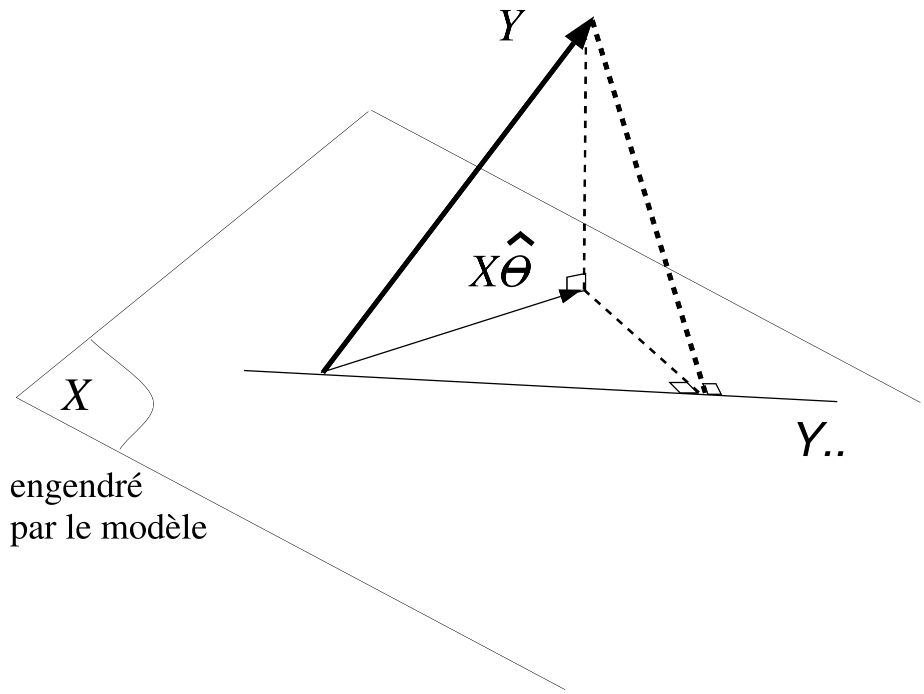
En effet si les $\gamma_{ij}$ sont nuls:

$$\left.\begin{array}{l} \theta_{11} = \mu + \alpha_1 + \beta_1 + \cancel{\gamma_{11}} \\ \underline{\theta_{12} = \mu + \alpha_1 + \beta_2 + \cancel{\gamma_{12}}} \\ \theta_{11} - \theta_{12} = \beta_1 - \beta_2 \\ \theta_{21} = \mu + \alpha_2 + \beta_1 + \cancel{\gamma_{21}} \\ \underline{\theta_{22} = \mu + \alpha_2 + \beta_2 + \cancel{\gamma_{22}}} \\ \theta_{21} - \theta_{22} = \beta_1 - \beta_2 \end{array}\right\} \theta_{11} - \theta_{12} = \theta_{21} - \theta_{22}$$

On pourra compléter (en groupe) le tableau suivant dont les données justifient le modèle additif:

| | FACTEUR 2 | | |
|---|---|---|---|
| FACTEUR 1 | 5 | 4 | 1 |
| | 7 | ? | ? |

# DECOMPOSITION DE LA SOMME DES CARRES EXPLIQUEE PAR LE MODELE




| Cas équirépété | Cas non-équirépété |
|---|---|
| $n_{ij}$ = constante | $n_{ij}$ = variable |
| ↓ | ↓ |
| **Décomposition UNIQUE** | **Décompositions MULTIPLES** |

On étudie maintenant comment se décompose (dans le cas général) la somme des carrés associée au modèle.

Cette somme est la projection du vecteur des observations dans le plan de la variation expliquée par le modèle (les $\widehat{\theta_{ij}}$).
Ce vecteur $(X\hat{\theta})$ s'écarte d'autant plus de la moyenne $Y_{..}$ que les differences entre les $\widehat{\theta_{ij}}$ sont grandes (qu'il y a un effet fort des couples $(\alpha_i, \beta_j)$).

## DECOMPOSITION DE LA SOMME DES CARRES EXPLIQUEE PAR LE MODELE

### Cas équirépété à $r$ répétitions

| Variation | Somme des carrés | DDL | Carré moyen | F |
|---|---|---|---|---|
| 1er facteur | $\sum_{ijk}\left(\widehat{\theta_{i.}}-\widehat{\theta_{..}}\right)^2$ | $I-1$ | $CM_1$ | $\dfrac{CM_1}{CM_r}$ |
| 2e facteur | $\sum_{ijk}\left(\widehat{\theta_{.j}}-\widehat{\theta_{..}}\right)^2$ | $J-1$ | $CM_2$ | $\dfrac{CM_2}{CM_r}$ |
| Inter-action | $\sum_{ijk}\left(\widehat{\theta_{ij}}-\widehat{\theta_{i.}}-\widehat{\theta_{.j}}+\widehat{\theta_{..}}\right)^2$ | $(I-1)(J-1)$ | $CM_I$ | $\dfrac{CM_I}{CM_r}$ |
| Résiduelle | $\sum_{ijk}\left(Y_{ijk}-Y_{ij.}\right)^2$ | $(r-1)IJ$ | $CM_r$ | |

Dans le cas équirépété, les sommes de carrés ont une expression simple basée sur les moyennes.

La table d'analyse de variance est **unique**.

**NB:** Les sommes se font sur **l'ensemble** des observations ($i$, $j$ et $k$).

# EXEMPLE D'APPLICATION DE L'ANALYSE DE LA VARIANCE DANS LE CAS EQUIREPETE

Etude de la teneur en huile de populations de tournesol

Données expérimentales

| Origine | AFRIQUE | HONGRIE | MAROC |
|---|---|---|---|
| Testeur 1 | 43.54<br>45.30 | 44.25<br>42.55 | 47.28<br>49.40 |
| Testeur 2 | 47.21<br>47.73 | 44.34<br>46.49 | 47.75<br>49.47 |

Le modèle est illustré par cet exemple de situation expérimentale :
***à construire au tableau avec les stagiaires***

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

$$\swarrow \qquad \downarrow \qquad \downarrow \qquad \searrow$$

Effet testeur, origine, Interaction, Résidu

# ANALYSE DE LA VARIANCE AVEC *SAS*

## Procédure GLM , cas équirépété

## Exemple Tournesol

```
        General Linear Models Procedure
Dependent Variable: HUILE
                   Sum of      Mean
Source     DF     Squares    Square  F Value   Pr > F


Model       5   47.1814417  9.4362883    6.18   0.0233


Error       6    9.1666500  1.5277750


Corrected 11   56.3480917
Total
    R-Square         C.V.     Root MSE    HUILE Mean
    0.837321     2.671010      1.23603     46.275833

Source     DF   Type I SS  Mean Square F Value Pr > F

TESTEUR     1   9.4874083    9.4874083   6.21   0.0470
ORIGINE     2  33.7458167   16.8729083  11.04   0.0097
TESTEUR*    2   3.9482167    1.9741083   1.29   0.3415
ORIGINE
```

Sortie de la procédure GLM ("General Linear Models") du logiciel SAS.

## La situation la plus courante :
## LE CAS NON-EQUIREPETE

- Hypothèse nulle : *la réponse moyenne $\theta_{i.}$ ne dépend pas de la modalité i du facteur*

$$H_0 : \theta_{1.} = \theta_{2.} = \cdots = \theta_{I.} = \theta_{..}$$
$$\alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$$

- Somme des carrés :

$$SC(\alpha) = \sum_{i=1}^{I} \omega_i \left( \widehat{\theta_{i.}} - \frac{\sum_i \omega_i \widehat{\theta_{i.}}}{\sum_i \omega_i} \right)^2$$

où

$$\frac{1}{\omega_i} = \frac{1}{J^2} \sum_{j=1}^{J} \frac{1}{n_{ij}}$$

Cette décomposition correspond aux sommes de carrés de type III du logiciel SAS.

C'est l'hypothèse habituellement considérée.

**NB:** Les tests des différents facteurs sont indépendants les uns des autres.

Les sommes des carrés tiennent compte des effectifs $n_{ij}$.
Leur somme n'est **pas égale** à la somme des carrés totale.

Le type III n'est efficace que si aucun des $n_{ij}$ n'est nul.
Dans le cas contraire, utiliser les sommes de carrés de type IV.

*Remarque pour le formateur:* Ces sommes font appel à une théorie différente (fonctions estimables).

## CAS NON-EQUIREPETE
## TESTS SEQUENTIELS

- Hypothèse nulle : *l'ajout d'un facteur d'indice j n'apporte pas une part d'explication significative par rapport au modèle*

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

- Somme de carrés séquentielles :

| Source | Somme de carrés |
|---|---|
| $\alpha$ | $SC\,(\alpha/\mu)$ |
| $\beta$ | $SC\,(\beta/\mu, \alpha)$ |
| $\gamma$ | $SC\,(\gamma/\mu, \alpha, \beta)$ |

Cette décomposition correspond aux sommes de carrés de type I de SAS, ainsi qu'à celles données par les logiciels MODLI et ANVARM.

NB : *le test dépend de l'ordre des facteurs dans le modèle*

On teste donc:

$$\begin{cases} \theta_{11} = \theta_{12} = \cdots = \theta_{1\mathrm{J}} = \theta_{1\cdot} \\ \theta_{21} = \theta_{22} = \cdots = \theta_{2\mathrm{J}} = \theta_{2\cdot} \\ \vdots \\ \theta_{\mathrm{I}1} = \theta_{\mathrm{I}2} = \cdots = \theta_{\mathrm{IJ}} = \theta_{\mathrm{I}\cdot} \\ \end{cases}$$

Les sommes des carrés séquentielless sont calculées, pas à pas, par la réduction de somme de carrés résiduelle obtenue lors de l'ajout d'un facteur supplémentaire.

La somme de carrés de $\beta$ tient compte de $\alpha$ et non le contraire; $\alpha$ n'est pas ajusté de $\beta$.

*Remarque pour le formateur:*

Les SCE type II de SAS correspondent à: $\left| \begin{array}{ll} \alpha & \mathrm{SC}(\alpha/\mu, \beta) \\ \beta & \mathrm{SC}(\beta/\mu, \alpha) \\ \mu & \mathrm{SC}(\gamma/\mu, \alpha, \beta) \end{array} \right.$

# MODELE COMPLET A DEUX FACTEURS

## Exemple Carottes

General Linear Models Procedure

Dependent Variable: GERMINAT

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 5 | 400.0000000 | 80.0000000 | 6.00 | 0.0103 |
| Error | 9 | 120.0000000 | 13.3333333 | | |
| Corrected Total | 14 | 520.0000000 | | | |

| R-Square | C.V. | Root MSE | GERMINAT Mean |
|---|---|---|---|
| 0.769231 | 24.34322 | 3.651484 | 15.0000000 |

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| SOL | 1 | 52.5000000 | 52.5000000 | 3.94 | 0.0785 |
| VARIETE | 2 | 124.7340426 | 62.3670213 | 4.68 | 0.0405 |
| SOL*VARIETE | 2 | 222.7659574 | 111.3829787 | 8.35 | 0.0089 |

| Source | DF | Type II SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| SOL | 1 | 83.9007092 | 83.9007092 | 6.29 | 0.0334 |
| VARIETE | 2 | 124.7340426 | 62.3670213 | 4.68 | 0.0405 |
| SOL*VARIETE | 2 | 222.7659574 | 111.3829787 | 8.35 | 0.0089 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| SOL | 1 | 123.7714286 | 123.7714286 | 9.28 | 0.0139 |
| VARIETE | 2 | 192.1276596 | 96.0638298 | 7.20 | 0.0135 |
| SOL*VARIETE | 2 | 222.7659574 | 111.3829787 | 8.35 | 0.0089 |

Dans l'expérience sur la germination des carottes, les données ne sont pas équilibrées:

$$\text{type I} \neq \text{type III}$$

Le type II correspond au type I par permutation des facteurs (les "rotations" de MODLI ou ANVARM).

$$\begin{aligned} \text{Test habituel:} \quad H_0 &= \text{égalité des } \gamma_{ij} \\ H_0 &= \text{égalité des } \alpha_i \text{(sols)} \\ H_0 &= \text{égalité des } \beta_i \text{ (variétés)} \end{aligned}$$

$\longrightarrow$ voir les SC TYPE III

On conclut au rejet de chacune de ces trois hypothèses.

# ANALYSE DE VARIANCE AVEC SAS DANS LE CAS EQUIREPETE

## Exempe Tournesol

General Linear Models Procedure

Dependent Variable: HUILE

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 5 | 47.1814417 | 9.4362883 | 6.18 | 0.0233 |
| Error | 6 | 9.1666500 | 1.5277750 | | |
| Corrected Total | 11 | 56.3480917 | | | |

| R-Square | C.V. | Root MSE | HUILE Mean |
|---|---|---|---|
| 0.837321 | 2.671010 | 1.23603 | 46.275833 |

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| TESTEUR | 1 | 9.4874083 | 9.4874083 | 6.21 | 0.0470 |
| ORIGINE | 2 | 33.7458167 | 16.8729083 | 11.04 | 0.0097 |
| TESTEUR*ORIGINE | 2 | 3.9482167 | 1.9741083 | 1.29 | 0.3415 |

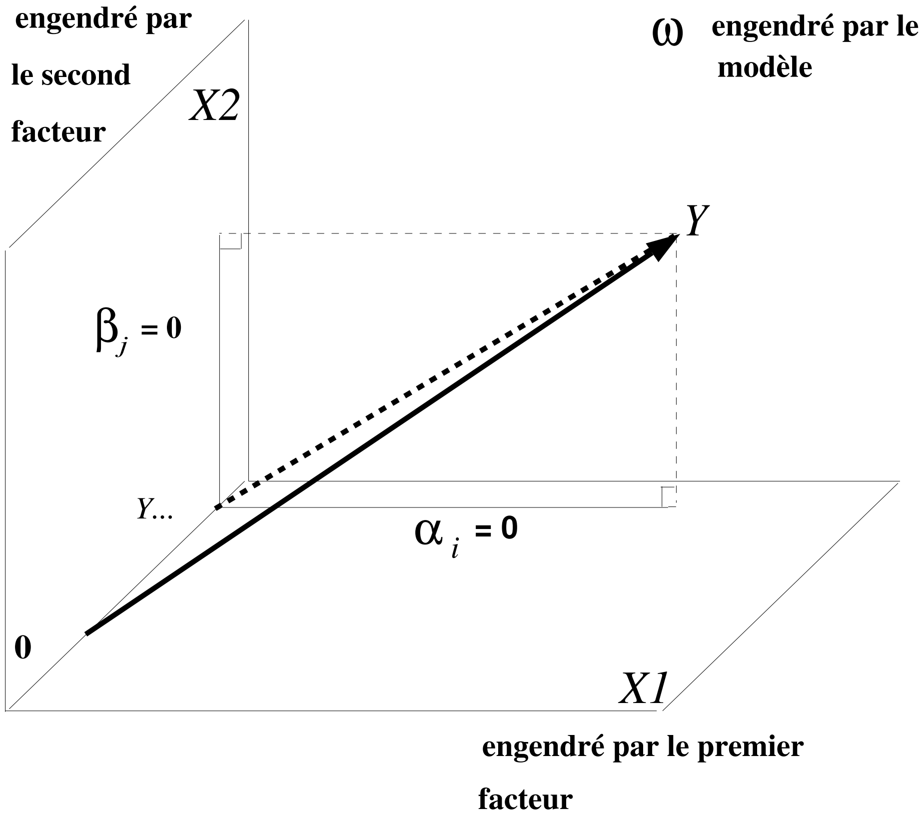
| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| TESTEUR | 1 | 9.4874083 | 9.4874083 | 6.21 | 0.0470 |
| ORIGINE | 2 | 33.7458167 | 16.8729083 | 11.04 | 0.0097 |
| TESTEUR*ORIGINE | 2 | 3.9482167 | 1.9741083 | 1.29 | 0.3415 |

**Equirépétition:**
L'ordre des facteurs n'influe pas sur les estimations.

$\Longrightarrow$ même résultats du type I et du type III.

Pour un risque de 5%, on conclut à une différence entre les origines d'une part et entre les deux testeurs d'autre part.

# CAS EQUIREPETE : REPRESENTATION GEOMETRIQUE



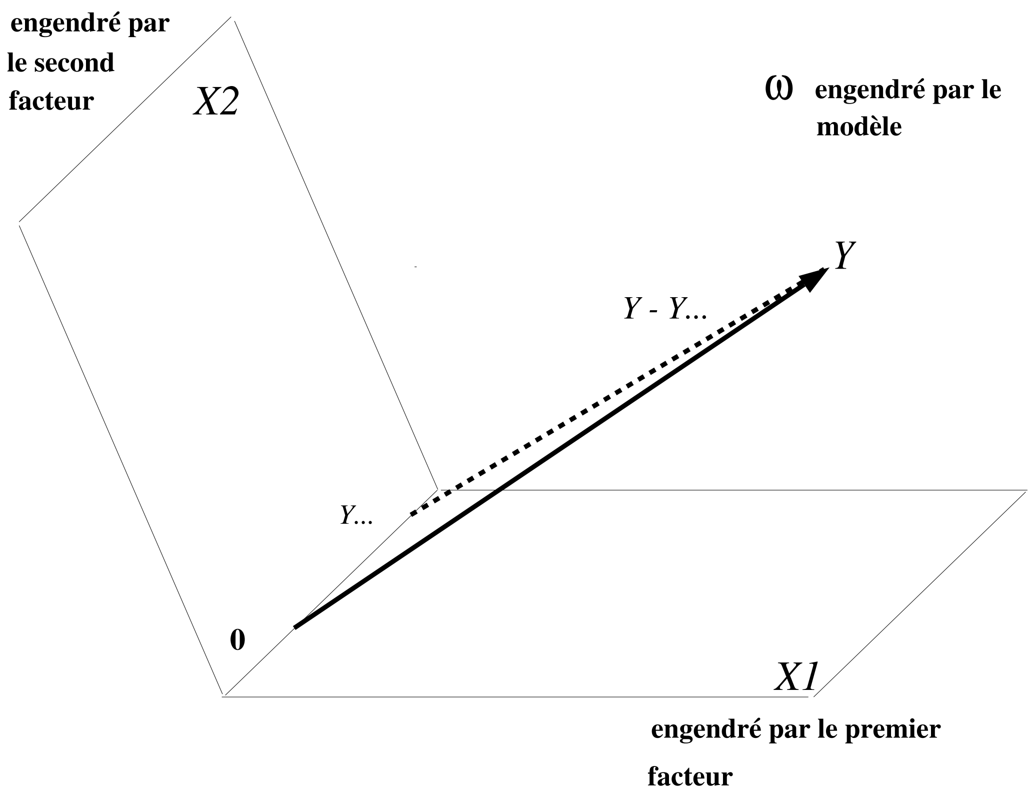

$$EQUIREPETITION \begin{cases} ORTHOGONALITE \\ DECOMPOSITION\ \ UNIQUE \end{cases}$$

Même estimation de la somme des carrés d'un facteur, que l'on projette
d'abord sur le "plan" engendré par ce facteur
ou non

$\Downarrow$

Décomposition unique

# CAS NON-EQUIREPETE : REPRESENTATION GEOMETRIQUE



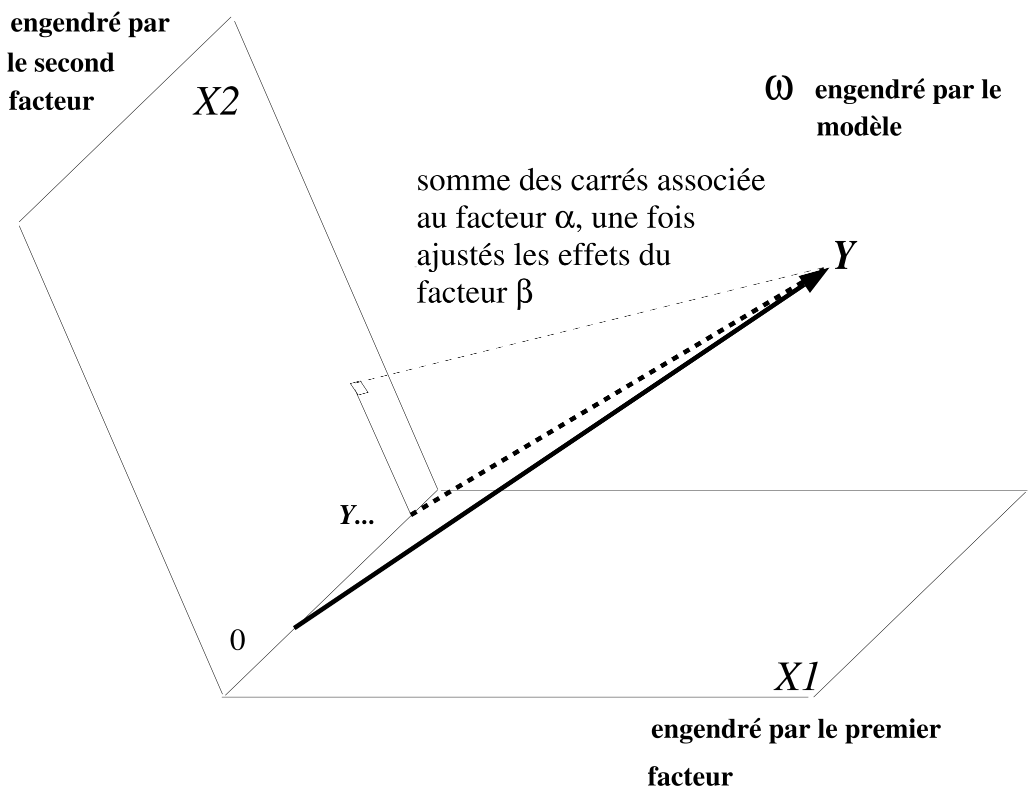
- $\boxed{\textbf{Sans rabat:}}$ $\overrightarrow{O\widehat{Y}}$ se situe dans l'espace engendré par le modèle ($\overrightarrow{OY}$ est hors de cet espace).
  $\overrightarrow{OY_{...}}$ est le terme de centrage.

- **Rabat gauche:** Somme des carrés du facteur $\alpha$ ajustée des effets de $\beta$.
  Projection sur $X_2$.
$$\Rightarrow \left\|0\widehat{Y}\right\| = \|0Y_{...}\| + SC_{n.a.}(\beta) + SC_{a.}(\alpha)$$

- **Rabat droit:** Somme des carrés du facteur $\beta$ ajustée des effets de $\alpha$.
  Projection sur $X_1$.
$$\Rightarrow \left\|0\widehat{Y}\right\| = \|0Y_{...}\| + SC_{n.a.}(\alpha) + SC_{a.}(\beta)$$

Il y a deux décompositions

En général, la non-equirépétition conduit à la non-orthogonalité.


engendré par
le second
facteur
X2
ω engendré par le
modèle
somme des carrés associée
au facteur α, une fois
ajustés les effets du
facteur β
Y
Y...
0
X1
engendré par le premier
facteur

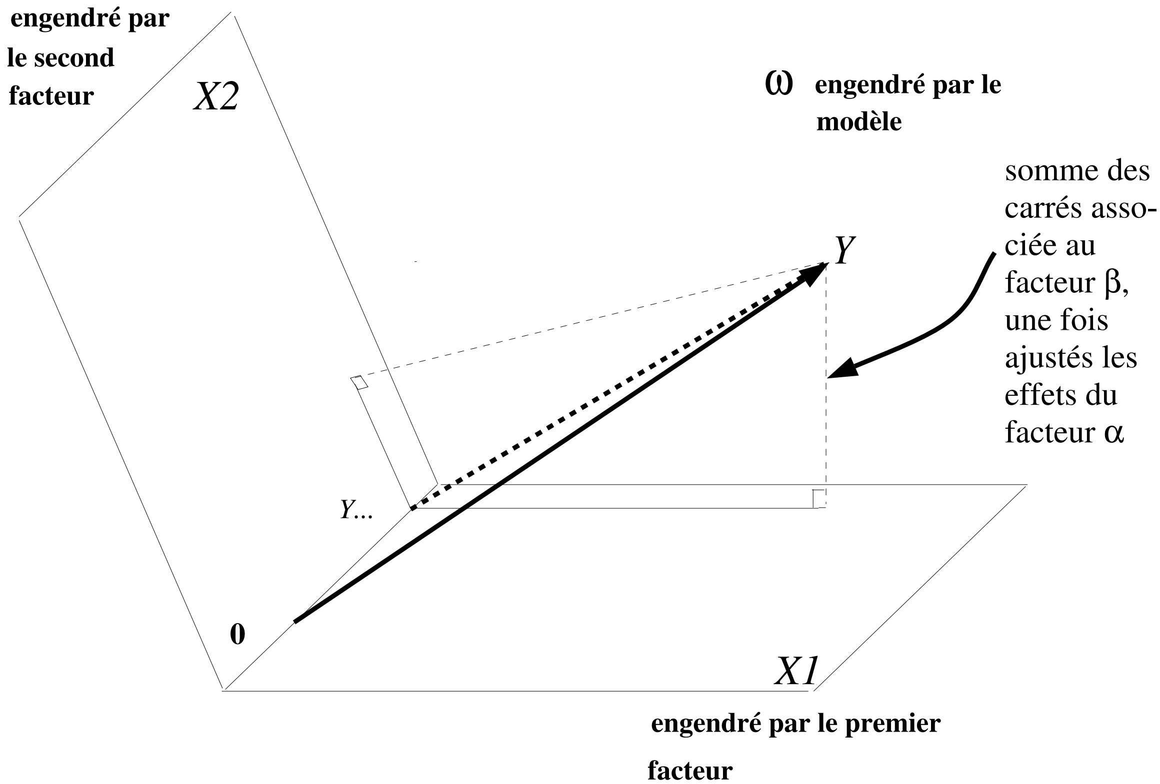
- **Sans rabat:** $\overrightarrow{O\widehat{Y}}$ se situe dans l'espace engendré par le modèle ($\overrightarrow{OY}$ est hors de cet espace).
  $\overrightarrow{OY}_{...}$ est le terme de centrage.

- $\boxed{\textbf{Rabat gauche:}}$ Somme des carrés du facteur $\alpha$ ajustée des effets de $\beta$.
  Projection sur $X_2$.
  $$\Rightarrow \left\|0\widehat{Y}\right\| = \|0Y_{...}\| + SC_{n.a.}(\beta) + SC_{a.}(\alpha)$$

- **Rabat droit:** Somme des carrés du facteur $\beta$ ajustée des effets de $\alpha$.
  Projection sur $X_1$.
  $$\Rightarrow \left\|0\widehat{Y}\right\| = \|0Y_{...}\| + SC_{n.a.}(\alpha) + SC_{a.}(\beta)$$

Il y a deux décompositions

En général, la non-equirépétition conduit à la non-orthogonalité.


engendré par
le second
facteur
X2
ω engendré par le
modèle
somme des
carrés asso-
ciée au
facteur β,
une fois
ajustés les
effets du
facteur α
Y
Y...
0
X1
engendré par le premier
facteur

- **Sans rabat:** $\overrightarrow{O\widehat{Y}}$ se situe dans l'espace engendré par le modèle ($\overrightarrow{OY}$ est hors de cet espace).
  $\overrightarrow{OY}_{...}$ est le terme de centrage.

- **Rabat gauche:** Somme des carrés du facteur $\alpha$ ajustée des effets de $\beta$.
  Projection sur $X_2$.

$$\Rightarrow \left\| 0\widehat{Y} \right\| = \|0Y_{...}\| + SC_{n.a.}(\beta) + SC_{a.}(\alpha)$$

- $\boxed{\textbf{Rabat droit:}}$ Somme des carrés du facteur $\beta$ ajustée des effets de $\alpha$.
  Projection sur $X_1$.

$$\Rightarrow \left\| 0\widehat{Y} \right\| = \|0Y_{...}\| + SC_{n.a.}(\alpha) + SC_{a.}(\beta)$$

Il y a deux décompositions

En général, la non-equirépétition conduit à la non-orthogonalité.

## DIFFERENTS TYPES DE CONDITIONS SUPPLEMENTAIRES

**Sur les effets principaux :**

$\sum_i \alpha_i = 0$ $\qquad$ $\sum_j \beta_j = 0$

$\sum_i n_{i+} \alpha_i = 0$ $\qquad$ $\sum_j n_{+j}\ \beta_j = 0$

$\alpha_I = 0$ $\qquad$ $\beta_J = 0$

**S'il y a des interactions :**

$\sum_j \gamma_{ij} = 0$ pour tout $i$ $\qquad$ $\sum_i \gamma_{ij} = 0$ pour tout $j$

$\sum_j n_{i+} \gamma_{ij} = 0$ " $\qquad$ $\sum_i n_{+j}\ \gamma_{ij} = 0$ "

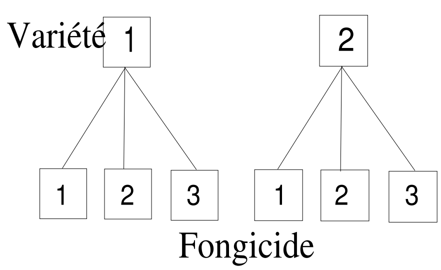
$\sum_j n_{ij} \gamma_{ij} = 0$ " $\qquad$ $\sum_i n_{ij} \gamma_{ij} = 0$ "

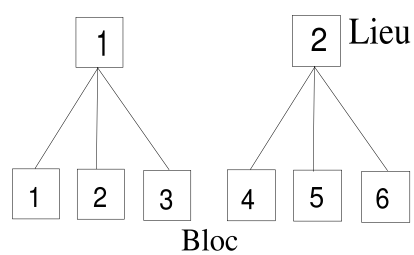
- Choix habituel (cf. l'exemple des carottes).

- Option prise dans Amance

- Option prise par MODLI et SAS
  (Dans certains cas, c'est l'effet du premier niveau des facteurs qui peut être mis à zéro).

| FACTEURS CROISES | FACTEURS HIERARCHISES |
|---|---|
| variété × lieu | bloc dans lieu |
| variété × fongicide | sous-bloc dans bloc |
| variété × testeur | n° de descendant dans famille génétique |
| concentration × température | n° de lapin dans n° de portée dans n° de lapine |
| Chaque fongicide appliqué sur chaque variété (le fongicide 1 est le même pour toutes les variétés) | Pas de correspondance entre le bloc 1 d'un lieu à l'autre |



- *(moitié droite cachée)*

  Les modèles d'analyse de variance rencontrés jusqu'ici s'appliquent à des facteurs dits **croisés**:

  – Chaque modalité (ou niveau) d'un facteur est présente pour l'ensemble des modalités des autres facteurs.

- *(moitié droite découverte)*

  Il existe des expériences dans lesquelles les facteurs sont dits **hierarchisés** (ou emboités).

  – Une modalité donnée d'un facteur hiérarchisé n'est présente que pour une seule des modalités du facteur hiérarchisant.

  La numérotation des niveaux du facteur hiérarchisé peut être arbitraire.

## MODELE D'ANALYSE DE VARIANCE A DEUX FACTEURS HIERARCHISES

2 facteurs A et B
Facteur B *emboîté* dans le facteur A
B= bloc, A= lieu

$$\boxed{Y_{ijk} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk}}$$

$Y$ réponse observée
$\mu$ moyenne de $Y$
$\alpha_i$ effet du niveau $i$ de A
$\beta_{ij}$ effet du niveau $j$ de B pour le niveau $i$ de A ($\beta_{11}$ : effet du bloc 1 pour le lieu 1)
$\varepsilon_{ijk}$ erreur

Ecriture symbolique
SAS Y=A B(A)
MODLI Y=MU + A + A.B + EPS

**Vocabulaire:**

emboité (dans) = hiérarchisé (à)

**L'effet propre** du facteur B n'existe pas, il est pris en compte uniquement par son **interaction** avec A.

($\beta_{ij}$ est l'effet de B dans A ou interaction B.A)

Indiquer au logiciel que l'on est dans ce cas:

- SAS a une syntaxe particulière.
  (mais on peut aussi écrire: $Y$ =A A$*$B).
- MODLI n'a pas de syntaxe particulière, mais il ne faut pas indiquer l'effet propre de B.

# MODELES HIERARCHISES : ANALYSE DE VARIANCE

Dependent Variable: PRODLAIT

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 20 | 1018.85333 | 50.94267 | 41.52 | 0.0001 |
| Error | 21 | 25.76500 | 1.22690 | | |
| Corrected Total | 41 | 1044.61833 | | | |

| R-Square | C.V. | Root MSE | PRODLAIT Mean |
|---|---|---|---|
| 0.975335 | 3.629680 | 1.10766 | 30.516667 |

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| FOURRAGE | 2 | 104.617619 | 52.3088 | 42.63 | 0.0001 |
| VACHE(FOURRAGE) | 18 | 914.235714 | 50.7908 | 41.40 | 0.0001 |

*Exemple:*
On étudie la production laitière de vaches consommant trois fourrages différents.
Il y a:

- 3 lots de 7 vaches,
- 2 répétitions par animal.

*Schéma à reproduire au tableau:*

| FOURRAGE | VACHE |
|---|---|
| 1 | 1 |
| | 2 |
| | 3 |
| | 4 |
| | 5 |
| | 6 |
| | 7 |
| 2 | 8 |
| | 9 |
| | $\vdots$ |
| | 14 |
| 3 | 15 |
| | 16 |
| | $\vdots$ |
| | 21 |

ddl (VACHE) = 3 lots $\times$ (7-1) vaches = 18

On conclut à l'existence:

- de différences entre les fourrages,
- de différences entre les animaux pour un fourrage donné.

# CAS PARTICULIER : ECHANTILLONS D'UNE SEULE OBSERVATION

- La somme des carrés des écarts totale se confond avec la somme factorielle totale :

$$\widehat{\theta_{ij}} = Y_{ij}$$

- Le modèle s'écrit donc :

$$Y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

- Il n'est jamais possible de tester l'interaction puisqu'il n'y a pas de base de comparaison (la SC résiduelle est nulle).
- Les effets principaux ne peuvent être testés qu'en l'**absence d'interaction** (modèle additif).