



# Analyse de la variance

David Causeur

*Laboratoire de Mathématiques Appliquées*

*Agrocampus Ouest*

*IRMAR CNRS UMR 6625*

*<http://www.agrocampus-ouest.fr/math/causeur/>*

# Plan de la présentation

- 1 Introduction
- 2 Approche exploratoire
  - Graphiques
  - Rapport de corrélation
- 3 Approche inférentielle
  - Le modèle d'analyse de la variance à un facteur
  - Test de l'effet d'un facteur
  - Estimation des paramètres
  - Prise en compte d'un effet « individu »
- 4 Bilan et perspectives



# Problématique

- L'appréciation globale diffère-t-elle d'un produit à un autre ?



# Problématique

- L'appréciation globale diffère-t'elle d'un produit à un autre ?
- Les variétés ont-elles toutes le même rendement ?

# Problématique

- L'appréciation globale diffère-t-elle d'un produit à un autre ?
- Les variétés ont-elles toutes le même rendement ?
- Les modes de consommation se différencient-ils selon la classe d'âge ?
- ...

# Problématique

- L'appréciation globale diffère-t'elle d'un produit à un autre ?
- Les variétés ont-elles toutes le même rendement ?
- Les modes de consommation se différencient-ils selon la classe d'âge ?
- ...

**Etude de l'effet d'une **variable qualitative (facteur)** sur une **variable quantitative****



## Différences entre génotypes

|   | $x = \text{Type génétique}$<br>(4 modalités) | $y = \text{Teneur en viande}$<br>maigre (kg/quintal) |
|---|--|--|
| 1 | LWP  | 62.70  |
| 2 | LWP  | 57.38  |
| 3 | LWP  | 53.41  |
| 4 | PAL  | 60.09  |
| 5 | PAL  | 59.43  |
| 6 | LWLF   | 54.95  |
| 7 | P  | 68.00  |
| 8 | P  | 63.54  |
| 9 | P  | 64.45  |
| ⋮ | ⋮  | ⋮  |

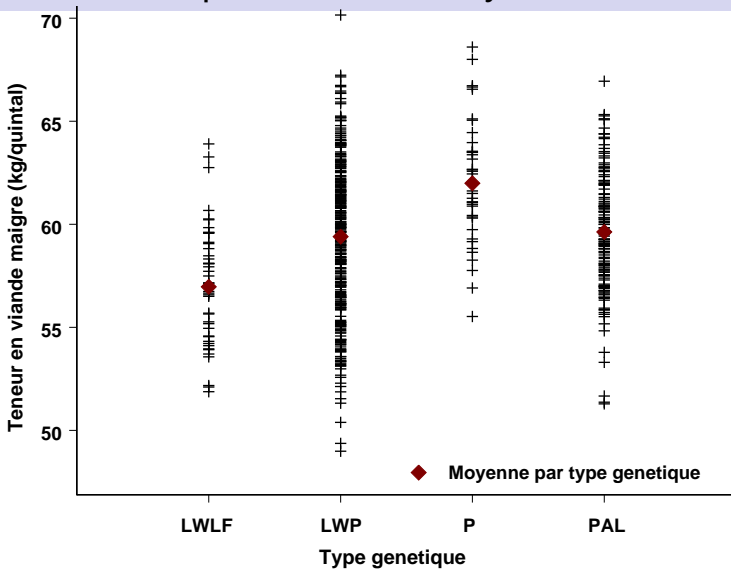


# Plan de la présentation

- 1 Introduction
- 2 Approche exploratoire
  - Graphiques
  - Rapport de corrélation
- 3 Approche inférentielle
  - Le modèle d'analyse de la variance à un facteur
  - Test de l'effet d'un facteur
  - Estimation des paramètres
  - Prise en compte d'un effet « individu »
- 4 Bilan et perspectives

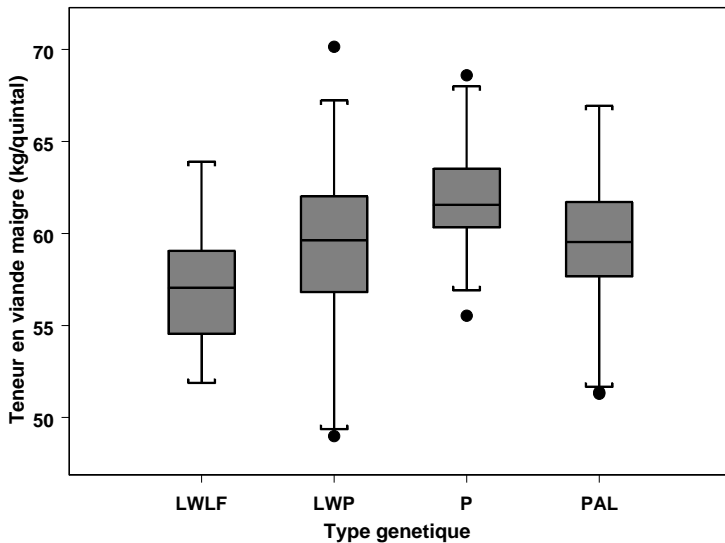


## Comparaison des moyennes



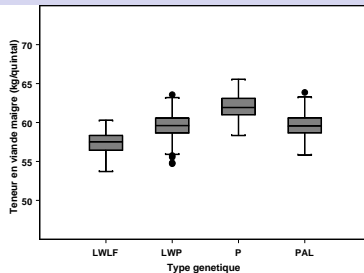
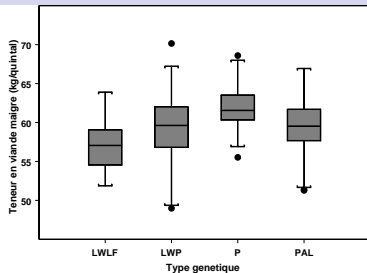


# Boîtes à moustaches



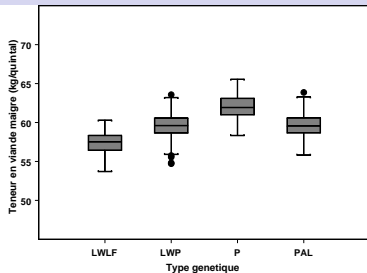
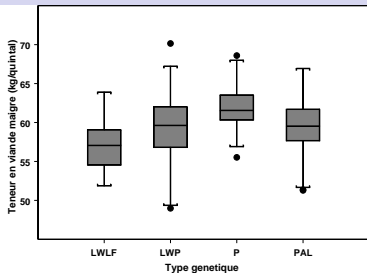


# Intensité du lien





## Intensité du lien

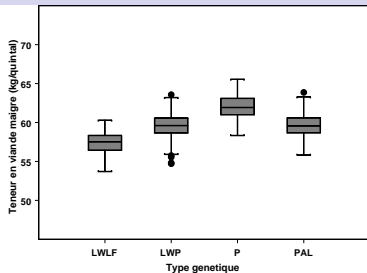
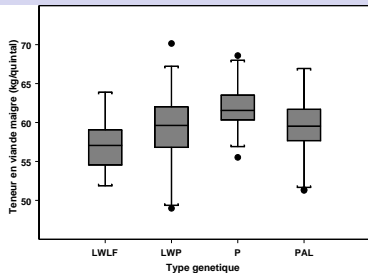


**Intuition :**

le lien *Teneur en viande maigre* × *Type génétique*  
est plus fort  
à qu'à



## Intensité du lien



**Intuition :**

le lien *Teneur en viande maigre* × *Type génétique*  
est plus fort  
à droite qu'à gauche



## Variabilité non-imputable au facteur

|               | Données                            | Moyenne | Écart – type |
|---------------|------------------------------------|---------|--------------|
| 1ère modalité | $y_{11}, y_{12}, \dots, y_{1,n_1}$ |         |              |
| 2ème modalité | $y_{21}, y_{22}, \dots, y_{2,n_2}$ |         |              |
| ⋮             | ⋮                                  |         |              |
| lème modalité | $y_{l1}, y_{l2}, \dots, y_{l,n_l}$ |         |              |

## Variabilité non-imputable au facteur

|               | Données                            | Moyenne  | Écart – type |
|---------------|------------------------------------|--|--------------|
| 1ère modalité | $y_{11}, y_{12}, \dots, y_{1,n_1}$ | $y_{1\bullet}$   |              |
| 2ème modalité | $y_{21}, y_{22}, \dots, y_{2,n_2}$ | $y_{2\bullet}$   |              |
| ⋮             | ⋮                                  |  |              |
| lème modalité | $y_{l1}, y_{l2}, \dots, y_{l,n_l}$ | $y_{l\bullet}$   |              |
| Globalement   |                                    | $y_{\bullet\bullet} = \sum_{i=1}^l \frac{n_i}{n} y_{i\bullet}$ |              |



## Variabilité non-imputable au facteur

|               | Données                            | Moyenne  | Écart – type |
|---------------|------------------------------------|--|--------------|
| 1ère modalité | $y_{11}, y_{12}, \dots, y_{1,n_1}$ | $y_{1\bullet}$   | $s'_1$       |
| 2ème modalité | $y_{21}, y_{22}, \dots, y_{2,n_2}$ | $y_{2\bullet}$   | $s'_2$       |
| ⋮             | ⋮                                  |  |              |
| lème modalité | $y_{l1}, y_{l2}, \dots, y_{l,n_l}$ | $y_{l\bullet}$   | $s'_l$       |
| Globalement   |                                    | $y_{\bullet\bullet} = \sum_{i=1}^l \frac{n_i}{n} y_{i\bullet}$ |              |



## Variabilité non-imputable au facteur

|               | Données                            | Moyenne  | Écart – type |
|---------------|------------------------------------|--|--------------|
| 1ère modalité | $y_{11}, y_{12}, \dots, y_{1,n_1}$ | $y_{1\bullet}$   | $s'_1$       |
| 2ème modalité | $y_{21}, y_{22}, \dots, y_{2,n_2}$ | $y_{2\bullet}$   | $s'_2$       |
| ⋮             | ⋮                                  |  |              |
| lème modalité | $y_{l1}, y_{l2}, \dots, y_{l,n_l}$ | $y_{l\bullet}$   | $s'_l$       |
| Globalement   |                                    | $y_{\bullet\bullet} = \sum_{i=1}^l \frac{n_i}{n} y_{i\bullet}$ |              |

$$SC_{\text{Intra}} = \sum_{j=1}^{n_1} (y_{1j} - y_{1\bullet})^2 + \sum_{j=1}^{n_2} (y_{2j} - y_{2\bullet})^2 + \dots + \sum_{j=1}^{n_l} (y_{lj} - y_{l\bullet})^2$$

## Variabilité non-imputable au facteur

|               | Données                            | Moyenne  | Écart – type |
|---------------|------------------------------------|--|--------------|
| 1ère modalité | $y_{11}, y_{12}, \dots, y_{1,n_1}$ | $y_{1\bullet}$   | $s'_1$       |
| 2ème modalité | $y_{21}, y_{22}, \dots, y_{2,n_2}$ | $y_{2\bullet}$   | $s'_2$       |
| ⋮             | ⋮                                  |  |              |
| lème modalité | $y_{l1}, y_{l2}, \dots, y_{l,n_l}$ | $y_{l\bullet}$   | $s'_l$       |
| Globalement   |                                    | $y_{\bullet\bullet} = \sum_{i=1}^l \frac{n_i}{n} y_{i\bullet}$ |              |

$$\begin{aligned}
 SC_{\text{Intra}} &= \sum_{j=1}^{n_1} (y_{1j} - y_{1\bullet})^2 + \sum_{j=1}^{n_2} (y_{2j} - y_{2\bullet})^2 + \dots + \sum_{j=1}^{n_l} (y_{lj} - y_{l\bullet})^2 \\
 &= (n_1 - 1)s_1'^2 + (n_2 - 1)s_2'^2 + \dots + (n_l - 1)s_l'^2
 \end{aligned}$$

# Variabilité imputable au facteur

|               | Données   | Moyenne | Écart – type |
|---------------|---|---------|--------------|
| 1ère modalité | $\underbrace{y_{1\bullet}, y_{1\bullet}, \dots, y_{1\bullet}}_{\times n_1}$ |         |              |
| 2ème modalité | $\underbrace{y_{2\bullet}, y_{2\bullet}, \dots, y_{2\bullet}}_{\times n_2}$ |         |              |
| ⋮             | ⋮   |         |              |
| lème modalité | $\underbrace{y_{l\bullet}, y_{l\bullet}, \dots, y_{l\bullet}}_{\times n_l}$ |         |              |



## Variabilité imputable au facteur

|               | Données   | Moyenne  | Écart – type |
|---------------|---|--|--------------|
| 1ère modalité | $\underbrace{y_{1\bullet}, y_{1\bullet}, \dots, y_{1\bullet}}_{\times n_1}$ | $y_{1\bullet}$   |              |
| 2ème modalité | $\underbrace{y_{2\bullet}, y_{2\bullet}, \dots, y_{2\bullet}}_{\times n_2}$ | $y_{2\bullet}$   |              |
| ⋮             | ⋮   |  |              |
| lème modalité | $\underbrace{y_{l\bullet}, y_{l\bullet}, \dots, y_{l\bullet}}_{\times n_l}$ | $y_{l\bullet}$   |              |
| Globalement   |   | $y_{\bullet\bullet} = \sum_{i=1}^l \frac{n_i}{n} y_{i\bullet}$ |              |



## Variabilité imputable au facteur

|               | Données   | Moyenne  | Écart – type |
|---------------|---|--|--------------|
| 1ère modalité | $\underbrace{y_{1\bullet}, y_{1\bullet}, \dots, y_{1\bullet}}_{\times n_1}$ | $y_{1\bullet}$   | 0            |
| 2ème modalité | $\underbrace{y_{2\bullet}, y_{2\bullet}, \dots, y_{2\bullet}}_{\times n_2}$ | $y_{2\bullet}$   | 0            |
| ⋮             | ⋮   |  |              |
| lème modalité | $\underbrace{y_{l\bullet}, y_{l\bullet}, \dots, y_{l\bullet}}_{\times n_l}$ | $y_{l\bullet}$   | 0            |
| Globalement   |   | $y_{\bullet\bullet} = \sum_{i=1}^l \frac{n_i}{n} y_{i\bullet}$ |              |



## Variabilité imputable au facteur

|               | Données   | Moyenne  | Écart – type |
|---------------|---|--|--------------|
| 1ère modalité | $\underbrace{y_{1\bullet}, y_{1\bullet}, \dots, y_{1\bullet}}_{\times n_1}$ | $y_{1\bullet}$   | 0            |
| 2ème modalité | $\underbrace{y_{2\bullet}, y_{2\bullet}, \dots, y_{2\bullet}}_{\times n_2}$ | $y_{2\bullet}$   | 0            |
| ⋮             | ⋮   |  |              |
| lème modalité | $\underbrace{y_{l\bullet}, y_{l\bullet}, \dots, y_{l\bullet}}_{\times n_l}$ | $y_{l\bullet}$   | 0            |
| Globalement   |   | $y_{\bullet\bullet} = \sum_{i=1}^l \frac{n_i}{n} y_{i\bullet}$ |              |

$$SC_{\text{Inter}} = \sum_{j=1}^{n_1} (y_{1\bullet} - y_{\bullet\bullet})^2 + \sum_{j=1}^{n_2} (y_{2\bullet} - y_{\bullet\bullet})^2 + \dots + \sum_{j=1}^{n_l} (y_{l\bullet} - y_{\bullet\bullet})^2,$$



## Variabilité imputable au facteur

|               | Données   | Moyenne  | Écart – type |
|---------------|---|--|--------------|
| 1ère modalité | $\underbrace{y_{1\bullet}, y_{1\bullet}, \dots, y_{1\bullet}}_{\times n_1}$ | $y_{1\bullet}$   | 0            |
| 2ème modalité | $\underbrace{y_{2\bullet}, y_{2\bullet}, \dots, y_{2\bullet}}_{\times n_2}$ | $y_{2\bullet}$   | 0            |
| ⋮             | ⋮   |  |              |
| lème modalité | $\underbrace{y_{l\bullet}, y_{l\bullet}, \dots, y_{l\bullet}}_{\times n_l}$ | $y_{l\bullet}$   | 0            |
| Globalement   |   | $y_{\bullet\bullet} = \sum_{i=1}^l \frac{n_i}{n} y_{i\bullet}$ |              |

$$\begin{aligned}
 \text{SC}_{\text{Inter}} &= \sum_{j=1}^{n_1} (y_{1\bullet} - y_{\bullet\bullet})^2 + \sum_{j=1}^{n_2} (y_{2\bullet} - y_{\bullet\bullet})^2 + \dots + \sum_{j=1}^{n_l} (y_{l\bullet} - y_{\bullet\bullet})^2, \\
 &= n_1 (y_{1\bullet} - y_{\bullet\bullet})^2 + n_2 (y_{2\bullet} - y_{\bullet\bullet})^2 + \dots + n_l (y_{l\bullet} - y_{\bullet\bullet})^2
 \end{aligned}$$



# Décomposition de la variabilité

## Équation d'analyse de la variance

$$\underbrace{\sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - y_{\bullet\bullet})^2}_{\text{SCET}} = \underbrace{\sum_{i=1}^I n_i (y_{i\bullet} - y_{\bullet\bullet})^2}_{\text{SC}_{\text{inter}}} + \underbrace{\sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - y_{i\bullet})^2}_{\text{SC}_{\text{intra}}}$$

# Décomposition de la variabilité

## Équation d'analyse de la variance

$$\underbrace{\sum_{i=1}^I \sum_{j=1}^{n_j} (y_{ij} - y_{\bullet\bullet})^2}_{\text{SCET}} = \underbrace{\sum_{i=1}^I n_i (y_{i\bullet} - y_{\bullet\bullet})^2}_{\text{SC}_{\text{inter}}} + \underbrace{\sum_{i=1}^I \sum_{j=1}^{n_j} (y_{ij} - y_{i\bullet})^2}_{\text{SC}_{\text{intra}}}$$

## Cas limites

- $\text{SC}_{\text{inter}} = 0 \Leftrightarrow y_{i\bullet} = y_{\bullet\bullet}, \forall i$
- $\text{SC}_{\text{intra}} = 0 \Leftrightarrow y_{ij} = y_{i\bullet}, \forall (i, j)$



# Rapport de corrélation

## Part de variabilité de $Y$ imputable au facteur

$$\begin{aligned}\eta^2 &= \frac{SC_{\text{inter}}}{SC_{\text{ET}}} \\ &= 1 - \frac{SC_{\text{intra}}}{SC_{\text{ET}}}\end{aligned}$$



# Rapport de corrélation

## Part de variabilité de $Y$ imputable au facteur

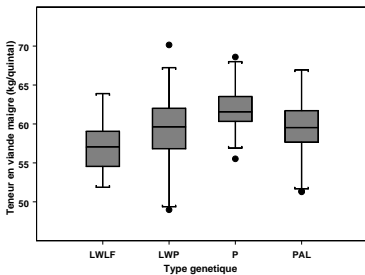
$$\begin{aligned}\eta^2 &= \frac{SC_{\text{inter}}}{SC_{\text{ET}}} \\ &= 1 - \frac{SC_{\text{intra}}}{SC_{\text{ET}}}\end{aligned}$$

### Propriétés

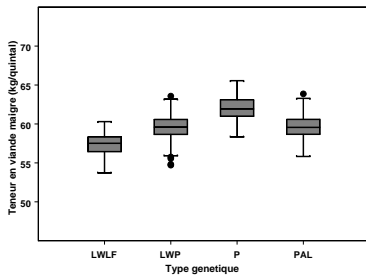
- $0 \leq \eta^2 \leq 1$ ,
- $\eta^2 = 0 \Leftrightarrow SC_{\text{inter}} = 0$
- $\eta^2 = 1 \Leftrightarrow SC_{\text{intra}} = 0$



# Exemples



$$\eta^2 = 7,8\%$$



$$\eta^2 = 26,8\%$$

# Plan de la présentation

- 1 Introduction
- 2 Approche exploratoire
  - Graphiques
  - Rapport de corrélation
- 3 Approche inférentielle
  - Le modèle d'analyse de la variance à un facteur
  - Test de l'effet d'un facteur
  - Estimation des paramètres
  - Prise en compte d'un effet « individu »
- 4 Bilan et perspectives

# Approche inférentielle

## Objectifs

- Tester l'**existence** d'une relation entre  $x$  (qualitative) et  $Y$  (quantitative)

*L'espérance de  $Y$  dépend-t-elle des modalités de  $x$  ?*

# Approche inférentielle

## Objectifs

- Tester l'**existence** d'une relation entre  $x$  (qualitative) et  $Y$  (quantitative)

*L'espérance de  $Y$  dépend-t'elle des modalités de  $x$  ?*

- Comparer les modalités du facteur

*Si une relation existe, quelles modalités de  $x$  correspondent à la plus grande espérance ?*

*à la plus faible espérance ?*



# Modèle d'analyse de la variance

$Y$  : variable à expliquer quantitative

$x$  : variable explicative qualitative (facteur)



# Modèle d'analyse de la variance

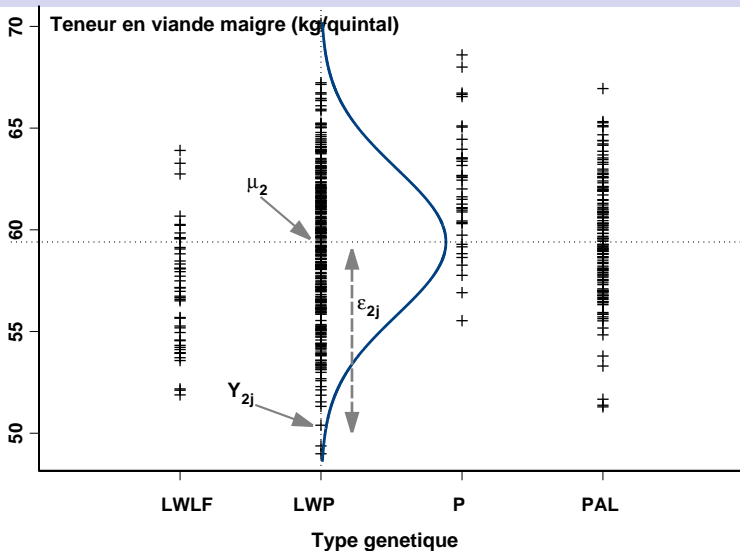
$Y$  : variable à expliquer quantitative

$x$  : variable explicative qualitative (facteur)

Pour une modalité  $i$  du facteur, les  $Y_{ij}$  sont des variables aléatoires **indépendantes**



# Modèle d'analyse de la variance



# Modèle d'analyse de la variance

$Y$  : variable à expliquer quantitative

$x$  : variable explicative qualitative (facteur)

Pour une modalité  $i$  du facteur, les  $Y_{ij}$  sont des variables aléatoires **indépendantes**

$$Y_{ij} = \underbrace{\mu_i}_{\text{Variations dues au facteur}} + \underbrace{\varepsilon_{ij}}_{\text{Variations non-dues au facteur}}$$
$$\varepsilon_{ij} \sim \mathcal{N}(0; \sigma)$$

# Paramètres du modèle

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

## Paramètres du modèle (hormis $\sigma$ )

- $I$  coefficients du modèle :  $\mu_i$

# Paramètres du modèle

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

## Paramètres du modèle (hormis $\sigma$ )

- Effet moyen  $\mu$
- $I$  effets principaux :  $\alpha_i$

# Paramètres du modèle

Effet moyen

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

## Paramètres du modèle (hormis $\sigma$ )

- Effet moyen  $\mu$
- $I$  effets principaux :  $\alpha_i$

## Paramètres du modèle

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

Effet moyen

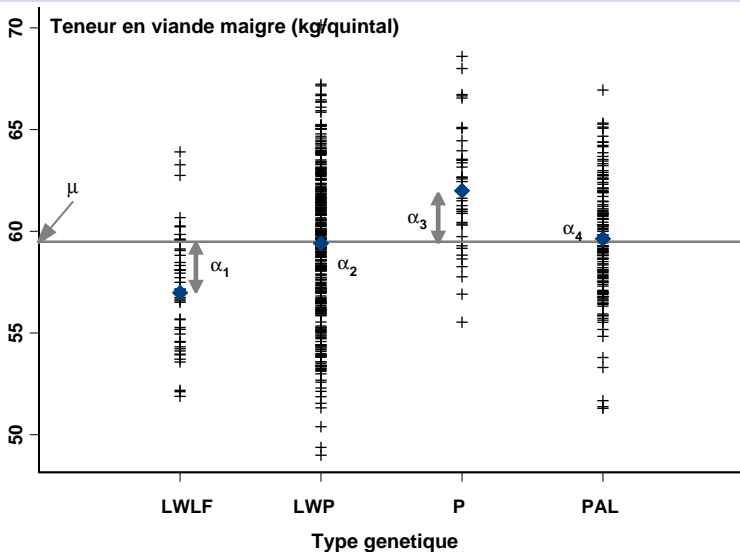
Effets principaux

### Paramètres du modèle (hormis $\sigma$ )

- Effet moyen  $\mu$
- $I$  effets principaux :  $\alpha_j$



# Paramètres du modèle



## Paramètres du modèle

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

Effet moyen

Effets principaux :  $\alpha_1 + \alpha_2 + \dots + \alpha_l = 0$

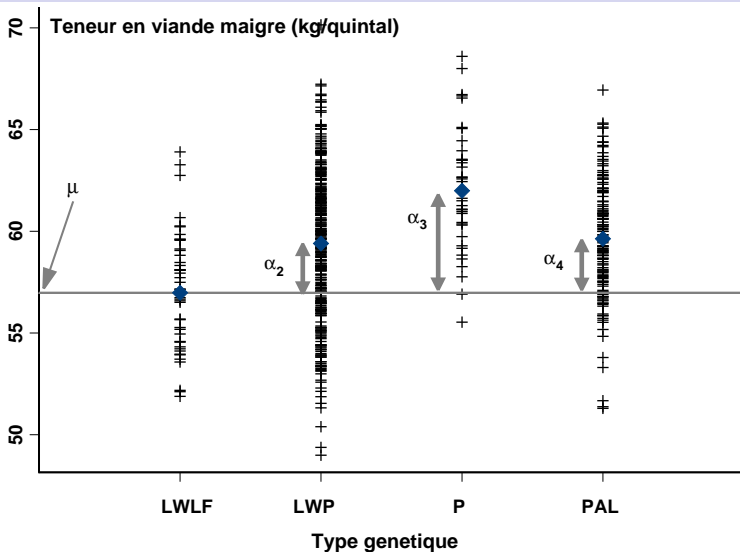
### Paramètres du modèle (hormis $\sigma$ )

- Effet moyen  $\mu$
- $l$  effets principaux :  $\alpha_j$

(en fait,  $l - 1$  paramètres indépendants)



# Paramètres du modèle



## Paramètres du modèle

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

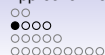
Effet moyen

Effets principaux :  $\alpha_1 = 0$

### Paramètres du modèle (hormis $\sigma$ )

- Effet moyen  $\mu$
- $I$  effets principaux :  $\alpha_j$

(en fait,  $I - 1$  paramètres indépendants)



## Effet d'un facteur

$H_0$  : Le facteur n'a pas d'effet sur  $Y$

$H_1$  : Le facteur a un effet sur  $Y$

ou encore ...

$H_0$  :  $\alpha_j = 0, \forall i$

$H_1$  :  $\exists i, \alpha_j \neq 0$

# Analyse de la variance du modèle

## Décomposition de la variabilité de $Y$

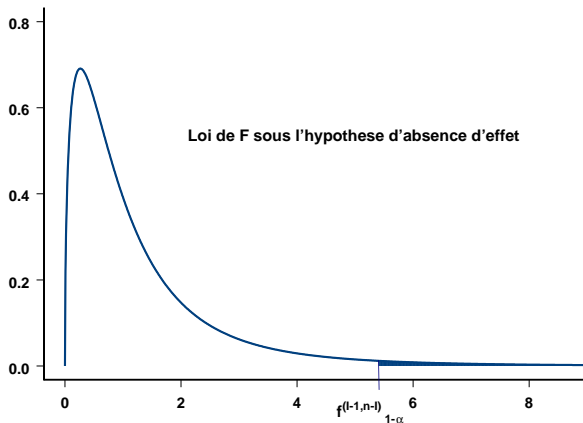
$$\underbrace{\sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - y_{\bullet\bullet})^2}_{\text{SCET}} = \underbrace{\sum_{i=1}^I n_i (y_{i\bullet} - y_{\bullet\bullet})^2}_{\text{SC}_{\text{inter}}} + \underbrace{\sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - y_{i\bullet})^2}_{\text{SC}_{\text{intra}}}$$

## Sous l'hypothèse d'absence d'effet

$$F = \frac{\frac{\text{SC}_{\text{inter}}}{I-1}}{\frac{\text{SC}_{\text{intra}}}{n-I}} = \frac{\text{CMF}}{\text{CMR}} \sim \mathcal{F}_{I-1, n-I}$$



# Loi de la statistique F sous $H_0$



# Illustration

## Table d'analyse de la variance

|                | Degrés de liberté | SC      | CM      | F     | Probabilité critique |
|----------------|-------------------|---------|---------|-------|----------------------|
| Type génétique | 3                 | 546.81  | 182.27  | 16.35 | 3.29e-010            |
| Résiduelle     | 578               | 6444.67 | 11.1499 |       |                      |



# Estimateurs des coefficients

$$\hat{\mu}_i = Y_{i\bullet}$$



# Estimateurs des coefficients

$$\hat{\mu}_i = Y_{i\bullet}$$

## Propriétés

- Non-biais :  $\mathbb{E}(\hat{\mu}_i) = \mu_i$
- Variance :  $\text{Var}(\hat{\mu}_i) = \sigma^2/n_i$

# Estimateurs des coefficients

$$\hat{\mu}_i = Y_{i\bullet}$$

## Propriétés

- Non-biais :  $\mathbb{E}(\hat{\mu}_i) = \mu_i$
- Variance :  $\text{Var}(\hat{\mu}_i) = \sigma^2/n_i$

## Variance résiduelle

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{n_1} (Y_{1j} - Y_{1\bullet})^2 + \sum_{j=1}^{n_2} (Y_{2j} - Y_{2\bullet})^2 + \dots + \sum_{j=1}^{n_l} (Y_{lj} - Y_{l\bullet})^2}{n-l}$$

# Estimateurs des coefficients

$$\hat{\mu}_i = Y_{i\bullet}$$

## Propriétés

- Non-biais :  $\mathbb{E}(\hat{\mu}_i) = \mu_i$
- Variance :  $\text{Var}(\hat{\mu}_i) = \sigma^2/n_i$

## Variance résiduelle

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{n_1} (Y_{1j} - Y_{1\bullet})^2 + \sum_{j=1}^{n_2} (Y_{2j} - Y_{2\bullet})^2 + \dots + \sum_{j=1}^{n_l} (Y_{lj} - Y_{l\bullet})^2}{n-l}$$

SC<sub>Intra</sub> ←

# Estimateurs des effets (contrainte de somme nulle)

$$\begin{array}{rcl} \hat{\mu} + \hat{\alpha}_1 & = & \hat{\mu}_1 = Y_{1\bullet}, \\ \hat{\mu} + \hat{\alpha}_2 & = & \hat{\mu}_2 = Y_{2\bullet}, \\ \vdots & & \vdots \\ \hat{\mu} + \hat{\alpha}_I & = & \hat{\mu}_I = Y_{I\bullet}, \end{array}$$

---

# Estimateurs des effets (contrainte de somme nulle)

$$\begin{aligned}
 \hat{\mu} + \hat{\alpha}_1 &= \hat{\mu}_1 = Y_{1\bullet}, \\
 \hat{\mu} + \hat{\alpha}_2 &= \hat{\mu}_2 = Y_{2\bullet}, \\
 &\vdots \\
 \hat{\mu} + \hat{\alpha}_I &= \hat{\mu}_I = Y_{I\bullet}, \\
 \hline
 \hat{\mu} + \sum_{i=1}^I \hat{\alpha}_i &= \frac{\sum_{i=1}^I Y_{i\bullet}}{\sum_{i=1}^I 1} \Rightarrow \hat{\mu} = \frac{1}{I} \sum_{i=1}^I Y_{i\bullet} \\
 &\hat{\alpha}_i = Y_{i\bullet} - \hat{\mu}
 \end{aligned}$$

# Estimateurs des effets (contrainte de somme nulle)

$$\begin{aligned}
 \hat{\mu} + \hat{\alpha}_1 &= \hat{\mu}_1 = Y_{1\bullet}, \\
 \hat{\mu} + \hat{\alpha}_2 &= \hat{\mu}_2 = Y_{2\bullet}, \\
 &\vdots \\
 \hat{\mu} + \hat{\alpha}_I &= \hat{\mu}_I = Y_{I\bullet}, \\
 \hline
 \hat{\mu} + \sum_{i=1}^I \hat{\alpha}_i &= \frac{\hat{\mu}_I = Y_{I\bullet}}{\sum_{i=1}^I Y_{i\bullet}} \Rightarrow \hat{\mu} = \frac{1}{I} \sum_{i=1}^I Y_{i\bullet} \\
 &\hat{\alpha}_i = Y_{i\bullet} - \hat{\mu}
 \end{aligned}$$

**Lorsque**  $n_1 = n_2 = \dots = n_I$ , alors

$$\begin{aligned}
 \hat{\mu} &= Y_{\bullet\bullet} \\
 \hat{\alpha}_i &= Y_{i\bullet} - Y_{\bullet\bullet}
 \end{aligned}$$

# Illustration

|             | Coefficients | Écart-type | Statistique t | Probabilité critique |
|-------------|--------------|------------|---------------|----------------------|
| Effet moyen | 59.49        | 0.20       | 300.37        | 0.00                 |
| LWLF        | -2.53        | 0.40       | -6.31         | 0.00                 |
| LWP         | -0.09        | 0.24       | -0.40         | 0.69                 |
| P           | 2.49         | 0.42       | 5.90          | 0.00                 |
| PAL         | 0.13         | 0.28       | 0.47          | 0.64                 |

**Écart-type résiduel : 3.34**

$$R^2 = 7.8\%$$



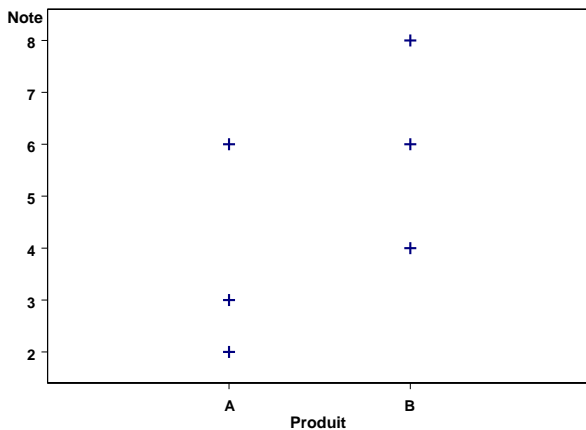
## Exemple « analyse sensorielle »

**Données brutes** : données appariées

| Produits | Juges          |                |                |
|----------|----------------|----------------|----------------|
|          | J <sub>1</sub> | J <sub>2</sub> | J <sub>3</sub> |
| A        | 2              | 3              | 6              |
| B        | 4              | 6              | 8              |



## Exemple « analyse sensorielle »



## Exemple « analyse sensorielle »

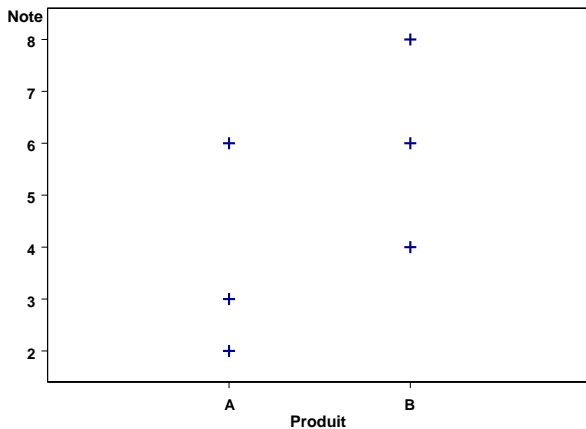
**Données brutes** : données appariées

| Produits | Juges          |                |                |
|----------|----------------|----------------|----------------|
|          | J <sub>1</sub> | J <sub>2</sub> | J <sub>3</sub> |
| A        | 2              | 3              | 6              |
| B        | 4              | 6              | 8              |

|            | Degrés de liberté | Sommes des carrés | Carrés moyens | F    | Probabilité critique |
|------------|-------------------|-------------------|---------------|------|----------------------|
| Produit    | 1                 | 8.17              | 8.17          | 1.96 | 0.2341006            |
| Résiduelle | 4                 | 16.67             | 4.17          |      |                      |

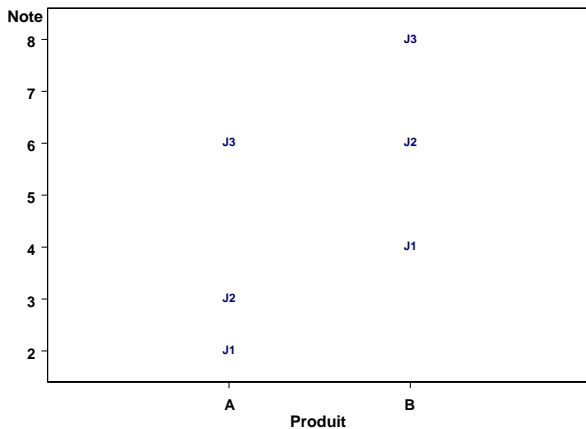


## Prise en compte de l'effet « juge »





## Prise en compte de l'effet « juge »





# Prise en compte de l'effet « Juge »

## Données brutes

| Produits | Juges          |                |                | Moyennes |
|----------|----------------|----------------|----------------|----------|
|          | J <sub>1</sub> | J <sub>2</sub> | J <sub>3</sub> |          |
| A        | 2              | 3              | 6              | 3.67     |
| B        | 4              | 6              | 8              | 6        |
| Moyennes | 3              | 4.5            | 7              |          |

# Prise en compte de l'effet « Juge »

## Données brutes

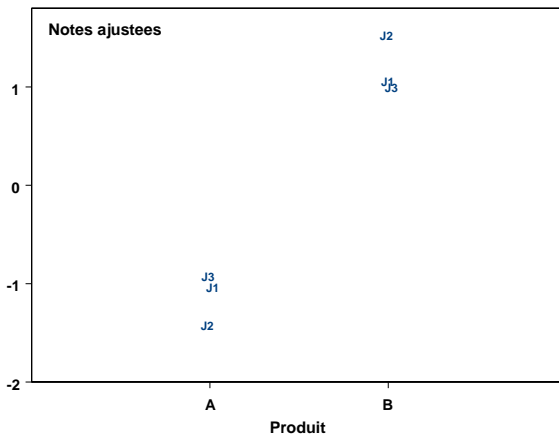
| Produits | Juges          |                |                | Moyennes |
|----------|----------------|----------------|----------------|----------|
|          | J <sub>1</sub> | J <sub>2</sub> | J <sub>3</sub> |          |
| A        | 2              | 3              | 6              | 3.67     |
| B        | 4              | 6              | 8              | 6        |
| Moyennes | 3              | 4.5            | 7              |          |

## Données ajustées de l'effet « juge »

| Produits | Juges          |                |                | Moyennes |
|----------|----------------|----------------|----------------|----------|
|          | J <sub>1</sub> | J <sub>2</sub> | J <sub>3</sub> |          |
| A        | -1             | -1.5           | -1             | -1.17    |
| B        | 1              | 1.5            | 1              | 1.17     |
| Moyennes | 0              | 0              | 0              |          |



## Données ajustées de l'effet « Juge »



## Modèle d'analyse de la variance à 2 facteurs

Pour la  $i$ ème modalité du 1er facteur et la  $j$ ème modalité du 2ème facteur

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0; \sigma),$$
$$\sum_{i=1}^I \alpha_i = 0, \quad \sum_{j=1}^J \beta_j = 0$$

**Remarque :**

- IJ données,
- $1 + (I - 1) + (J - 1) = I + J - 1$  paramètres

# Décomposition de la variabilité

$$\underbrace{\sum_{i=1}^I \sum_{j=1}^J (y_{ij} - y_{\bullet\bullet})^2}_{\text{SCET}} = \underbrace{\sum_{i=1}^I J (y_{i\bullet} - y_{\bullet\bullet})^2}_{\text{SC}_{1\text{er}} \text{ facteur}} + \underbrace{\sum_{j=1}^J I (y_{\bullet j} - y_{\bullet\bullet})^2}_{\text{SC}_{2\text{ème}} \text{ facteur}} + \underbrace{\sum_{i=1}^I \sum_{j=1}^J (y_{ij} - y_{i\bullet} - y_{\bullet j} + y_{\bullet\bullet})^2}_{\text{SC}_{\text{résiduelle}}}$$

# Tests des effets

## Effet de 1er facteur

$$F_1 = \frac{SC_{1er\ facteur}}{I-1} \bigg/ \frac{SC_{résiduelle}}{IJ-I-J+1}$$

$$F_1 \sim \text{si pas d'effet } \mathcal{F}_{I-1, IJ-I-J+1}$$

# Tests des effets

## Effet de 1er facteur

$$F_1 = \frac{SC_{1er\ facteur}}{I-1} \bigg/ \frac{SC_{résiduelle}}{IJ-I-J+1}$$

$$F_1 \sim_{\text{si pas d'effet}} \mathcal{F}_{I-1, IJ-I-J+1}$$

## Effet de 2ème facteur

$$F_2 = \frac{SC_{2ème\ facteur}}{J-1} \bigg/ \frac{SC_{résiduelle}}{IJ-I-J+1}$$

$$F_2 \sim_{\text{si pas d'effet}} \mathcal{F}_{J-1, IJ-I-J+1}$$



# Illustration

|            | Degrés de liberté | Sommes des carrés | Carrés moyens | F  | Probabilité critique |
|------------|-------------------|-------------------|---------------|----|----------------------|
| Juge       | 2                 | 16.33             | 8.17          | 49 | 0.0200               |
| Produit    | 1                 | 8.17              | 8.17          | 49 | 0.0198               |
| Résiduelle | 2                 | 0.33              | 0.17          |    |                      |



# Estimateurs des effets

$$\hat{\mu} = Y_{..}$$

$$\hat{\alpha}_j = Y_{j\bullet} - Y_{..}$$

$$\hat{\beta}_j = Y_{\bullet j} - Y_{..}$$

# Estimateurs des effets

$$\hat{\mu} = Y_{\bullet\bullet}$$

$$\hat{\alpha}_i = Y_{i\bullet} - Y_{\bullet\bullet}$$

$$\hat{\beta}_j = Y_{\bullet j} - Y_{\bullet\bullet}$$

## Valeurs ajustées et Résidus

$$\begin{aligned}\hat{Y}_{ij} &= \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j \\ &= Y_{i\bullet} + Y_{\bullet j} - Y_{\bullet\bullet}\end{aligned}$$

$$\hat{\epsilon}_{ij} = Y_{ij} - Y_{i\bullet} - Y_{\bullet j} + Y_{\bullet\bullet}$$

# Estimateur de $\sigma$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^I \sum_{j=1}^J (Y_{ij} - Y_{i\bullet} - Y_{\bullet j} + Y_{\bullet\bullet})^2}{IJ - I - J + 1}$$

# Estimateur de $\sigma$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^I \sum_{j=1}^J (Y_{ij} - Y_{i\cdot} - Y_{\cdot j} + Y_{\cdot\cdot})^2}{IJ - I - J + 1}$$

SC<sub>résiduelle</sub>



## Effet produit

**Écart-type résiduel : 2.041**

|             | Coefficients | Écart-type | Statistique t | Probabilité critique |
|-------------|--------------|------------|---------------|----------------------|
| Effet moyen | 4.83         | 0.83       | 5.80          | 0.0044               |
| A           | -1.17        | 0.83       | -1.40         | 0.2341               |
| B           | 1.17         | 0.83       | -1.40         | 0.2341               |

# Effet produit

**Écart-type résiduel : 0.408**

|                | Coefficients | Écart-type | Statistique t | Probabilité critique |
|----------------|--------------|------------|---------------|----------------------|
| Effet moyen    | 4.83         | 0.17       | 29.00         | 0.0012               |
| A              | -1.17        | 0.17       | -7.0000       | 0.0198               |
| B              | 1.17         | 0.17       | -7.0000       | 0.0198               |
| J <sub>1</sub> | -1.83        | 0.24       | -7.78         | 0.0161               |
| J <sub>2</sub> | -0.33        | 0.24       | -1.41         | 0.2929               |
| J <sub>3</sub> | 2.17         | 0.24       | 9.04          | 8.3e-4               |

# Plan de la présentation

- 1 Introduction
- 2 Approche exploratoire
  - Graphiques
  - Rapport de corrélation
- 3 Approche inférentielle
  - Le modèle d'analyse de la variance à un facteur
  - Test de l'effet d'un facteur
  - Estimation des paramètres
  - Prise en compte d'un effet « individu »
- 4 Bilan et perspectives



# Bilan et perspectives

## Analyse de la variance

- Comparaison de plusieurs populations



# Bilan et perspectives

## Analyse de la variance

- Comparaison de plusieurs populations
- Prise en compte d'un facteur de variabilité *secondaire*

# Bilan et perspectives

## Analyse de la variance

- Comparaison de plusieurs populations
- Prise en compte d'un facteur de variabilité *secondaire*

## Extensions

- Hiérarchisation de plusieurs sources de variabilité (module *Modèle Linéaire et Analyse des Données*)

# Bilan et perspectives

## Analyse de la variance

- Comparaison de plusieurs populations
- Prise en compte d'un facteur de variabilité *secondaire*

## Extensions

- Hiérarchisation de plusieurs sources de variabilité (module *Modèle Linéaire et Analyse des Données*)
- Planification expérimentale (module *INSFA 4*)