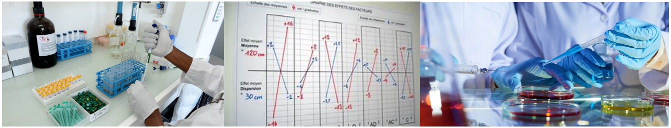
# Plans d'expériences

## M2 Chimie ICAP Arômes - Parfums - Cosmétiques

Elodie Brunel

Université de Montpellier – Faculté des Sciences

2024-2025



## Séance 2 : Analyse de la Variance

# Données - Exemple

- Analyse sensorielles de plusieurs produits (parfums) par des juges (experts)
- Score attribué à chaque produit pour différents critères (l'intensité, la durée, la qualité, la note de tête, la note de cœur et la note de fond).

| Juge | Produit | intensité | durée | qualité | $\cdots$ |
|---|---|---|---|---|---|
| J1 | P4 | 4 | 3 | 2 | $\cdots$ |
| J1 | P1 | 8 | 5 | 2 | $\cdots$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| J2 | P4 | 7 | 2 | 1 | $\cdots$ |
| J2 | P2 | 4 | 3 | 5 | $\cdots$ |

# Données - Exemple

Exemple : 2 produits, 2 juges et 2 répétitions.

| | J1 | J2 | moyenne |
|---|---|---|---|
| Prod. 1 | 4<br>5 | 3<br>4 | 4 |
| Prod. 2 | 2<br>2 | 1<br>3 | 2 |
| moyenne | 3.25 | 2.75 | 3 |

| Juge | Prod. | Score |
|---|---|---|
| J1 | P1 | 4 |
| J1 | P1 | 5 |
| J1 | P2 | 2 |
| J1 | P2 | 2 |
| J2 | P1 | 3 |
| J2 | P1 | 4 |
| J2 | P2 | 1 |
| J2 | P2 | 3 |

# Notations

- $Y$ variable réponse quantitative (Score)
- $X_1$, $X_2$ facteurs ou variables qualitatives à $I$ et $J$ modalités ou niveaux ($X_1$ : Produit 1 ou 2 ; $X_2$ : Juge 1 ou 2)
- $n_{ij}$ : nombre de répétitions pour le couple $(i, j)$.

| obs | $X_1$ : Produit $i$ | $X_2$ : Juge $j$ | $Y$ : Score $y_{ijk}$ |
|---|---|---|---|
| 1 | 1 | 1 | $y_{111} = 4$ |
| 2 | 1 | 1 | $y_{112}$=5 |
| 3 | 1 | 2 | $y_{121} = 2$ |
| 4 | 1 | 2 | $y_{122} = 2$ |
| 5 | 2 | 1 | $y_{211} = 3$ |
| 6 | 2 | 1 | $y_{212} = 4$ |
| 7 | 2 | 2 | $y_{221} = 1$ |
| 8 | 2 | 2 | $y_{222} = 3$ |

$y_{ijk}$ : observation $k$ du score du juge $i$ pour le produit $j$.

On cherche à savoir :

- s'il existe un un effet "produit" sur le score.
- s'il existe un effet "juge" sur la note.
- s'il existe une interaction entre les deux facteurs.

Interaction : l'effet d'un facteur sur $Y$ diffère selon les modalités de l'autre.

Exemple :
pas d'interaction : l'écart entre les scores de deux juges reste constant quel que soit le produit évalué.
présence d'interaction : selon le produit l'écart de scores entre deux juges varie.

# Modèle d'Analyse de la variance à deux facteurs avec interaction :

Pour $i = 1, \cdots, I$, $j = 1, \cdots, J$ et $k = 1, \cdots, n_{ij}$, le modèle s'écrit :

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

$$\begin{array}{rl} \text{avec} & \varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2) \\ \text{et} & \varepsilon_{ijk} \text{ non corrélés.} \end{array}$$

- $\mu$ moyenne globale de $Y$ (du score) quel que soit le juge ou le produit
- $\alpha_i$ effet du niveau $i$ du facteur $X_1$ (effet" produit")
- $\beta_j$ effet du niveau $j$ du facteur $X_2$ (effet "juge")
- $\gamma_{ij}$ effet de l'interaction des facteurs $X_1$ et $X_2$ sur les niveaux $i$ et $j$.
- $\varepsilon_{ijk}$ erreur du modèle de variance commnune $\sigma^2$.

Le modèle linéaire va fournir un cadre statistique qui permet de mettre en place des tests d'hypothèse pour tester les effets des facteurs.

# Modèle d'Analyse de la variance à deux facteurs avec interaction :

Ecriture du modèle pour notre exemple avec interaction :

$$\left\{\begin{array}{lllllll}
y_{111} & = & \mu & +\alpha_1 & +\beta_1 & +\gamma_{11} & +\varepsilon_{111} \\
y_{112} & = & \mu & +\alpha_1 & +\beta_1 & +\gamma_{11} & +\varepsilon_{112} \\
y_{121} & = & \mu & +\alpha_1 & +\beta_2 & +\gamma_{12} & +\varepsilon_{121} \\
y_{122} & = & \mu & +\alpha_1 & +\beta_2 & +\gamma_{12} & +\varepsilon_{122} \\
y_{211} & = & \mu & +\alpha_2 & +\beta_1 & +\gamma_{21} & +\varepsilon_{211} \\
y_{212} & = & \mu & +\alpha_2 & +\beta_1 & +\gamma_{21} & +\varepsilon_{212} \\
y_{221} & = & \mu & +\alpha_2 & +\beta_2 & +\gamma_{22} & +\varepsilon_{221} \\
y_{222} & = & \mu & +\alpha_2 & +\beta_2 & +\gamma_{22} & +\varepsilon_{222}
\end{array}\right.$$

Ecriture matricielle du modèle avec interaction :

$$\underbrace{\begin{pmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \end{pmatrix}}_{Y} = \underbrace{\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{21} \\ \gamma_{22} \end{pmatrix}}_{\beta} + \underbrace{\begin{pmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \varepsilon_{122} \\ \varepsilon_{211} \\ \varepsilon_{212} \\ \varepsilon_{221} \\ \varepsilon_{222} \end{pmatrix}}_{\varepsilon}$$

c'est-à-dire :

$$Y = X\beta + \varepsilon \text{ avec } \mathbb{E}(\varepsilon) = 0 \text{ et } \mathrm{Var}(\varepsilon) = \sigma^2 I_8$$

# Contraintes

Il y a $1 + I + J + IJ$ paramètres mais seulement $IJ$ paramètres "libres" , par ajout de contraintes pour l'identifiabilité du modèle :

effet du facteur $X_1$ :

$$\sum_{i=1}^{I} \alpha_i = 0$$

effet du facteur $X_2$ :

$$\sum_{j=1}^{J} \beta_j = 0$$

interactions :

$$\forall i, \sum_{j=1}^{J} \gamma_{ij} = 0 \text{ et } \forall j, \sum_{i=1}^{I} \gamma_{ij} = 0$$

# Contraintes sur l'exemple :

Il y a 9 paramètres dans le modèle avec les contraintes suivantes :

effet du facteur $X_1$ :

$$\alpha_1 + \alpha_2 = 0 \Longleftrightarrow \alpha_1 = -\alpha_2$$

effet du facteur $X_2$ :

$$\beta_1 + \beta_2 = 0 \Longleftrightarrow \beta_1 = -\beta_2$$

interactions :

$$\left\{ \begin{array}{lcl} \gamma_{11} + \gamma_{12} & = & 0 \\ \gamma_{21} + \gamma_{22} & = & 0 \\ \gamma_{11} + \gamma_{21} & = & 0 \\ \gamma_{12} + \gamma_{22} & = & 0 \end{array} \right. \Longleftrightarrow \left\{ \begin{array}{c} \gamma_{11} = -\gamma_{12} = -\gamma_{21} \\ \gamma_{21} = -\gamma_{22} = \gamma_{12} \end{array} \right.$$

Donc seulement 4 paramètres à estimer : $\mu$, $\alpha_1$, $\beta_1$, $\gamma_{11}$(les autres se déduisent).

Avec ces contraintes, le modèle devient :

$$\underbrace{\begin{pmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \end{pmatrix}}_{Y} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \gamma_{11} \end{pmatrix}}_{\beta} + \underbrace{\begin{pmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \varepsilon_{122} \\ \varepsilon_{211} \\ \varepsilon_{212} \\ \varepsilon_{221} \\ \varepsilon_{222} \end{pmatrix}}_{\varepsilon}$$

# Estimation des paramètres du modèle

On cherche à minimiser la somme des carrés des erreurs du modèle ;

$$\sum_{i,j,k} \varepsilon_{ijk} = (Y - X\beta)'(Y - X\beta)$$

La solution de ce problème est obtenue en résolvant :

$$X'Y = X'X\beta$$

- si $X'X$ est une matrice inversible, l'estimateur des moindres carrés de $\beta$ est :

$$\hat{\beta} = (X'X)^{-1}XY$$

- Propriétés : $\mathbb{E}(\hat{\beta}) = \beta$ (sans biais), $\mathrm{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$ et $\hat{\beta} - \beta \sim \mathcal{N}(0, \sigma^2(X'X)^{-1})$.

# Calcul des estimateurs des paramètres du modèle

On note :

– moyenne globale $\bar{Y} = \frac{1}{n}\sum_{i,j} y_{ij}$ moyenne de tous les $y_{ij}$.

$$\bar{Y} = \frac{1}{8}(4+5+2+2+3+4+1+3) = 3$$

– moyennes marginales de $X_1$ :

- $\bar{Y}_{1.} = \frac{1}{n_{1.}}\sum_j y_{1j}$ moyenne de tous les $y_{1j}$ (tous les scores du produit $P1$)

$$\bar{Y} = \frac{1}{4}(4+5+3+4) = 4$$

- $\bar{Y}_{2.} = \frac{1}{n_{2.}}\sum_j y_{2j}$ moyenne de tous les $y_{2j}$ (tous les scores de produit $P2$)

$$\bar{Y}_{2.} = \frac{1}{4}(2+2+1+3) = 2$$

# Calcul des estimateurs des paramètres du modèle

– moyennes marginales de $X_2$ :
  - $\bar{Y}_{.1} = \frac{1}{n_{.1}} \sum_i y_{i1}$ moyenne de tous les $y_{i1}$ (tous les scores du Juge $J1$)

$$\bar{Y}_{.1} = \frac{1}{4}(4 + 5 + 2 + 2) = 3.25$$

  - $\bar{Y}_{.2} = \frac{1}{n_{.2}} \sum_i y_{i2}$ moyenne de tous les $y_{i2}$ (tous les scores du Juge $J2$)

$$\bar{Y}_{.2} = \frac{1}{4}(3 + 4 + 1 + 3) = 2.75$$

Toutes ces formules peuvent se visualiser à l'aide d'un tableau :

$X_1$ : facteur 1 = Produit

$X_2$ : facteur 2 = Juge

| score $Y$ | J1 | J2 | moyenne (en lignes) |
|---|---|---|---|
| Prod. 1 | 4 | 3 | $\bar{Y}_{1.} = 4$ |
| | 5 | 4 | |
| Prod. 2 | 2 | 1 | $\bar{Y}_{2.} = 2$ |
| | 2 | 3 | |
| moyenne (en colonnes) | $\bar{Y}_{.1} = 3.25$ | $\bar{Y}_{.2} = 2.75$ | $\bar{Y} = 3$ |

# Calcul des estimateurs des paramètres du modèle

On estime :
$\mu$ par $\hat{\mu} = \bar{Y}$
$\alpha_1$ par $\hat{\alpha}_1 = \bar{Y}_{1.} - \bar{Y}$
$\beta_1$ par $\hat{\beta}_1 = \bar{Y}_{.1} - \bar{Y}$
Dans notre exemple, cela donne :

$$\begin{cases} \hat{\mu} = 3 \\ \hat{\alpha}_1 = 4 - 3 = 1 \text{ (et } \hat{\alpha}_2 = -1 \text{ )} \\ \hat{\beta}_1 = 4 - 3.75 = 0.25 \text{ (et } \hat{\beta}_2 = -0.25 \text{ )} \end{cases}$$

Enfin pour estimer l'interaction $\gamma_{11}$,
on écrit :

$$\begin{aligned} \bar{Y}_{11} &= \bar{Y} + \hat{\alpha}_1 + \hat{\beta}_1 + \hat{\gamma}_{11} \\ 4.5 &= 3 + 1 + 0.25 + \hat{\gamma}_{11} \end{aligned}$$

D'où, $\hat{\gamma}_{11} = 0.25$

# Tableau récapitulatif des formules des paramètres estimés

Attention ! ! ! Cas particulier du plan équilibré complet équilibré : on a fait toutes les "expériences" possibles (tous les juges ont donné une note à tous les produits) et le nombre de répétitions est identique = 2 pour chaque expérience.

- $\hat{\mu} = \bar{Y}$
- $\hat{\alpha}_i = \bar{Y}_{i.} - \bar{Y}$
- $\hat{\beta}_j = \bar{Y}_{.j} - \bar{Y}$
- $\hat{\gamma}_{ij} = \bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}$

# Avec R

On peut faire tous les calculs avec R et tester la significativité des facteurs :

```
model<- lm(score~ produit*juge, data= D)

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.0000     0.3062   9.798 0.000608 ***
produit1         1.0000     0.3062   3.266 0.030906 *
juge1            0.2500     0.3062   0.816 0.460051
produit1:juge1   0.2500     0.3062   0.816 0.460051
```

On retrouve bien les estimations des paramètres calculés "à la main" :

$$\underbrace{\hat{\mu} = 3}_{\text{"}\textit{intercept}\text{"}} \; ; \hat{\alpha}_1 = 1 \, ; \hat{\beta}_1 = 0.25 \, ; \hat{\gamma}_{11} = 0.25 \, .$$

# Décomposition de la variabilité

Cas complet et équirépété : les sommes des carrés s'additionnent :

$$SCT = SC_1 + SC_2 + SC_{12} + SCR$$

$$\begin{aligned} SCT = \sum_{i,j,k} \left(y_{ijk} - \bar{Y}\right)^2 \quad = \quad & \underbrace{\sum_{i,j,k} \left(\bar{Y}_{i.} - \bar{Y}\right)^2}_{SC_1 : facteur1} \\ & + \underbrace{\sum_{i,j,k} \left(\bar{Y}_{.j} - \bar{Y}\right)^2}_{SC_2 : facteur2} \\ & + \underbrace{\sum_{i,j,k} \left(\bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}\right)^2}_{SC_{12} : interaction} \\ & + \underbrace{\sum_{i,j,k} \left(Y_{ijk} - \bar{Y}_{ij}\right)^2}_{SCR : résidus} \end{aligned}$$

# Test global d'un effet : test de Fisher

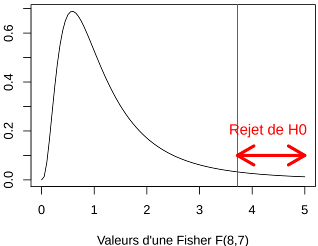
Exemple Test du facteur $X_1$ :

- $\mathcal{H}_0$ : tous les $\alpha_i = 0$ *contre* $\mathcal{H}_1$ : au moins un $\alpha_i \neq 0$
- statistique de test :

$$F = \frac{SC_1/ddl_1}{SCR/ddl_R}$$

avec $ddl_1$ le degré de liberté du facteur $X_1$ soit $ddl_1 = I - 1$ et $ddl_R$ le degré de liberté des résidus.

- Si $\mathcal{H}_0$ est vraie alors la loi de $F$ est une Fisher $\mathcal{F}(ddl_1, ddl_R)$
- Si $F > f_{1-\alpha}$ ou si la p-value $< \alpha$ on rejette $\mathcal{H}_0$ au risque $\alpha$.



# Test "ANOVA" de Fisher : tableau d'ANOVA avec formules

**Table ANOVA à deux facteurs avec interaction**

| Source | Degrés de liberté | Somme des Carrés | Carrés Moyens | $F$ - statistique de test | $p$-value |
|---|---|---|---|---|---|
| facteur 1 | $I-1$ | $SC_1$ | $\frac{SC_1}{I-1}$ | $F_{1,obs}$ | p-value |
| facteur 2 | $J-1$ | $SC_2$ | $\frac{SC_2}{J-1}$ | $F_{2,obs}$ | p-value |
| inter 1-2 | $(I-1)(J-1)$ | $SC_{12}$ | $\frac{SC_{12}}{(I-1)(J-1)}$ | $F_{12,obs}$ | p-value |
| Résidus | $ddl_R$ | $SCR$ | $\frac{SCR}{n-K}$ | | |

avec $ddl_R = n - K - 1$ et $K$ est la somme de tous les autres ddl.

$\hookrightarrow$ voir exercice d'application du TD.

# Test "ANOVA" de Fisher : exemple

**Table ANOVA à deux facteurs avec interaction**

| Source | Degrés de liberté | Somme des Carrés | Carrés Moyens | statistique de test $F$ | $p$-value |
|---|---|---|---|---|---|
| produit | 1 | $SC_1 = 8$ | $\frac{SC_1}{I-1} = 8$ | $F_{1,obs} = 10.6667$ | 0.03091 * |
| juge | 1 | $SC_2 = 0.5$ | $\frac{SC_2}{J-1} = 0.5$ | $F_{2,obs} = 0.6667$ | 0.46005 |
| interaction | 1 | $SC_{12} = 0.5$ | $\frac{SC_{12}}{(I-1)(J-1)} = 0.5$ | $F_{12,obs} = 0.6667$ | 0.46005 |
| Résidus | 4 | $SCR = 3$ | $\frac{SCR}{n-K} = 0.75$ | | |

avec $ddl_R = 8 - 3 - 1 = 4$.

$\hookrightarrow$ voir exercice d'application du TD.

# Test "ANOVA" de Fisher : avec R

On peut obtenir la Table ANOVA avec tous les tests des effets avec R en tapant les deux lignes de commandes suivantes :

```
model<- lm(score~ produit*juge, data= D)
anova(model)
```

```
Analysis of Variance Table

Response: score
             Df Sum Sq Mean Sq F value  Pr(>F)
produit       1    8.0    8.00 10.6667 0.03091 *
juge          1    0.5    0.50  0.6667 0.46005
produit:juge  1    0.5    0.50  0.6667 0.46005
Residuals     4    3.0    0.75
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
```

Tout ceci à condition que les hypothèses du modèle soient valides :
– Vérifier la normalité de la variable d'intérêt.
– Vérifier l'égalité des variances dans les groupes définis par les facteurs ...
– Vérifier la non-corrélation des erreurs ...

$\hookrightarrow$ voir comment faire en pratique en TP...