

Analyse de la Variance

- i) Principe de l'analyse de la variance
- ii) Conditions d'application de l'ANOVA
- iii) ANOVA à un seul facteur
 - Dispositif complètement aléatoire
 - Dispositif en blocs aléatoires complets
 - Dispositif suivant un carré latin

L'analyse de la variance

ANOVA

Introduction : Lorsqu'on fait une expérience, on fait souvent affecter des facteurs qu'on appelle souvent **traitement** ou **objets** (par exemple : dose d'azote, dose d'irrigation....) à des unités expérimentales (par exemple des parcelles, des animaux, des malades humains...etc).

On veut donc savoir l'effet de ces facteurs sur 1 ou plusieurs variables de ces unités expérimentales, ces variables peuvent être par exemple dans le cas de la production végétale les composantes du rendement (poids de milles graines, n^{bre} de pieds par m²....) et dans le cas de la production animale les performances zootechniques de la production animale (poids, production laitière) ou des taux de certains paramètres tels que la glycémie en biologie...etc.

Exemple: Dans chacune des 3 régions agricoles, on cultive 4 variétés de tomate en utilisant 2 sortes d'engrais.

L'ordre d'évaluation de la production des 3 régions après 1 année, on pose les 4 questions suivantes :

- 1- y a t il une différence significative entre les productions.
- 2- Si oui à quel facteur est elle due.
- 3- Peut on estimer l'influence de chacun des facteurs.
- 4- Y a-t-il une influence qui est due à la combinaison (interaction) des facteurs.

Sur la base d'un plan d'expérience (dispositif expérimental), on procède à l'analyse de la variance (ANOVA) pour répondre à ces questions.

Principe de l'analyse de la variance :

Le principe de l'ANOVA découle de l'hypothèse H_0 qui consiste à dire que toutes les moyennes des variables sont égales et que qu'il n'y a aucune différence significative entre elles, cela veut dire qu'on suppose que les facteurs étudiés n'ont pas d'effet sur les variables.

$$H_0 : m_1 = m_2 = m_3 = \dots\dots\dots=m_n.$$

Cela veut dans notre exemple précédent que les rendements sont égaux

Puis en fonction de chaque dispositif expérimental, on procède à l'établissement de l'ANOVA par le calcul de:

- La moyenne générale
- La somme des carrés des écarts des facteurs SCE (SCE_a, SCE_b, \dots),
- La moyenne des blocs s'il y en a.,
- La somme des carrés des écarts de l'interaction SCE_{ab} ,
- La somme des carrés des écarts résiduels SCE_r

- Les carrés moyens factoriels (CM_a, CM_b, \dots)
- Les carrés moyens inter factoriels CM_{ab}
- Les carrés moyens résiduels CM_r
- $F_{\text{observé}} \quad F_a = \frac{CM_a}{CM_r}; F_b = \frac{CM_b}{CM_r}; F_{ab} = \frac{CM_{ab}}{CM_r} \dots$
- Puis on résume tout dans un tableau appelé **Tableau de l'analyse de la variance**.

| Source de variabilité | ddl | SCE | CM | F_{obs} | P |
|------------------------------|-----|----------------|--------------|------------------|---|
| Variabilité factorielle | | SCE_a, SCE_b | CM_a, CM_b | F_a, F_b | |
| Variabilité de l'interaction | | SCE_{ab} | CM_{ab} | F_{ab} | |
| Variabilité du bloc | | SCE_i | CM_i | F_i | |
| Variabilité résiduelle | | SCE_r | CM_r | | |
| Variabilité totale | | SCE_t | CM_t | | |

Finalemment on doit comparer le $F_{\text{observé}}$ au $F_{\text{théorique}}$ qui doit être lu á partir de la table de Fischer-Snedecore.

Si $F_{\text{observé}} \geq F_{\text{théorique}} \Rightarrow$ on rejette H_0 indiquant une différence significative entre les moyennes des variables. On déclare que l'effet des traitements ou facteurs est significative au seuil de 5 % ou très hautement significative au seuil de 1%. Dans ce cas on doit procéder á la comparaison multiples des moyennes telle que la méthodes de la PPDS ou PPAS pour détecter á quel niveau la différence est plus significative et par conséquent le traitement le plus significatif.

Si $F_{\text{observé}} < F_{\text{théorique}} \Rightarrow H_0$ sera acceptée. Cela veut dire que les facteurs n'ont pas d'effet et dans ce cas on doit s'assurer de la **précision** de l'essai en étudiant l'écart type résiduel global pour accepter un tel résultat.

Conditions d'application de l'analyse de la variance :

Théoriquement, l'ANOVA ne peut s'applique que sur des données :

- qui sont normales c à d obéissent à une loi normale, la normalité d'une distribution est appréciée par un ajustement á la courbe en cloche de Gauss ou par le calcul du **coefficient d'asymétrie** qui doit être égal á zéro ($\beta_1 = 0$) et le **coefficient d'asymétrie** qui doit être égal á trois ($\beta_2 = 3$),

- qui sont indépendantes et qui n'ont aucun lien de dépendance ou corrélation, cela veut dire que l'erreur commise sur une unité expérimentale ne doit pas être liée à l'erreur faite sur une autre unité expérimentale voisine,
- qui ont la même variance (même dispersion) dans tous les traitements et dans tous les blocs, cela veut dire que les erreurs doivent être de même ordre de grandeur ou presque quelque soit le bloc ou le traitement.

L'égalité des variances qui est donnée sous forme d'une hypothèse H_0 est testée par la méthode de Bartlett qui consiste à calculer un paramètre qui suit une loi de χ^2 à ddl égal au n^{bre} de traitement ou au n^{bre} de bloc moins 1 ($K - 1$) avec une probabilité de calculer p. (bien sûr à un risque d'erreur α généralement égal à 5% ou 1%)

Si $p > \alpha \Rightarrow$ on conserve l'hypothèse H_0 (égalité des variances).

Si $p < \alpha \Rightarrow$ on rejette H_0 (les variances sont inégales).

Analyse de variance à un seul facteur :

1 Dispositif complètement aléatoire (randomisation totale) :

Si par exemple on cherche à savoir si l'alimentation a une influence sur la production laitière, alors on fait un plan d'expérience: On tire les vaches au hasard, on leur fait administrer un aliment qui est aussi choisi au hasard, ce type de plan est appelé **dispositif complètement aléatoire (ou complètement randomisé)**.

Supposons qu'il existe K aliments (A_i), on aura au minimum K populations (P_i) dans lesquelles on extrait K échantillons (E_i). Soit :

- m_i la moyenne de la population p_i
- m_1 la moyenne de la population p_1
- m_2 la moyenne de la population p_2
- .
- .
- .
- m_k la moyenne de la population p_k

On veut tester l'hypothèse $H_0 : m_1 = m_2 \dots = m_i = \dots = m_k$

A cet effet on va estimer les moyennes des populations par les moyennes des échantillons :

\bar{y}_i estime m_i

\bar{y}_1 estime m_1

\bar{y}_2 estime m_2

.

.

\bar{y}_K estime m_K

1-1. Présentation des données :

| | | | | | | | |
|----------------|-----------------|-----------------|--|-----------------|-------|-----------------|----------------|
| | A ₁ | A ₂ | | A _j | | A _k | |
| E ₁ | y ₁₁ | y ₁₂ | | y _{1j} | | y _{1K} | |
| E ₂ | Y ₂₁ | Y ₂₂ | | Y _{2j} | | Y _{2K} | |
| | | | | | | | |
| E _i | Y _{i1} | Y _{i2} | | Y _{ij} | | y _{iK} | |
| ⋮ | | | | | | | |
| E _n | y _{n1} | y _{n2} | | y _{nj} | | y _{nK} | |
| Moyenne | \bar{y}_1 | \bar{y}_2 | | \bar{y}_j | | \bar{y}_K | \bar{y} |
| Variance | \bar{S}_1^2 | \bar{S}_2^2 | | \bar{S}_j^2 | | \bar{S}_K^2 | S ² |

$$\bar{y}_1 = \frac{\sum y_{j1}}{n_{j1}}$$
 Production laitière moyenne des vaches qui ont consommé l'aliment A₁.

$$\bar{y} = \frac{\sum_{i=1}^n \sum_{j=1}^K y_{ij}}{n_i}$$
 Moyenne générale de la production laitière des vaches.

$$S_j^2 = \frac{\sum (y_j - \bar{y}_c)^2}{n_j - 1}$$
 Variance de la production laitière des vaches qui ont consommé l'aliment A_j.

$$S^2 = \frac{\sum_{i=1}^n \sum_{j=1}^K (y_{ij} - \bar{y})^2}{n - 1}$$
 Variance totale

1-2. Estimation du modèle :

Il s'agit d'expliquer la variance totale par les variations des facteurs qui entrent en jeu dans un dispositif donné. Le modèle de l'analyse de la variance á un facteur pour un dispositif complètement aléatoire se résume comme suit:

Variation totale = Variation factorielle + Variation résiduelle ou

$$SCE_t = SCE_a + SCE_r$$

Mathématiquement, on peut formulé ceci de la manière suivante:

$$\begin{aligned} y_{ij} - \bar{y} &= (\bar{y}_j - \bar{y}) + (y_{ij} - \bar{y}_j) \\ (y_{ij} - \bar{y})^2 &= (y_j - \bar{y})^2 + (y_{ij} - \bar{y}_j)^2 \\ \sum \sum (y_{ij} - \bar{y})^2 &= \sum \sum (y_j - \bar{y})^2 + \sum \sum (y_{ij} - \bar{y}_j)^2 \\ \sum \sum (y_{ij} - \bar{y})^2 &= \sum n_j (\bar{y}_j - \bar{y})^2 \oplus \sum \sum (y_{ij} - \bar{y}_j)^2 \\ SCE_t &= SCE_a + SCE_r \end{aligned}$$

Cette relation montre que la somme des carrés des écarts par rapport à la moyenne générale également appelé SCE_t peut être divisée en 2 composantes **additives** :

- Une SCE factorielle (SCE entre échantillon ou entre traitement) SCE_a .
- Et une SCE résiduelle (ou variation à l'intérieur des échantillons) SCE_r .

En divisant les SCE sur leurs ddl respectifs, on obtient ce qu'on appelle les carrés moyens qui serviront de base pour rejeter ou accepter H_0 .

$$CM_t = \frac{SCE_b}{ddl} \quad \text{carré moyen total ddl} = n - 1$$

$$CM_a = \frac{SCE_a}{K - 1} \quad \text{carré moyen factoriel}$$

$$CM_r = \frac{SCE_r}{n - K} \quad \text{carré moyen résiduel}$$

A partir de cela, on calcule $F_{\text{observé}}$:

$$F_{\text{obs}} = \frac{CM_a}{CM_r} \quad \text{une valeur observée d'une variable F de Fischer - Snedecor.}$$

$F_{\text{théorique}}$ est lu à partir de la table de Fischer - snedecor à $V_1 = K - 1$ et $V_2 = n - K$ de degré de liberté pour un seuil α donnée.

On rejette l'hypothèse H_0 lorsque $F_{\text{obs}} \geq F_{\text{théorique}}$.

1-3. Tableau de l'ANOVA :

| Source de variation | ddl | SCE | CM | F | p |
|---|-------|---------|--------|---------------------------|---|
| Variation entre traitement ou factorielle | K - 1 | SCE_a | CM_a | $F_a = \frac{CM_a}{CM_r}$ | |
| Variation résiduelle | n - K | SCE_r | CM_r | | |
| Variation totale | n - 1 | SCE_t | | | |

Si $F_{\text{obs}} < F_{\alpha, K-1, n-K}$ (ou $p > \alpha$) on accepte $H_0 \Rightarrow$ Il n'y a aucun effet des traitement sur les variables dans ce cas il faut voir la puissance de l'essai.

Si $F_{obs} \geq F_{\alpha, K-1, n-K}$ ($p < \alpha$) on rejette $H_0 \Rightarrow$ l'effet des traitements est significatif.

On passe à la CMM (comparaison multiple des moyennes).

Exemple: On désire comparer 3 types d'aliments en ce qui concerne leur effet sur la production laitière, à cet effet on prend 15 vaches, on sert le premier aliment A_1 aux 5 premières vaches choisi au hasard, l'aliment A_2 sera administré aux 5 autres vaches, l'aliment A_3 sera administré aux 5 dernières vaches et ceci au hasard.

| A_1 | A_2 | A_3 |
|-------|-------|-------|
| 38 | 42 | 30 |
| 40 | 45 | 32 |
| 41 | 43 | 41 |
| 35 | 44 | 34 |
| 36 | 39 | 33 |

Tester l'hypothèse que les aliments n'ont aucun effet sur la production laitière des vaches.

Solution: $\alpha = 5\%$ on pose H_0 : « les 3 aliments n'ont aucun effets sur la production laitière »

1) $\bar{y}_1 = 38,00$; $\bar{y}_2 = 42,60$; $\bar{y}_3 = 34,00$; $\bar{y} = 38,20$

2) $SCE_a = 5 [(38,00 - 38,20)^2 + (42,60 - 38,20)^2 + (34,00 - 38,20)^2] = 185,2$

$SCE_t = (38 - 38,20)^2 + (40 - 38,20)^2 + (41 - 38,20)^2 + \dots + (33 - 38,20)^2 = 302,4$

$SCE_r = SCE_t - SCE_a = 302,4 - 185,2 = 117,2$

3) $CM_a = \frac{SCE_a}{K-1} = \frac{185,2}{3-1} = 92,60$

$CM_r = \frac{SCE_r}{n-K} = \frac{117,2}{15-3} = 9,77$

4) $F_{obs} = \frac{CM_a}{CM_r} = \frac{92,60}{9,7} = 9,48$

| Source de variation | ddl | SCE | CM | F | P | ET | CV |
|-----------------------|-----|-------|-------|------|--------|------|-------|
| Variation factorielle | 2 | 185.2 | 92.60 | 9.48 | 0.0035 | | |
| Variation résiduelle | 12 | 117.2 | 9.77 | | | 3.13 | 8.2 % |
| Variation totale | 14 | 302.4 | | | | | |

$F_{\alpha} = 3.89$ (lu de la table de Fischer-Snédecure à $\alpha = 0,05$ et ddl: 2 et 12)

$F_{obs} > F_{\alpha} \longrightarrow H_0$ est rejetée : les traitements sont significatifs.

On remarque que F_{obs} est supérieur à $F_{théorique}$ (la probabilité $P = 0.0035 < 0,05$) ce qui indique que l'alimentation a bien un effet sur la production laitière, dans ce cas, on dit

que le traitement est significatif , il est même hautement significatif puisque P est aussi inférieur à 1% et le F_{obs} est largement supérieur à $F_{théorique}$ mais la question qui reste posée est: quel aliment parmi $A_1, A_2,$ et $A_3,$ qui donne la meilleur production laitières ? On doit passer à la comparaison multiple des moyennes CMM pour répondre à une telle question.

- Il faut voire aussi l'écart type résiduel qui est l'erreur globale de l'essai.

D'habitude c'est le 1^{er} paramètre à voir car il indique la précision de l'expérience, apparemment notre essai est assez précis pour accepter les résultats de l'ANOVA puisque $ET = 3.13$ qui est une valeur assez faible et ne représente que 8,2 % (coefficient de variation) de la moyenne.

2- Dispositif en blocs aléatoires complets :

Parfois il est souhaitable de classer les unités expérimentâtes suivant un certains N_{bre} de critères avant de faire l'expérience.

Si on garde le même exemple, On peut classer par exemple les vaches suivant l'âge, on aura déterminé les blocs une fois les unités classées.

$B_1 : \{ \text{âge entre 3 et 5} \}$

$B_2 : \{ \text{âge entre 5 et 9} \}$

$B_3 : \{ \text{âge plus de 9} \}.$

Supposons qu'on a K aliments à tester et l'hypothèse de base est toujours $H_0,$

$H_0: m_1 = m_2 = \dots = m_K,$ on suppose que les aliments n'ont aucun effet sur la production laitière.

Il faut que la taille de chaque **bloc** soit au moins égale au nombre de traitements, c à d qu'il soit **complet**.

Une fois les **blocs** constitués on commence l'expérience: On prend le bloc 1 et on affecte les traitements aux unités expérimentâtes d'une manière **aléatoire**. On fait la même chose aux 2 autres blocs. C'est pour cette raison qu'on parle du **dispositif en blocs aléatoires complets**.

Si on a constitué n_0 blocs et on a K traitements alors on aura $n_0 \times K$ unités expérimentales (vaches ou parcelles... par exemples).

2-1 Présentation des données :

| | T_1 | T_2 | | T_j | | T_K | Moyenne par Bloc |
|-------------------|----------|----------|-------|----------|-------|----------|------------------|
| BLOC ₁ | y_{11} | y_{12} | | y_{1j} | | y_{1k} | $\bar{y}_{.1}$ |
| BLOC ₂ | y_{21} | y_{22} | . | y_{2j} | . | y_{2k} | $\bar{y}_{.2}$ |
| | | | . | | . | | |

| | | | | | | | |
|------------------------|--------------------|--------------------|---|--------------------|---|--------------------|----------------------------|
| . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . |
| BLOC i | y_{i1} | y_{i2} | . | y_{noj} | . | y_{nok} | $\bar{y}_{\cdot i}$ |
| . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . |
| BLOC n_0 | y_{no1} | y_{no2} | . | y_{noj} | . | y_{nok} | $\bar{y}_{\cdot n_0}$ |
| . | . | . | . | . | . | . | . |
| Moyenne par Traitement | $\bar{y}_{1\cdot}$ | $\bar{y}_{2\cdot}$ | . | $\bar{y}_{j\cdot}$ | . | $\bar{y}_{k\cdot}$ | Moyenne générale \bar{y} |

n_0 : n^{bre} de bloc

K : n^{bre} de traitement

$n = n_0 * K$

$$\bar{y} = \frac{1}{n} \sum \sum y_{ij}$$

$$\bar{y}_{\cdot i} = \frac{1}{K} \sum y_{i\cdot} = \text{moyenne par bloc } i.$$

$$\bar{y}_{j\cdot} = \frac{1}{n_0} \sum y_{\cdot j} = \text{moyenne par traitement } j.$$

2-2 Estimation du modèle :

En plus du modèle précédent. On doit ajouter **l'effet bloc** qui correspond à la variabilité due au bloc (SCE_{Gi}) :

$$SCE_t = SCE_a + SCE_{Gi} + SCE_r$$

$$\sum \sum (y_{ij} - \bar{y})^2 = \sum n_0 (\bar{y}_j - \bar{y})^2 + \sum K (\bar{y}_i - \bar{y})^2 + \sum \sum (y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y})^2$$

En bloc aléatoire complet la variation totale peut se subdiviser en SCE_a , SCE_{Gi} et SCE_r . A partir de cela on peut déduire les carrés moyens suivants :

$$CM_t = \frac{SCE_t}{n-1} \quad : \text{ carré moyen total}$$

$$CM_a = \frac{SCE_a}{K-1} \quad : \text{ carré moyen factoriel}$$

$$CM_G = \frac{SCE_{G_i}}{n_0-1} \quad : \text{ carré moyen bloc}$$

$$CM_r = \frac{SCE_r}{(n_0-1)(K-1)} \quad : \text{ carré moyen résiduel}$$

avec n_0 : bloc et K : traitement

et finalement On peut calculer

$$F_a = \frac{CM_a}{CM_r}, \quad F_G = \frac{CM_G}{CM_r}$$

2-3 tableau de l'ANOVA :

| Source de variabilité | ddl | SCE | CM | F | F α |
|-----------------------|---------------------------|------------------|-----------------|----------------|------------------------------------|
| Variable factorielle | K- 1 | SCE _a | CM _a | F _a | F $\alpha_{(k-1), (k-1)(n_0-1)}$ |
| Variable du bloc | n ₀ -1 | SCE _G | CM _g | F _G | F $\alpha_{(n_0-1), (K-1)(n_0-1)}$ |
| Variable résiduelle | (K-1) (n ₀ -1) | SCE _r | CM _r | | |
| Variable totale | n - 1 | SCE _t | CM _t | | |

Exemple:

Les résultats suivants relatifs à la teneur en cendre (exprimé en %) ont été obtenus par un essai en bloc aléatoire complet destiné à comparer l'influence de 7 fumures différentes sur une même variété du tabac. En fonction de ces résultats doit-on conclure à l'existence de différences significatives dues aux fumures ?

| Fumures | BLOC 1 | BLOC 2 | BLOC 3 | Moyennes |
|----------|--------|--------|--------|---------------------------|
| 01 | 21.3 | 22,4 | 20.4 | 21,36 |
| 02 | 18.3 | 19.9 | 20.0 | 19,4 |
| 03 | 19.1 | 22.6 | 19.8 | 20,5 |
| 04 | 19.9 | 24.6 | 19.3 | 21,26 |
| 05 | 19.3 | 23.0 | 19.9 | 20,73 |
| 06 | 19.8 | 22.2 | 19.1 | 20,36 |
| 07 | 19.0 | 22.2 | 20.2 | 20,46 |
| Moyennes | 19,52 | 22,41 | 19,81 | Moyenne générale 20,58 |

Solution:

• $H_0 : \bar{Y}_{1\bullet} = \bar{Y}_{2\bullet} = \dots = \bar{Y}_{7\bullet} = \bar{Y}_{\bullet}$ 7traitements

$H_0 : \bar{Y}_{\bullet 1} = \bar{Y}_{\bullet 2} = \bar{Y}_{\bullet 3}$ 3blocs

• $\bar{Y}_{1\bullet} = 21,36$; $\bar{Y}_{2\bullet} = 19,4$; $\bar{Y}_{3\bullet} = 20,5$; $\bar{Y}_{4\bullet} = 21,26$; $\bar{Y}_{5\bullet} = 20,73$; $\bar{Y}_{6\bullet} = 20,52$
 $\bar{Y}_{7\bullet} = 20,6$

$\bar{Y}_{\bullet 1} = 19,52$; $\bar{Y}_{\bullet 2} = 22,41$; $\bar{Y}_{\bullet 3} = 19,81$

$\bar{Y} = 20,58$

• $SCE_t = (21,3 - 20,58)^2 + \dots + (20,2 - 20,58)^2 = 53,03$

$SCE_a = 3[(21,36 - 20,58)^2 + \dots + (20,4 - 20,58)^2] = 7,6647$

$SCE_G = 7[(19,52 - 20,58)^2 + (22,41 - 20,58)^2 + (19,81 - 20,58)^2] = 35,4578$

$SCE_r = SCE_t - (SCE_a + SCE_G)$

$SCE_r = 53,038 - (7,66 + 35,45) = 9,9155$

• $CM_a = \frac{SCE_a}{3-1} = \frac{7,66}{2} = 3,83$

$CM_G = \frac{SCE_G}{3-1} = \frac{35,45}{2} = 17,7289$

$CM_r = \frac{SCE_r}{(7-1)(3-1)} = \frac{9,9155}{12} = 0,8262$

$F_a = \frac{CM_a}{CM_r} = 4,63$

$F_G = \frac{CM_G}{CM_r} = 21,458$

| Source de variable | ddl | SCE | CM | F | F α |
|----------------------|-----|---------|---------|--------|------------|
| Variable factorielle | 6 | 7.6647 | 1.277 | 1.546 | 3.00 |
| Variable du bloc | 2 | 35.4578 | 17.7289 | 21.456 | 3.89 |
| Variable résiduelle | 12 | 9.9155 | 0.8262 | | |
| Variable totale | 20 | 53.038 | | | |

Puisque $F_{obs} = 1.54 < F_{theo} = 3 \Rightarrow$ les fumures sont identiques et qu'il n'y a aucune différence entre elles. H_0 est donc acceptée.

On remarque aussi que l'effet bloc est significatif puisque $21.45 > 3.89$. Les blocs sont donc hétérogène et n'ont pas été bien contrôlés.

3- dispositif suivant un carré latin :

Soit une population bovine **P** et soit **K** aliments, on veut tester l'efficacité des aliments, mais on craint que l'effet âge et l'effet race viennent nous fausser les conclusions

quand à l'efficacité des aliments. On veut donc isoler l'effet traitement (alimentation) des autres effets (âge et race).

D'une manière générale pour des problèmes de ce genre on classe les unités expérimentales (bovins ici) suivant 2 critères au lieu d'un critère.

Supposons qu'il y ait 3 niveaux pour le 1^{er} critère (3 âges) et 3 niveaux pour le 2^{ème} critère (3 races) et 3 traitements. Pour mener à bien l'expérience il faut utiliser $3 \times 3 \times 3 = 3^3 = 27$ unités expérimentales (3 vaches par exemple)

Malheureusement une opération de ce genre est très coûteuse en argent et en temps, de plus certaines combinaisons (âge et race) sont dans la pratique « inobservable » ! Devant de telles difficultés on a recours à l'**expérience suivant un carré latin**, l'idée est la suivante: Chaque traitement sera expérimenté une fois et une seule fois pour chaque niveau du 1^{er} critère (âge), et une fois et une seule fois pour chaque niveau du 2^{ème} critère (race).

Si les traitements sont représentés par A, B, C on peut résumer le carré latin par le tableau suivant:

| Âge \ Race | Race | | |
|------------|------|---|---|
| | 1 | 2 | 3 |
| 1 | A | B | C |
| 2 | C | A | B |
| 3 | B | C | A |

C'est un carré latin à 3 dimension (3 lignes et 3 colonnes).

On peut donc énoncer la définition suivante: « On appelle un carré latin, un tableau dont chaque nombre figure une fois et une seule fois par ligne et une fois et une seule fois par colonne ».

En expérimentation au lieu de parler de la ligne on parle du **bloc horizontal** (BH) et au lieu de parler de la colonne on parlera du **bloc vertical** (BV), l'intersection entre la ligne i et la colonne j sera appelée **cellule (i,j)**.

3-1 présentation des données :

Avec **r** blocs horizontaux et **r** blocs verticaux et **r** traitements, on représente les données d'un carré latin comme suit:

| BH \ BV | 1 | 2 | | j | | r |
|---------|-------------|---------------|---------|---------------|---------|---------------|
| 1 | $y_{11(1)}$ | $y_{12(2)}$ | \cdot | $y_{ij(j)}$ | | $y_{ir(r)}$ |
| \cdot | \cdot | \cdot | \cdot | \cdot | \cdot | \cdot |
| \cdot | \cdot | \cdot | \cdot | \cdot | \cdot | \cdot |
| \cdot | \cdot | \cdot | \cdot | \cdot | \cdot | \cdot |
| i | $y_{i1(1)}$ | $y_{i2(I+2)}$ | \cdot | $y_{ij(I+j)}$ | \cdot | $y_{ir(1)}$ |
| \cdot | \cdot | \cdot | \cdot | \cdot | \cdot | \cdot |
| \cdot | \cdot | \cdot | \cdot | \cdot | \cdot | \cdot |
| \cdot | \cdot | \cdot | \cdot | \cdot | \cdot | \cdot |
| r | $y_{r1(r)}$ | $y_{n2(r)}$ | \cdot | $y_{nj(r)}$ | \cdot | $y_{2r(r-1)}$ |
| | | | \cdot | | \cdot | |
| | | | \cdot | | \cdot | |
| | | | \cdot | | \cdot | |

$y_{11(1)}$: cellule (1,1) qui a utilisé le traitement 1 ou traitement 1 figurant dans le BH_1 et BV_1

On calcul de ce tableau les quantités suivantes :

- La moyenne des BH : $\bar{y}_{i..} = \frac{1}{r} \sum_{jK} y_{ijK}$
- La moyenne des B V : $\bar{y}_{.j.} = \frac{1}{r} \sum_{jK} y_{ijK}$
- La moyenne des traitements : $\bar{y}_{..K} = \frac{1}{r} \sum_{jK} y_{ijK}$
- La moyenne générale : $\bar{y} = \frac{1}{r^2} \sum y_{ijK}$

3-2 Estimation du modèle :

Le modèle du carré latin est un modèle additif comme ceux des dispositifs précédents. On peut dire que la variation totale est expliquée par la variation qui due au facteur étudié plus la variation due aux blocs horizontaux et variation due aux blocs verticaux, plus la variation résiduelle. Cela se traduit mathématiquement sous la forme suivante.

$$SCE_t = SCE_a + SCE_{BH} + SCE_{BV} + SCE_r$$

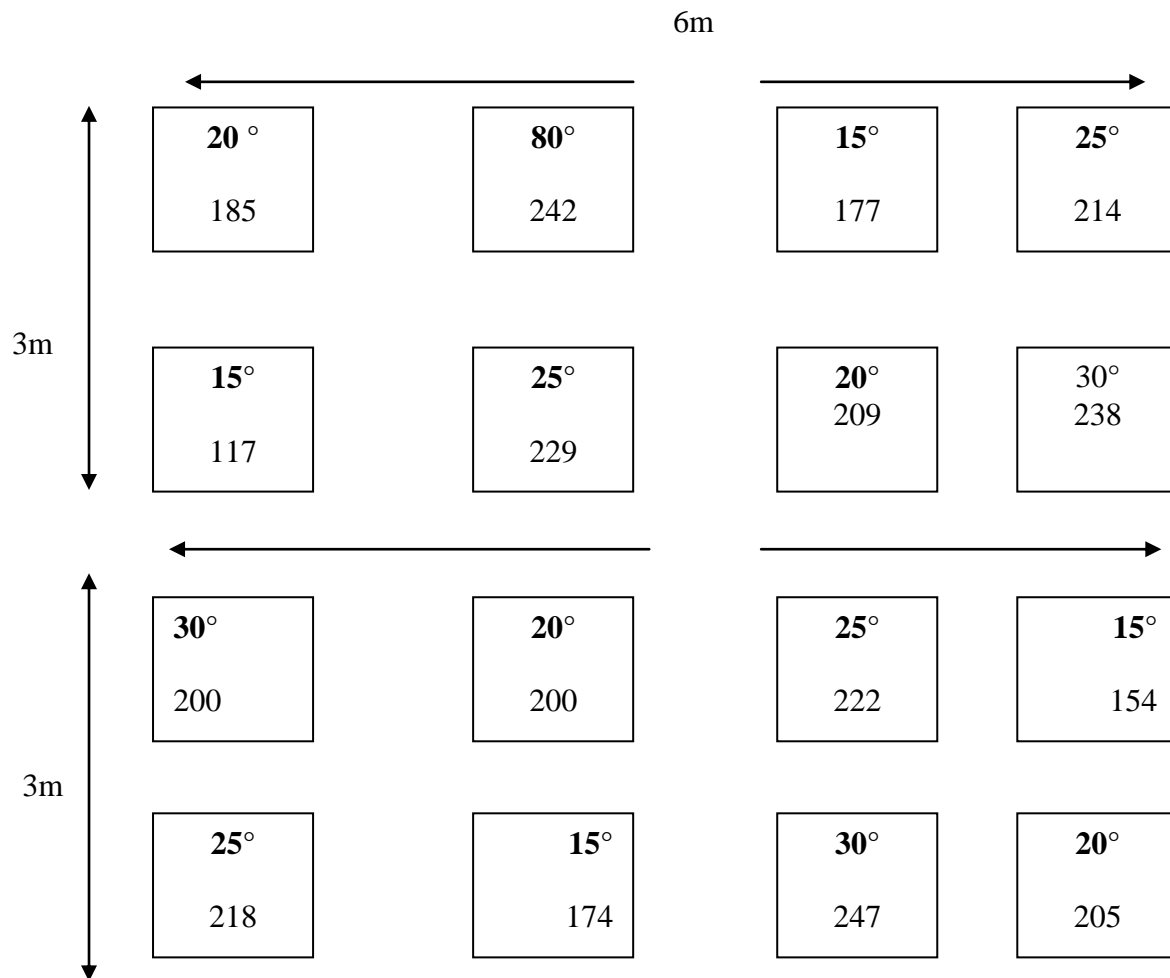
avec :

- $SCE_t = \sum_i^r \sum_K^r (y_{ijK} - \bar{y})^2$
- $SCE_a = \sum r(\bar{y}_{..K} - \bar{y})^2$
- $SCE_{BH} = \sum r(\bar{y}_{i..} - \bar{y})^2$
- $SCE_{BV} = \sum r(\bar{y}_{.j.} - \bar{y})^2$
- $SCE_r = SCE_t - [SCE_a + SCE_{BH} + SCE_{BV}]$
 $= \sum \sum (y_{ij} - \bar{y}_{i..} - \bar{y}_{..K} - \bar{y}_{.j.} + 2\bar{y})^2$
- $CM_a = \frac{SCE_a}{r-1}$
- $CM_{BH} = \frac{SCE_{BH}}{r-1}$
- $CM_{BV} = \frac{SCE_{BV}}{r-1}$
- $CM_t = \frac{SCE_t}{r^2-1}$
- $CM_r = \frac{SCE_r}{(r-1)(r-2)}$

3-3 Tableau de l'ANOVA :

| Source de variable | ddl | SCE | CM | F |
|----------------------------------|--------------------|-------------------|------------------|---------------------------------|
| Variable inter traitement | r - 1 | SCE _a | CM _a | $F_a = \frac{CM_a}{CM_r}$ |
| Variable inter blocs horizontaux | r - 1 | SCE _{BH} | CM _{BH} | $F_{BH} = \frac{CM_{BH}}{CM_r}$ |
| Variable inter blocs verticaux | r - 1 | SCE _{BV} | CM _{BV} | $F_{BV} = \frac{CM_{BV}}{CM_r}$ |
| Variable résiduelle | (r-1)(r-2) | SCE _r | CM _r | |
| Variable totale | r ² - 1 | SCE _t | CM _t | |

Exemple: Un essai de chauffage du sol réalisé en serre sur une variété décorative de **Ficus elastica R.** Les températures du sol prises en considération sont 15.20 .25 et 30° C à 12cm de profondeur. Toutes les autres condition de culture étant en principe uniformes. Les unités expérimentales sont des parcelles de 9 plantes cultivées en pots enfoncés enterre. On mesure l'accroissement moyen en hauteur de la plante en mm. Y-t-il une différence significative entres les températures ? le dispositif expérimental est indiqué comme suit:



Solution :

| BH \ BV | 1 | 2 | 3 | 4 | $\bar{y}_{i..}$ |
|-----------|-----|--------|--------|----------|-----------------|
| 1 | 185 | 242 | 177 | 214 | 204.5 |
| 2 | 117 | 229 | 209 | 238 | 198.25 |
| 3 | 200 | 200 | 222 | 154 | 194 |
| 4 | 218 | 174 | 247 | 205 | 211 |
| $y_{.j.}$ | 180 | 211.25 | 213.75 | 164.3125 | Y=201.9375 |

La moyenne des traitements :

$$\bar{y}_{15^{\circ}} = 155,5 \quad ; \quad \bar{y}_{20^{\circ}} = 199,75 \quad ; \quad \bar{y}_{25^{\circ}} = 220,75 \quad ; \quad \bar{y}_{30^{\circ}} = 231,75$$

$$H_{0\text{taritement}} : \bar{y}_{\bullet\bullet 1} = \bar{y}_{\bullet\bullet 2} = \bar{y}_{\bullet\bullet 3} = \bar{y}_{\bullet\bullet 4}$$

controle

$$BH : \bar{y}_{1\bullet\bullet} = \bar{y}_{2\bullet\bullet} = \bar{y}_{3\bullet\bullet} = \bar{y}_{4\bullet\bullet}$$

$$BV : \bar{y}_{\bullet 1\bullet} = \bar{y}_{\bullet 2\bullet} = \bar{y}_{\bullet 3\bullet} = \bar{y}_{\bullet 4\bullet}$$

$$\begin{aligned} \bullet \text{ SCE}_t &= (185 - 202)^2 + \dots + (205 - 202)^2 = 17743 \\ \text{SCE}_a &= 4[(155,5 - 202)^2 + \dots + (232 - 202)^2] = 13616 \\ \text{SCE}_{BH} &= 4[(204,5 - 202)^2 + \dots + (211 - 202)^2] = 661 \\ \text{SCE}_{BV} &= 4[(180 - 202)^2 + \dots + (202,75 - 202)^2] = 2833 \\ \text{SCE}_r &= \text{SCE}_t - [\text{SCE}_{BH} + \text{SCE}_{BV} + \text{SCE}_a] \\ &= 17743 - [13616 + 661 + 2833] \\ &= 633 \end{aligned}$$

$$\bullet \text{ CM}_t = \frac{\text{SCE}_t}{16 - 1} = \frac{17743}{15} = 1182,86$$

$$\text{CM}_a = \frac{\text{SCE}_a}{4 - 1} = 4539$$

$$\text{CM}_{BH} = \frac{\text{SCE}_{BH}}{4 - 1} = 220$$

$$\text{CM}_{BV} = \frac{\text{SCE}_{BV}}{4 - 1} = 944$$

$$\text{CM}_r = \frac{\text{SCE}_r}{(4 - 1)(4 - 2)} = 105,5$$

$$\bullet F_a = \frac{\text{CM}_a}{\text{CM}_r} = \frac{4539}{105,5} = 43$$

$$F_{BH} = \frac{\text{CM}_{BH}}{\text{CM}_r} = \frac{220}{105,5} = 2,08$$

$$F_{BV} = \frac{\text{CM}_{BV}}{\text{CM}_r} = \frac{944}{105,5} = 8,94$$

| Source de variable | ddl | SCE | CM | F OBS | F théo |
|---------------------|-----|--------|------|-------|--------|
| Temperature | 3 | 13616 | 4539 | 43 | 4.76 |
| B H | 3 | 661 | 220 | 2.09 | 4.76 |
| B V | 3 | 2833 | 944 | 8.94 | 4.76 |
| variable résiduelle | 6 | 633 | 105 | | |
| Totale | 15 | 177743 | | | |

Puisque $F_{\text{obs}} > F_{\text{théo}}$ ($43 > 4,76$) \Rightarrow H_0 qui supposait l'égalité des 4 températures est rejetée. Cela veut dire que la différence entre ces moyennes est **significative** ce qui indique que le facteur température a une influence très marquée sur l'accroissement de la plante étudiée.

Mais la question qui reste posée est: quelle température est la plus significative ?

Concernant les blocs horizontaux, on remarque que $2,09 < 4,76$ cela traduit le bon contrôle de ces blocs. Les BH semblent être homogènes, par contre les blocs verticaux apparaissent très hétérogènes puisque $8,94 > 4,76$, ce qui peut être expliqué par le fait que les BV ont été installés le long de deux serres différents.