Université Ibn Khaldou Tiaret

UIK

Faculté des Sciences de la Nature et de la Vie

STATISTIQUES ET EXPÉRIMENTATIONS AGRICOLES

COURS POLYCOPIÉS

DESTINÉS AUX CLASSES DE MASTER EN AGRICULTURE DE PRÉCISION

DANS LE CADRE D'UN PROJET ERASMUS+ CBHE

(*Capacity Building in Higher Education CBHE)*

‘' New Curricula in Precision Agriculture Using GIS Technologies and Sensing Data”

PRÉPARÉS PAR

OMAR YAMINA — MAÎTRE DE CONFÉRENCES A

MAATOUG M'HAMED — PROFESSEUR

ANNÉE 2021/2022

AVANT-PROPOS

Statistiques et expérimentations Agricoles : cours polycopiés destinés aux classes de master en Agriculture de Précision, mais il s'adresse également aux étudiants et aux chercheurs spécialistes en Sciences agronomiques et biologiques.

Une grande partie des références bibliographique et les séries des exemples sont basées sur les ouvrages de *Pierre Dagnelie* (1970, 1975, 1951 2012).

Nos remerciements adressés à toutes les personnes (Partenaires Européens et Algériens du projet Coupais) qui nous ont aidé dans la préparation de ce document.

TABLE DES MATIERES

Chapitre 01: Statistiques descriptives

1. Notions générales vocabulaires

1.1. Population

On appelle population l'ensemble d'éléments sur lequel on s'intéresse et porte notre étude.

1.2. Individu « unité statistique »

Les éléments de la population sont appelés les individus ou unités statistiques.

1.3. Caractère « variable statistique »

Sur les unités statistiques ou individus, on mesure un caractère ou une variable, la série de ces observations forme ce que l'on appelle **une variable statistique.**

La variable prend toujours une seule valeur sur chaque unité. Les variables sont désignées par une lettre (X, Y, Z).

Par exemple, les notes des étudiants à l'examen de Statistique, les mentions qu'ils ont obtenues à leur bac, leur sexe, les couleurs de leurs yeux.

1.4. Modalité

Les modalités d'une variable statistique sont les différentes valeurs que peut prendre celle-ci. Elles sont les différentes situations dans lesquelles les individus peuvent se trouver à l'égard du caractère considéré.

Exemple : Variable est situation familiale, Modalités sont *célibataire, marié, divorcé.*

Variable est « catégories socio-professionnelles » les Modalités sont « *Employés, ouvriers, retraités,...* ».

1.5. Types de variables

Nous distinguons deux catégories de variables : les variables qualitatifs et les variables quantitatifs.

1.5.1. Variable qualitative

La variable est dite qualitative quand les modalités ou les valeurs sont des catégories et sont désignées par des noms, exemple : le sexe, couleur des yeux….

On distingue deux types de variables qualitatives : les variables qualitatives **ordinales** et les variables qualitatives **nominales.**

– Variable qualitative nominale

La variable est dite qualitative nominale quand les modalités ne peuvent pas être ordonnées.

Exemple: Couleur des yeux ou variable sexe.

– Variable qualitative ordinale

La variable est dite qualitative ordinale quand les modalités peuvent être ordonnées dans un certain ordre naturel.

Exemple: Mention du bac.

1.5.2. Variable quantitative

Une variable est dite quantitative si toutes ses valeurs possibles sont numériques et est mesurée par un nombre.

On distingue 2 types de variables quantitatives: les variables quantitatives discrètes et les variables quantitatives continues.

– Variable quantitative discrète

Une variable est dite discrète, si l'ensemble des valeurs possibles est dénombrable.

Par exemple le nombre d'enfants par ménage ne peut être que 0, ou 1, ou 2, ou 3, ... ; il ne peut jamais prendre une valeure strictement comprise entre 0 et 1.

– Variable quantitative continue

Une variable est dite continue, si l'ensemble des valeurs possibles est continu. Les variables quantitatives continues peuvent prendre toute valeur dans un intervalle.

2. Paramètres de position

2.1. Mode

Le mode est la valeur qui a le plus grand effectif ; il est noté *Mo*.

Si on reprend la variable 'Etat civil', dans le tableau statistique suivant :

Tableau 01: Statistique de l'état civil

Xi	*ni*
C	9
M	7
V	2
D	2
n =	20

C: célibataire, M: marié V: veuf, D: divorcé

Le mode est *C* : célibataire avec 9.

– Le mode peut être calculé pour tous les types de variable, quantitative et qualitative.

– Le mode n'est pas nécessairement unique.

– Quand une variable continue est découpée en classes, on peut définir une classe modale (classe correspondant à l'effectif le plus élevé).

2.2. Médiane

La médiane (notée M_e) d'une variable quantitative est la valeur de cette variable qui permet de scinder la population étudiée en deux sous-populations de même effectif.

Plus précisément, il y a autant d'individus pour lesquels on a observé une valeur supérieure à M_e que d'individus pour lesquels on a observé une valeur inférieure à M_e.

2.3. Moyenne

La *moyenne* ne peut être définie que sur une variable *quantitative*.

On dispose d'une population de N individus et on observe $x_1, x_2, ..., x_N$ les valeurs d'une variable quantitative discrète X pour ces individus.

a) Moyenne arithmétique

Elle est notée par $\bar{x}$ et elle est définie de la manière suivante:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Appliquons cette définition au calcul de la moyenne de la série suivante :

N = {4, 0, 1, 1, 2, 2, 2, 3, 3, 4, 2, 3, 4, 5, 2, 1, 3, 3, 4, 5}

On a donc :

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$
$$= \frac{4+0+1+1+2+2+2+3+3+4+2+3+4+5+2+1+3+3+4+5}{20} = \frac{54}{20}$$
$$= 2{,}7$$

b) Moyenne arithmétique pondérée

La moyenne arithmétique est dite moyenne pondérée quand chaque valeur de la variable est multipliée (pondérée) par un coefficient, chaque valeur x_i de la variable intervient dans le calcul de la moyenne autant de fois qu'elle a été observée.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} n_i x_i$$

Exemple

Tableau 02: Soit la distribution suivante :

x_i	n_i
0	1
1	3
2	5
3	5
4	4
5	2

La moyenne arithmétique pondérée est

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} n_i x_i = \frac{(0 \times 1) + (1 \times 3) + (2 \times 5) + (3 \times 5) + (4 \times 4) + (5 \times 2)}{20}$$

$$= \frac{0 + 3 + 10 + 15 + 16 + 10}{20} = \frac{54}{20} = 2{,}7$$

c) La moyenne arithmétique pour un caractère quantitatif continu (classes)

Lorsque l'on veut calculer la moyenne d'une distribution par classes de valeurs, celle-ci s'obtient en prenant la formule de la moyenne pondérée et en remplaçant dans cette formule "x_i" par "c_i", où c_i représente le centre de la classe j, c'est-à-dire la moyenne arithmétique des extrémités de classe.

Exemple :

Tableau 03: Soit la distribution statistique suivante :

Classes	n_i
[0 - 2[	4
[2 - 4[	10
[4 - 6]	6

Pour calculer la moyenne, nous devons déterminer les centres de classe, puis faire la somme des "nj x cj" et diviser par n. Autrement dit, nous devons appliquer la formule:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} n_i c_i$$

La notation c_i indique le centre de classe, il est égal à la moyenne des extrémités de classe. On a donc le tableau de calcul suivant :

Tableau 04: Moyenne des extrémités de classe

Classes	n_i	**c_i (moyenne des extrémités de classe)**	**n_i x c_i**
[0 - 2[	4	1	4
[2 - 4[	10	3	30
[4 - 6]	6	5	30
			64

Et finalement :

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} n_i c_i = \frac{1}{20}\sum_{i=1}^{N} [(4 \times 1) + (10 \times 3) + (6 \times 5)] = \frac{4 + 30 + 30}{20} = \frac{64}{20} = 3{,}2$$

d) Moyenne quadratique

Elle est notée par m_2 et elle est définie de la manière suivante :

$$m_2 = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} = \sqrt{\sum_{i=1}^{k} f_i\, x_i^2}$$

Exemple:

Une étude statistique menée sur une population de ménages a montré que 30% de ces ménages ont 01 enfants, 40% 02 enfants, 15% 03 enfants, 10% 04 enfants, et 5% 05 enfants.

Ainsi, la moyenne quadratique de la variable « Nombre d'enfants par ménage », vaut :

$$m_2 = (0{,}3 \text{ x } 1^2 + 0{,}\ 4 \text{ x } 2^2 + 0{,}15 \text{ x } 3^2 + 0{,}1 \text{ x } 4^2 + 0{,}05 \text{ x } 5^2)^{1/2} = 2{,}47.$$

e) Moyenne harmonique

Elle est notée par m_{-1} et elle est définie de la manière suivante :

$$m_{-1} = \frac{N}{\sum_{i=1}^{N}\frac{1}{x_i}} = \frac{1}{\sum_{i=1}^{k}\frac{f_i}{x_i}}$$

La moyenne harmonique peut être utilisée chaque fois qu'il est possible d'attribuer un sens réel aux inverses des données (taux d'équipement, pouvoir d'achat, calcul d'indice,...).

f) Moyenne géométrique

La moyenne géométrique ne peut être définit que lorsque les observations x_1,. .. , x_N sont tous des nombres réels positifs. Si tel est le cas, la moyenne géométrique de ces observations est notée par M_g, et elle est définie par :

$$\mathrm{M}_g = \sqrt[N]{x_1 x_2 \ldots\, x_N} = \sqrt[N]{x_1^{n1} \ldots\, x_K^{nk}} = x_1^{f1} \ldots \ldots x_K^{fk}$$

Exemple:

Supposons que pendant une décennie, les salaires aient été multipliés par 2 et que pendant la décennie suivante ils aient été multipliés par 4 ; alors pour la période de l'ensemble de ces deux décennies le coefficient multiplicateur est 2 x 4 = 8.

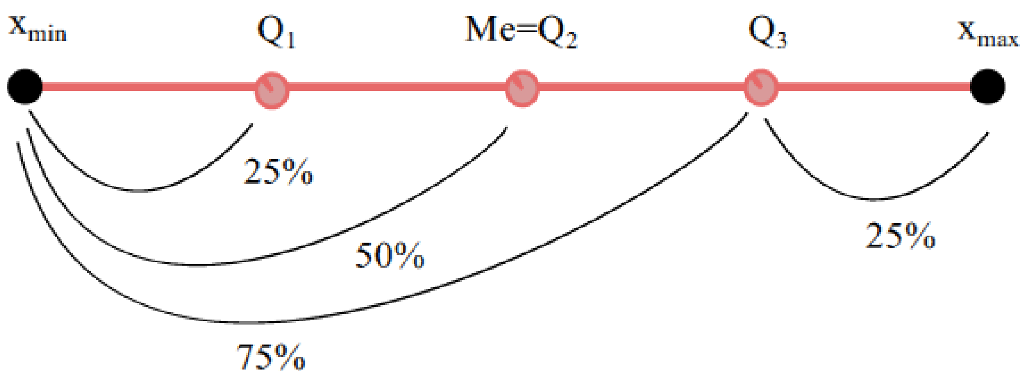
Le coefficient multiplicateur moyen par décennie pour cette période de vingt ans est, par définition, le coefficient μ qui ne change pas d'une décennie à l'autre, et qui permet une multiplication par 8 des salaires entre le début et la fin de la période.

0n a donc μ = 8 = 2 x 4, d'où μ = $\sqrt{2 \times 4}$ = 2, 83.

2.4. Quantile

Les quartilesde X sont ses trois quantiles $x_{0,25}$; $x_{0,5}$ et $x_{0,75}$.

$Q_1 = x_{0,25}$, s'appelle le premier quartile ; un quart des valeurs prises par X sont inférieures ou égales à Q_1. $Q_2 = x_{0,5} = \mathrm{M_e}$ est la médiane. $Q_3 = x_{0,75}$ s'appelle le troisième quartile ; un quart des valeurs prises par X sont supérieures ou égales à Q_3



L'intervalle interquartile (IIQ) est la différence entre le troisième quartile et le premier quartile ; il s'écrit :

$$\boxed{IIQ = Q_3 - Q_1}$$

L'intervalle interquartile sert à apprécier la dispersion de X, de façon absolue, ou bien par comparaison avec une autre variable quantitative, à condition que cette dernière soit exprimée dans la même unité que X.
En effet, les valeurs Q_1 et Q_3 délimitent une plage au sein de laquelle 50% des valeurs de X sont concentrées. Plus IIQ est grand, plus X est dispersée

3. Paramètres de dispersion

On dispose d'une population de N individus, et on observe x_1,.. . , x_N les valeurs d'une variable quantitative discrète X pour ces individus.

3.1. Etendue

L'étendue est la différence entre la plus grande et la plus petite des valeurs observées :

$$e = x_{max} - x_{min}$$

Le calcul de l'étendue est très simple, il donne une première idée de la dispersion des observations

3.2. Variance

La variance est la somme des carrés des écarts à la moyenne divisée par le nombre d'observations.

$$var\ (x) = \sum_{i=1}^{n} f_i\ (\bar{x} - x_i)^2$$

On dit que la variance est la moyenne des carrés des écarts à la moyenne $\bar{x}$.
Les « écarts à la moyenne » sont les $(\bar{x} - x_i)$, les « carrés des écarts à la moyenne » sont donc les $(\bar{x} - x_i)^2$. En faisant la moyenne de ces écarts, on trouve la variance.

3.3. Écart-type

L'écart-typeest la racine carrée de la variance:

$$\sigma_x = \sqrt{var\ (x)}$$

- **L'écart moyen absolu**

L'écart moyen absolu est la somme des valeurs absolues des écarts à la moyenne divisée par le nombre d'observations :

$$e_{moy} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$$

- **L'écart médian absolu**

L'écart médian absolu est la somme des valeurs absolues des écarts à la médiane divisée par le nombre d'observations :

$$e_{med} = \frac{1}{n}\sum_{i=1}^{n}|x_i - Médiane|$$

4. Représentation graphique

4.1. Histogramme des fréquences (ou effectifs)

Nous pouvons représenter le tableau statistique par un histogramme. Nous reportons les classes sur l'axe des abscisses et, au-dessus de chacune d'elles, nous traçons un rectangle dont l'aire est proportionnelle à la fréquence fi (ou l'effectif ni) associée. Ce graphique est appelé l'histogramme des fréquences.

Exemple:

Étudions l'exemple de la variable couleurs des yeux ; on peut résumer tout cela dans le tableau récapitulatif suivant :

Tableau 05: Couleur des yeux

Couleur	Bleu	Marron	Noir	Vert
Effectif	4	3	4	4

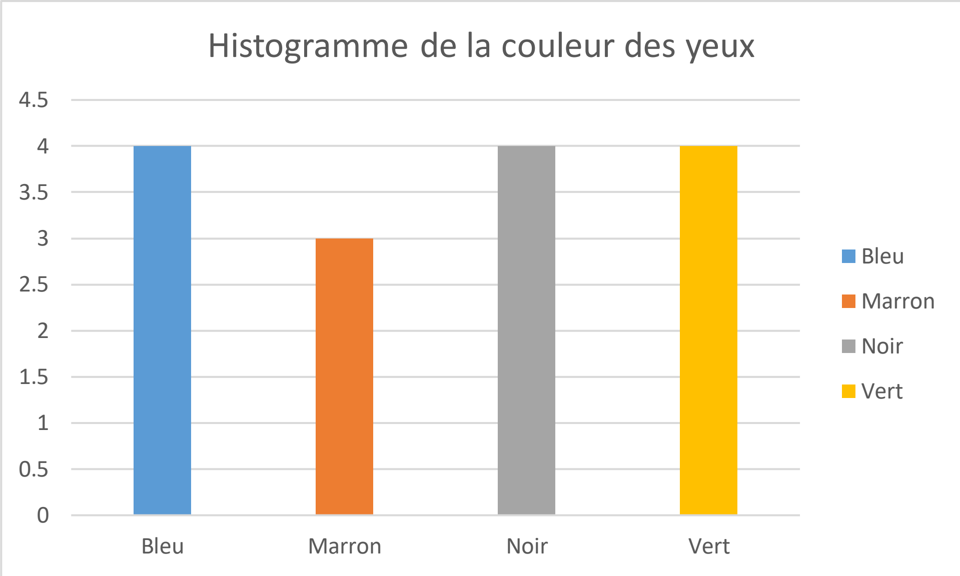


Figure 01: Histogramme des fréquences

4.2. Diagramme en bâtons

C'est un ensemble de bâtons ayant pour abscisse les valeurs $x_1,x_2,x_3.......x_n$ du caractère qui sont souvent des nombres entiers. On fait correcpondre à chaque point représentatif, un trait parallèle à l'axe des ordonnées et de longueur proportionnelle à l'effectif ni de la classe correspondante. Chaque couple (x_i, n_i) est représenté par un "bâton" d'abscisse x_i et d'ordonnée n_i. Le graphique obtenu (Fig 02) est appelé **diagramme en bâtons**.

Souvent, on complète le diagramme en traçant le **polygone statistique**, ligne polygonale obtenue en joignant les extrémités des bâtons.

Exemple:

Le tableau ci-dessous indique le nombre de médailles obtenues par la France et ses pays voisins lors des jeux Olympiques de Pékin en août 2008.

Tableau 06: Nombre de médailles lors des jeux Olympiques de Pékin 2008

Pays	Nombre de médailles
Allemagne	41
Belgique	2
Espagne	18
France	40
Grande-Bretagne	47
Italie	27
Luxembourg	0
Suisse	6

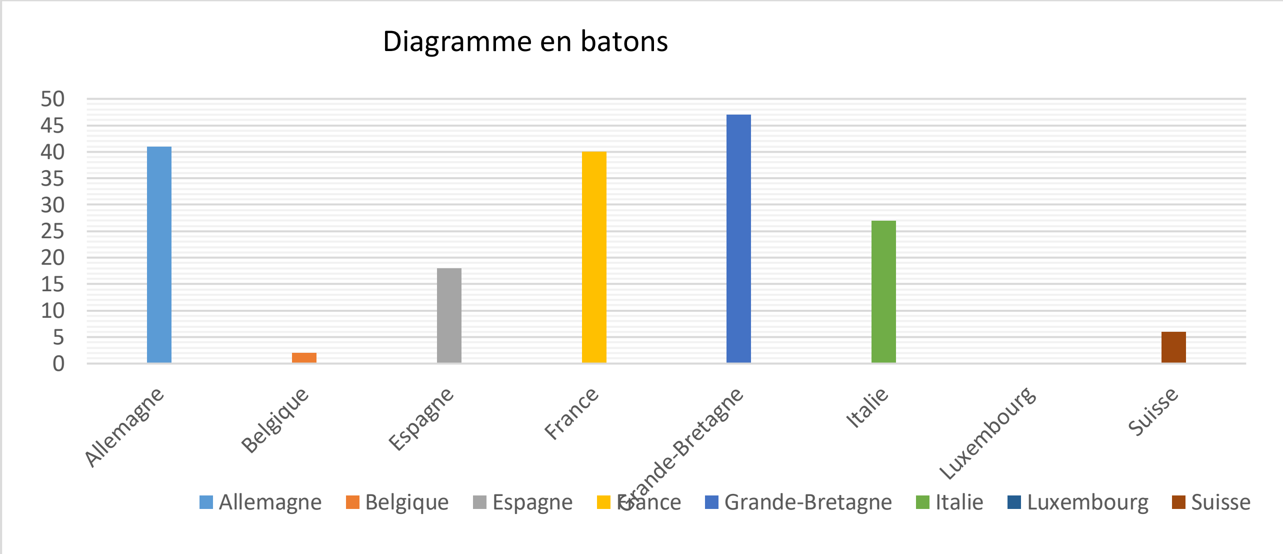


Figure 02: Diagramme en bâtons

4.3. Les diagrammes circulaires ou semi-circulaires

Les diagrammes circulaires ou semi-circulaires (fig 03) permettent de mettre en évidence la répartition de données suivant plusieurs catégories.

Exemple : Le diagramme circulaire ci-dessous représente la répartition des salariés de l'entreprise d'après le salaire mensuel.

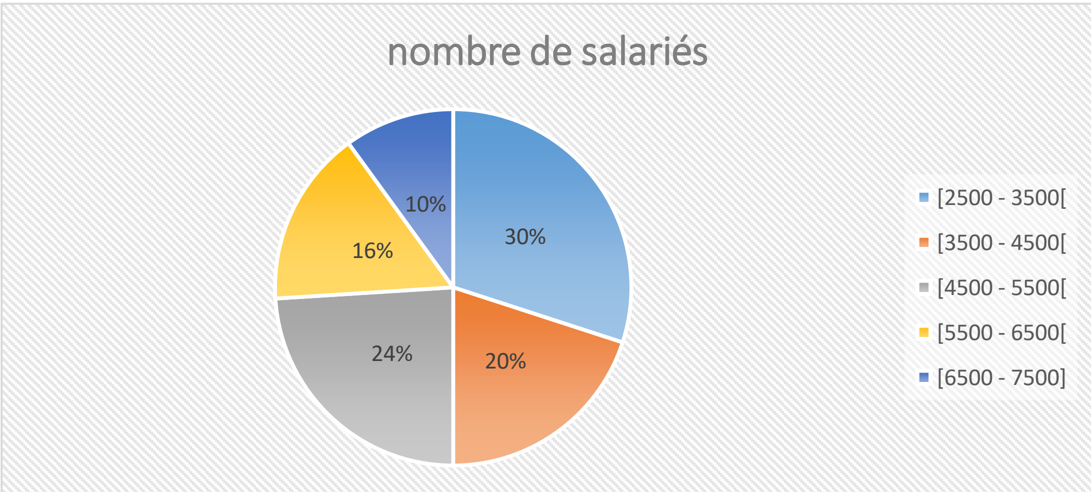


Figure 03: Diagramme circulaire

4.4. Box plots

Il est possible de résumer, sous la forme d'un graphique, l'information fournie par l'étendue, ainsi que par les trois quartiles et les intervalles qui les séparent. Ce graphique porte le nom de boîte à moustaches(fig 04) , ou encore de boîte à pattes ou diagramme en boîte (boxplot en anglais).

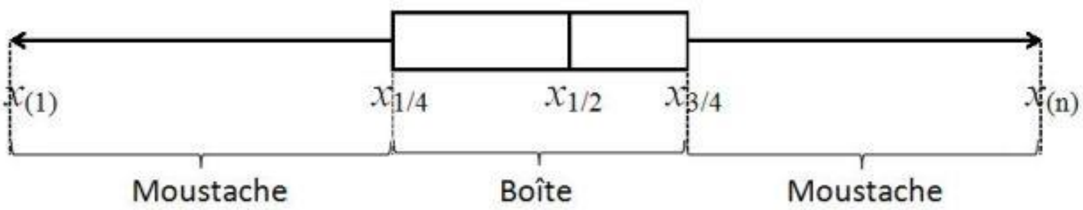
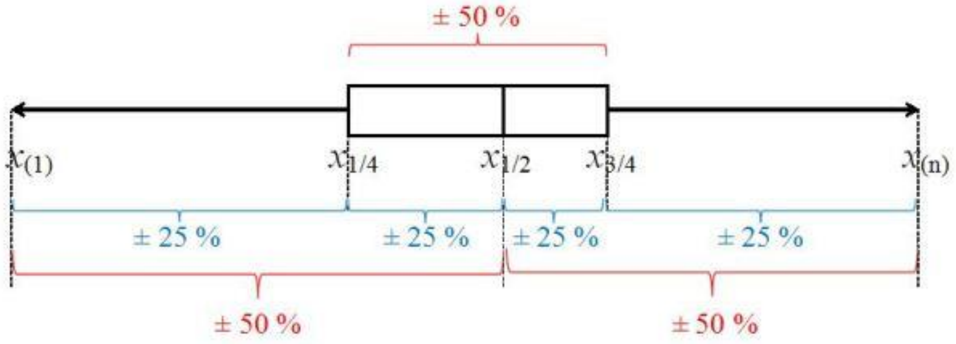


Figure 04: Boîte à moustache

Une boîte à moustaches nous indique de façon simple et visuelle quelques traits marquants de la série observée:



La médiane nous renseigne sur le milieu de la série ;
Les largeurs des deux parties de la boîte rendent compte de la dispersion des valeurs situées au centre de la série (la boîte contient 50% (environ) de l'ensemble des observations : 25% à gauche de la médiane et 25% à sa droite) ;
La longueur des moustaches renseigne sur la dispersion des valeurs situées au début de la série ordonnée (les valeurs les plus petites correspondant à 25% des observations) ou à la fin de celle-ci (les valeurs les plus grandes correspondant aussi à 25% des observations) ;
De façon générale, la boîte et les moustaches seront d'autant plus étendues que la dispersion de la série statistique est grande

Chapitre 02: Comparaison des moyennes

2.1. Hypothèses

Hypothèse nulle (H_0):

- La moyenne observée $\overline{x}$ est un estimateur de la moyenne μ_0

m étant la moyenne théorique

$\mu_0 = m$

2.2. Test bilatéral

Dans cet exemple, le test est bilatéral car la masse ne doit pas s'écarter de la norme d'un certain pourcentage, ni dans un sens ni dans autre.

On veut tester l'hypothèse nulle H_0 que la moyenne de la population $\overline{x}$ est égale à une norme *m*.

On accepte H_0 au niveau $\alpha = 0{,}05$ si pour l'échantillon donné:

$$-1{,}96 \leq \frac{\overline{x} - m}{\frac{\delta}{\sqrt{n}}} \leq +\,1{,}96$$

Si H_0 est vraie, la quantité $\frac{|\overline{x} - m|}{\frac{\delta}{\sqrt{n}}}$ suit une loi de normale si la distribution de la variable suit une loi normal $N\,(\mathrm{m}, \frac{\delta}{\sqrt{n}})$.

2.3. Test de conformité, comparaison d'une fréquence à une norme

On veut savoir si un échantillon de taille *n* est représentatif d'une population P.

Soit une population P dont la fréquence des éléments possédant le caractère C est *p*.

On considère un échantillon de taille n (n >30): la fréquence des éléments possédant le caractère C est *f*.

On compare $t_C = \frac{|p - f|}{\delta}$ à *t* théorique, noté $t_{\text{théo}}$ donnée par la table, pour α donné.

On formule l'hypothèse nulle :

(H_0) la fréquence observée *f* est égale à la norme *p*.

$t\mathrm{c} < t_{\text{théo}}$: on accepte (H_0). la différence peut être dûe aux fluctuations.

L'échantillon est représentatif.

$t\mathrm{c} > t_{\text{théo}}$: on rejette (H_0) . le hasard ne suffit d'expliquer la différence entre *p* et *f*.

L'échantillon n'est pas représentatif.

2.4. Test d'homogénéité, comparaison de deux fréquences

Pour savoir si deux échantillons de taille n_1 et n_2 ont même fréquence d'apparition du caractère C.

On considère deux échantillons de taille n_1 et n_2 avec les fréquences respectives d'apparition du caractère C, f_1 et f_2.

On utilise la variable réduite:

$$t_c = \frac{|\,f_1 - f_2|}{\sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{avec } f = \overline{\pi} = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$$

Et $$S_d = \sqrt{\overline{\pi}\,(1 - \overline{\pi})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$ l'estimateur

On formule l'hypothèse nulle:

(H_0): les fréquences d'apparition sont égales; $f_1 = f_2$. La différence observée peut être dûe aux fluctuations.

$t_c < t_{\text{théo}}$: on accept H_0. les deux échantillons sont homogènes.

$t_c > t_{\text{théo}}$: on rejette H_0. la différence est significative.

2.5. Test du X^2 (test d'homogénéité)

Ce test consiste à vérifier si une distribution expérimentale d'une carctère peut être ajustée à une distribution théorique.

Les valeurs du caractère étudiè sont réparties en *n* classes numérotés 1,2,.....i,*n* (*i* variant de 1 à *n*).

On calcule $\boldsymbol{X^2_c}$ que l'on compare au x^2 théorique, noté $x^2_{\textit{théo}}$, donné par les tables.

$$X^2_c = \sum_{i=1}^{n} \frac{(n_1 - n_2)^2}{N_i} = \sum_{i=1}^{n} \frac{(n_i - N_{Pi})^2}{N_{Pi}}$$

.

n: nombre de classes d'échantillons; $1 \leq i \leq n$
n_i: effectif de la classe *i*
N: effectif de l'échantillon considéré
P_i: fréquence théorique pour la classe *i*
N_i: N_{Pi}: effectif de l'échantillon théorique *i* de *N* membres
On formule l'hypothèse nulle:
(H_0): l'échantillon ne fournit pas d'argument contre la théorie
$X^2_C < X^2_{théo}$: On rejette l'hypothèse nulle.
$X^2_C > X^2_{théo}$: On accepte l'hypothèse nulle.

.

Chapitre 03: Analyse de la variance

3.1. Analyse de la variance* :

L'analyse *de la variance* a pour objectif de comparer des ensembles de plus de deux moyennes, en identifiant les sources de variation qui peuvent expliquer les différences existant entre elles.

3.1.1. Analyse de la variance à un seul critère de classification :

L'analyse de la variance à un critère de classification est parfois appelée aussi analyse de la variance à deux composantes, en raison du fait que la variation totale y est divisée en deux parties (variation factorielle et variation résiduelle). Elle concerne des ensembles de moyennes qui ne présentent aucune structure particulière, liée par exemple `a l'existence de deux ou plusieurs facteurs sous-jacents.

- *Décomposition de la variation totale*

Supposons que nous avons p échantillons d'effectifs $n_i (i = 1, \ldots, p)$, et que l'effectif total est $n.$:

Soit : $n. = \sum_{i=1}^{p} n_i$

L'observation est désignée par le symbole x_{ik} $(i = 1, \ldots, p$ et $k = 1, \ldots, n_i)$, la valeur x_{ik} étant donc la $k^{ème}$ observation du $i^{ème}$ échantillon.

Nous avons donc:

$$\boxed{\bar{x}_i = \frac{1}{n_i} \sum_{k=1}^{n_i} x_{ik}} \quad \text{et} \quad \boxed{\bar{x} = \frac{1}{n.} \sum_{i=1}^{p} \sum_{k=1}^{n_i} x_{ik} = \frac{1}{n.} \sum_{i=1}^{p} (n_i \bar{x}_i)}$$

$\bar{x}_i$ est la moyenne des p échantillons, et $\bar{x}$ est une moyenne générale

Il est possible de subdiviser les écarts entre les observations individuelles et la moyenne générale en deux composantes additives :

$$\boxed{x_{ik} - \bar{x} = (\bar{x}_i - \bar{x}) + (x_{ik} - \bar{x}_i)}$$

La variation totale = variation partielle, appelée aussi variation factorielle + variation résiduelle ou dans les échantillons.

En sommant pour toutes les valeurs observées, on obtient l'équation d'analyse de la variance :

$$\sum_{i=1}^{p} \sum_{k=1}^{n_i} (x_{ik} - \bar{x})^2 = \sum_{i=1}^{p} \sum_{k=1}^{n_i} (\bar{x}_i - \bar{x})^2 + \sum_{i=1}^{p} \sum_{k=1}^{n_i} (x_{ik} - \bar{x}_i)^2$$

La somme des carrés des écarts totales = une somme des carrés des écarts factorielles ou entre échantillons + une somme des carrés des écarts résiduelles ou dans les échantillons.

$$\sum_{i=1}^{p} \sum_{k=1}^{n_i} (x_{ik} - \bar{x})^2 = \sum_{i=1}^{p} \sum_{k=1}^{n_i} (\bar{x}_i - \bar{x})^2 + \sum_{i=1}^{p} \sum_{k=1}^{n_i} (x_{ik} - \bar{x}_i)^2$$

Les nombres de degrés de liberté peuvent être associés aux différentes sommes des carrés des écarts. Ces nombres de degrés de liberté sont aussi additifs et se présentent de la manière suivante.

$$\boxed{n. - 1 = (p - 1) + (n. - p)}.$$

En divisant les sommes des carrés des écarts par leurs nombres de degrés de liberté respectifs, on définit des quantités appelées carrés moyens, savoir un carré moyen total, un carré moyen factoriel ou entre échantillons, et un carré moyen résiduel ou dans les échantillons.

$$\boxed{\mathrm{CM}_t = \mathrm{SCE}_t/(n. - 1)}, \quad \boxed{\mathrm{CM}_a = \mathrm{SCE}_a/(p - 1)} \quad \text{et} \quad \boxed{\mathrm{CM}_r = \mathrm{SCE}_r/(n. - p)}.$$

L'ensemble des résultats peut être présenté sous la forme d'un tableau d'analyse de la variance :

Tableau 07: Tableau d'analyse de la variance à un critère de classification

Sources de variation	Degrés de liberté-	Sommes des carrés des encarts	Carrés moyens
Différences entre échantillons Différences entre observations (dans les échantillons)	$p - 1$ $n. - p$	*SCEa* *SCEr*	*CMa* *CMr*
Totaux	$n. - 1$	*SCEt*	

3.1.2. Modèle fixe

Pour le modelé fixe, l'analyse de la variance à un critère de classification, nous désignerons par X_{ik} les variables aléatoires qui correspondent aux valeurs observées x_{ik} , parmi les moyennes des p populations, et par m une moyenne :

$$m = \frac{1}{n_.} \sum_{i=1}^{p} (n_i \, m_i) \, .$$

On peut définir un modèle théorique, qui se présente comme suit :

$$\boxed{X_{ik} - m = (m_i - m) + (X_{ik} - m_i)} \quad \text{ou} \quad \boxed{X_{ik} - m = a_i + D_{ik}},$$

*a*s ont les écarts factoriels, non aléatoires, liés au facteur contrôlé,

D_{ik} sont les écarts résiduels, aléatoires, indépendants les uns des autres de moyenne nulle et de même variance σ^2

$$\boxed{X_{ik} = m + a_i + D_{ik}}$$,ƆE$_r$

L'hypothèse nulle est l'hypothèse d'égalité des moyennes et aussi l'hypothèse de nullité des effets principaux :

$$H_0 : m_1 = m_2 = . . . = m_p \quad \text{ou} \quad a_1 = a_2 = . . . = a_p = 0 \, .$$

Pour le test de l'hypothèse nulle émise ci-dessus, il faut le calcul de la valeur de F (une variable F de FISHER-SNEDECOR) :

$$\boxed{F_{obs} = CM_a / CM_r \, .}$$

Le rejet de l'hypothèse nulle H_0, au niveau de probabilité α quand :

$P(F \geq F_{obs}) \leq \alpha$ ou $F_{obs} \geq F_{1-a}$, avec $p - 1$ et $n. - p$ degrés de liberté.

Le rejet de l'hypothèse nulle, relative a` un ensemble de p moyennes, soulève la question de savoir quelles sont les moyennes qui diffèrent significativement les unes des autres.

Exemple

Nous voulons tester quatre types de carburateurs : *A1*, *A2*, *A3* et *A4*. Pour chaque type de carburateur nous disposons de six pièces qui sont montées successivement en parallèle sur quatre voitures que nous supposons avoir des caractéristiques parfaitement identiques.

Le tableau ci-dessous indique pour chacun des essais la valeur d'un paramètre lié à la consommation :

Tableau 08: Test d'essais de 04 types de carburateurs

Essai	*A1*	*A2*	*A3*	*A4*

1	21	23	18	20
2	24	23	19	21
3	25	32	28	25
4	20	23	19	15
5	34	32	24	29
6	17	15	14	9

Nous consultons alors le tableau de l'analyse de la variance

Facteur	*Type*	*Niveaux*	*Valeurs*
Carburant	*fixe*	*4*	*A1 A2 A3 A4*

Analyse de la variance pour Consomma, en utilisant la SC ajustée pour les tests

Source	DL	SC séq	SC ajust	CM ajust	F	P
Carburant	3	100,83	100,83	33,61	0,89	0,464
Erreur	20	757,00	757,00	37,85		
Total	23	857,83				

La $p-$valeur associée, par la statistique de Fisher, à l'hypothèse nulle H_0 est de 0, 464. Elle est strictement supérieure à α = 5 % : le test n'est pas significatif à ce seuil. Nous ne pouvons rejeter l'hypothèse nulle H_0 d'absence d'effet du facteur Carburateur sur la Consommation. Ainsi au niveau α = 5 %, le facteur

3.1.3. Le modèle aléatoire

Dans le cas du modèle aléatoire, les populations pour lesquelles des observations sont réalisées sont choisies au hasard au sein d'un ensemble très vaste, infini ou quasi infini. Les moyennes de ces populations doivent donc être considérées comme des variables aléatoires et nous les d´enseignerons par des lettres majuscules :

Le modèle théorique s'écrit :$X_{i,k} = \mu + \alpha_i + \alpha_{ik}$, $i = 1...I$, $k = 1...K$,

où $X_{i,j}$ est la valeur prise par la réponse X dans la condition A_i

$$\boxed{X_{ik} - m = A_i + D_{ik}} \quad \text{ou} \quad \boxed{X_{ik} = m + A_i + D_{ik}}.$$

Dans ce modèle, les écarts factoriels :

$$A_i = M_i - m,$$

sont ´également aléatoires, de distribution normale, de moyenne nulle et de varianceσ^2_{Ai}égale à σ^2_M:

$$m_A = 0 \quad \text{et} \quad \sigma^2_A = \sigma^2 M$$

Ces écarts factoriels sont aussi indépendants des écarts résiduelsD_{ik}.

$$\boxed{\mathrm{E(CM}_t) = \sigma^2 + n'\,(p-1)\,\sigma_A^2/(n_. - 1)}, \quad \boxed{\mathrm{E(CM}_a) = \sigma^2 + n'\,\sigma_A^2}$$

$$\boxed{\mathrm{E(CM}_r) = \sigma^2},$$

L'hypothèse à tester est sensiblement différente :

$$\mathrm{H}_0 : \sigma_M^2 = 0 \quad \text{ou} \quad \sigma_A^2 = 0$$

La variable F de $p-1$ et $n_. - p$ degrés de liberté et la variable χ^2 ayant $p-1$ degrés de liberté.

En cas des effectifs inégaux en remplaçant n par la moyenne harmonique n'' des effectifs n_i :

$$n'' = p / \sum_{i=1}^{p} \frac{1}{n_i}$$

Et F_{obs} par une valeur modifiée F'_{obs} définie de la manière suivante:

$$F'_{obs} = n''\,\mathrm{SCE}_{\bar{x}_i} / \left[(p-1)\,\mathrm{CM}_r\right]$$

$SCE_{\bar{x}_i}$ étant la somme des carrés des écarts non pondérée des moyennes $\bar{x}_i$

La relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_F + SC_R$$

Le test d'hypothèse sera comme suit :

$$H_0 : \sigma^2_A = 0 \qquad \text{contre} \qquad H_1 : \sigma^2_A \neq 0$$

Sous l'hypothèse nulle H_0 précédente d'absence d'effet du facteur *A* et lorsque les conditions de validité du modèle sont respectées, f est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I - 1$ et $n - I$ degrés de liberté.

Nous résumons les informations dans le tableau de l'ANOVA ci-dessous :

Tableau 09: Tableau de l'ANOVA

Source	Variation	Degrés de liberté	Carré Moyen	F	Décision
Facteur	sc_F	$I-1$	$s_F^2 = \dfrac{sc_F}{I-1}$	$f = \dfrac{s_F^2}{s_R^2}$	$\mathcal{H}_0$ ou $\mathcal{H}_1$
Résiduelle	sc_R	$n-I = I(J-1)$	$s_R^2 = \dfrac{sc_R}{n-I}$		
Totale	sc_{TOT}	$n-1 = IJ-1$			

Exemple

On s'intéresse à l'ensemble des prairies d'une région donnée et on souhaite identifier l'importance, absolue ou relative, de la variabilité de la production fourragère, d'une part d'une prairie à l'autre, et d'autre part d'un endroit à l'autre à l'intérieur des différentes prairies. Dans ce but, on a tout d'abord choisi au hasard trois prairies dans l'ensemble du territoire considère, puis au sein de chacune de ces trois prairies, cinq petites parcelles de 2 m^2.

Dans les trois prairies constituent trois unités du premier degré et les 15 petites parcelles 15 unités du deuxième degré.

Dans chacune des 15 parcelles, on a mesuré les rendements en matière sèche à une date donnée. Les valeurs observées, exprimées en tonnes par hectare, figurent dans le tableau 10 et les résultats de l'analyse de la variance dans le tableau11 .

Tableau 10: Etude de l'homogénéité des rendements fourragers d'un ensemble de prairies : rendements observés, en tonnes de matière sèche par hectare.

Parcelles	Prairie 1	Prairie 2	Prairie 3
1	2,06	1,59	1,92
2	2,99	2,63	1,85
3	1,98	1,98	2,14
4	2,95	2,25	1,33
5	2,70	2,09	1,83

NB : Dans la pratique, les calculs des SEC sont les mêmes.

Tableau 11: Tableau d'analyse de la variance.

Sources de variation	Degrés de liberté	Sommes des carrés des écarts	Carrés moyens	F
Différences entre prairies	2	1,3182	0,6591	4,23*
Différences entre parcelles (dans les prairies)	12	1,8711	0,1559	
Totaux	14	3,1893		

La probabilité de dépasser la valeur 4,23 étant égale à 0,041, pour une variable F de Fisher-Snedecor à 2 et 12 degrés de liberté, les différences entre prairies doivent être considérées comme juste significatives. C'est ce qu'indique d'ailleurs l'astérisque qui est associée, conventionnellement, a la valeur F_{obs} .

3.2. Analyse de la variance à deux facteurs

L'analyse de la variance à deux critères de classification, peut être considérée comme une généralisation de l'analyse à un critère, qui permet de tenir compte simultanément de deux facteurs sous-jacents, et non plus d'un seul facteur.

3.2.1. Les modèles croisés à effectifs égaux : aspects descriptifs

3.2.1.1. La décomposition de la variation totale

Considérons pq échantillons ou séries d'observations de même effectif n, et désignons les observations individuelles par x_{ijk} , les indices i, j et k étant relatifs respectivement aux différentes modalités du premier critère de classification $(i = 1,\dots\dots,p)$, aux différentes modalités du deuxième critère de classification $(j = 1,\dots\dots,q)$, et aux différentes observations d'un même échantillon ou d'une série $(k = 1,\dots\dots,n)$.

La moyenne étant :

$$\bar{x}_{ij.} = \frac{1}{n}\sum_{k=1}^{n} x_{ijk}$$

On peut calculer une moyenne pour chacune des modalité : $(i = 1,\dots\dots,p)$ et $(j = 1,\dots\dots,q)$

$$\bar{x}_{i..} = \frac{1}{qn}\sum_{j=1}^{q}\sum_{k=1}^{n} x_{ijk} = \frac{1}{q}\sum_{j=1}^{q}\bar{x}_{ij.}$$ et $$\bar{x}_{.j.} = \frac{1}{pn}\sum_{i=1}^{p}\sum_{k=1}^{n} x_{ijk} = \frac{1}{p}\sum_{i=1}^{p}\bar{x}_{ij.}$$

La moyenne générale :

$$\bar{x}_{...} = \frac{1}{p\,q\,n}\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n} x_{ijk} = \frac{1}{p\,q}\sum_{i=1}^{p}\sum_{j=1}^{q}\bar{x}_{ij.} = \frac{1}{p}\sum_{i=1}^{p}\bar{x}_{i..} = \frac{1}{q}\sum_{j=1}^{q}\bar{x}_{.j.}$$

La décomposition sera donc :

$$\begin{aligned} x_{ijk} - \bar{x}_{...} &= (\bar{x}_{ij.} - \bar{x}_{...}) + (x_{ijk} - \bar{x}_{ij.}) \\ &= (\bar{x}_{i..} - \bar{x}_{...}) + (\bar{x}_{.j.} - \bar{x}_{...}) + (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x}_{...}) + (x_{ijk} - \bar{x}_{ij.}) \end{aligned}$$

Et l'équation de l'analyse de variance sera donc :

$$\begin{aligned} \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{...})^2 &= q\,n\sum_{i=1}^{p}(\bar{x}_{i..} - \bar{x}_{...})^2 + p\,n\sum_{j=1}^{q}(\bar{x}_{.j.} - \bar{x}_{...})^2 \\ &+ n\sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x}_{...})^2 + \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{ij.})^2 \end{aligned}$$

D'une façon simplifiée :

$$\boxed{\mathrm{SCE}_t = \mathrm{SCE}_a + \mathrm{SCE}_b + \mathrm{SCE}_{ab} + \mathrm{SCE}_r}$$

avec des degrés de libertés :

$$\boxed{p\,q\,n - 1 = (p-1) + (q-1) + (p-1)(q-1) + p\,q\,(n-1)}$$

Le tableau de l'analyse de variance sera donc :

Tableau 12: Analyse de la variance

Sources de variation	Degrés de liberté	Sommes des carrés des écarts	Carrés moyens
Facteur *a*	$p - 1$	SCE_a	CM_a
Facteur *b*	$q - 1$	$SCEb$	CM_b
Interaction	$(p - 1)(q - 1)$	$SCEab$	CM_{ab}
Variation résiduelle	$p\,q\,(n - 1)$	SCE_r	CM_r
Totaux	$pq\,n - 1$	SCE_t	

Exercice :

Tableau 13: Teneurs en P_2O_5 , en mg par 100 g de terre sèche, et moyennes observées, pour deux types de sols et trois types de sondes.

	Sonde 1 (j = 1)	Sonde2 (j = 2)	Sonde 3 (j = 3)	
Sol 1 (i = 1)	43 45 46 53	41 42 43 44	42 44 46 48	
Sol 2 (i = 2)	40 40 40 43	35 37 40 40	37 39 40 40	
		$\bar{x}_{ij.}$		$\bar{x}_{i.}$
Sol 1	46,75	42,50	45,00	44,75
Sol 2	40,75	38,00	39,00	39,25
$\bar{x}_{.j.}$	43,75	40,25	42,00	42,00 ($\bar{x}_{...}$)

Ce tableau met immédiatement en évidence une composante particulièrement importante liée aux types de sols et basée sur les écarts :

$$\bar{x}_{1..} - \bar{x}_{...} = 2{,}75 \quad \text{et} \quad \bar{x}_{2..} - \bar{x}_{...} = -2{,}75$$

Les composantes moins importantes :

$$\bar{x}_{.1.} - \bar{x}_{...} = 1{,}75\,, \quad \bar{x}_{.2.} - \bar{x}_{...} = -1{,}75 \quad \text{et} \quad \bar{x}_{.3.} - \bar{x}_{...} = 0{,}00$$

Les différentes sommes sont :

$X_{11.}$ = 187, $X_{12.}$ = 170, $X_{13.}$ = 180, $X_{21.}$ = 163, $X_{22.}$ = 152, $X_{23.}$ = 156,

$X_{1..}$ = 537, $X_{2..}$ = 471, $X_{.1.}$ = 350, $X_{.2.}$ = 322, $X_{.3.}$ = 336, $X_{...}$ = 1.008.

Les sommes des carrés des écarts relatives aux six séries d'observations sont aussi :

SCE_{11} = 56,8, SCE_{12} = 5,0, SCE_{13} = 20,0,

SCE_{21} = 6,8, SCE_{22} = 18,0, SCE_{23} = 6,0.

On obtient don le tableau de l'analyse de variance :

Sources de variation	Degrés de liberté	Sommes des carrés des ´écarts	Carrés moyens
Types de sols	1	181,5	181,5
Types de sondes	2	49,0	24,5
Interaction	2	3,0	1,5
Variation résiduelle	18	112,5	6,25
Totaux	23	346,0	

- **Exemple d'un test d'ANOVA par l'utilisation du logiciel Statistica (www. Statistica.com) :**

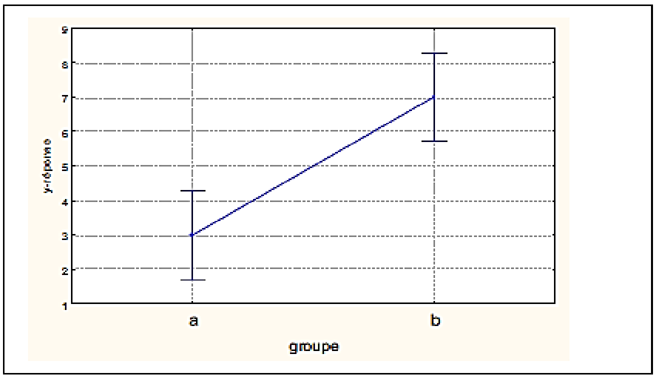
id	groupe	sexe	y-réponse
1	a	hom	1
2	a	hom	2
3	a	hom	3
4	b	hom	5
5	b	hom	6
6	b	hom	7
7	a	fem	3
8	a	fem	4
9	a	fem	5
10	b	fem	7
11	b	fem	8
12	b	fem	9

moyenne	groupe a	groupe b
hommes	2	6
femmes	4	8
tous	3	7

ANOVA

	SS	df	MS	F	p
Intercept	300	1	300	150	0.0000
groupe	48	1	48	24	0.0006
erreur	20	10	2		

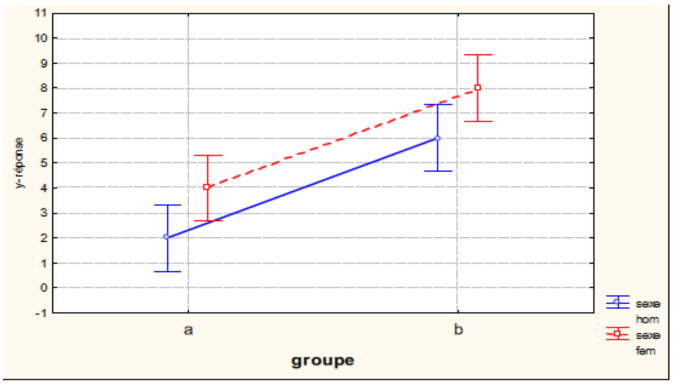
Les groupes sont différents mais une partie de cette différence est dûe au facteur sexe



- **Cas des effets principaux et effets d'interaction**

Modèle complet y = général + groupe + sexe + groupe x sexe

ANOVA					
	df	SS	MS	F	p
Intercept	1	300	300	300	0.0000
groupe	1	48	48	48	0.0001
sexe	1	12	12	12	0.0085
groupe*sexe	1	0	0	0	1.0000
erreur	8	8	1		
total	11	68			



On remarque un effet Groupe et effet sexe qui sont significatifs.

Chapitre 04: Principes d'expérimentation.

Planification des expériences et analyse de leurs résultats

.

4.1. Protocole expérimental

On peut considérer que les déférents éléments de base d'un protocole expérimental sont :

1. la définition du ou des *buts* et des *conditions* de réalisation de l'expérience (expérience en laboratoire ou en champ par exemple);
2. la définition du ou des *facteurs* (température, pression, etc.) dont on désire étudier l'influence, de ses ou de leurs modalités et des combinaisons de ces modalités, auxquelles nous associerons les notions de *traitement* ou d'*objet* et de *plan* ou de *structure des traitements* ou *objet*;
3. la définition des individus ou, d'une manière plus générale, des *unités expérimentales* qu'on se propose d'observer (telles que des patients ou des groupes de patients, des animaux ou des plantes ou des groupes d'animaux ou de plantes, des pièces de bois ou de métal, etc.);
4. la définition des *observations* (rendements par exemple) qui devront être réalisées et des modalités de collecte de ces observations;
5. la manière dont les différentes modalités ou combinaisons de modalités du ou des facteurs devront être affectées aux différentes unités expérimentales et la répartition des unes et des autres dans l'espace ou dans le temps, ce qui constitue le dispositif expérimental;
6. des informations générales relatives à l'analyse des résultats (telles qu'un schéma d'analyse de la variance et/ou la formulation d'une ou plusieurs équations de régression, y compris la spécification précise du ou des modèles envisagés).

4.2. Notion de facteur

1° Facteur, facteurs qualitatifs et quantitatifs

En matière d'expérimentation, on appelle *facteur 1* toute série d'éléments de même nature qui peuvent être comparés au cours d'une expérience, tels qu'une série de variétés, un ensemble de produits phytosanitaires, différentes doses d'un même engrais, différentes températures, différentes pressions, etc.

2° Variantes, niveaux, modalités

Les différents éléments individuels qui sont associés à chacun des facteurs sont appelés variantes, niveaux ou modalités. Le terme « *variante* » convient mieux dans le cas des facteurs qualitatifs (différentes variétés par exemple), et le terme « *niveau* » dans le cas des facteurs quantitatifs (différentes températures par exemple), tandis que le vocable *« modalités »* s'adapte bien aux deux situations.

3° Facteurs contrôlés et non contrôlés, facteurs constants

Les facteurs qui sont électivement étudiés au cours d'une expérience sont aussi appelés facteurs contrôlés ou maîtrisés. Ils s'opposent aux facteurs non contrôlés ou non maîtrisés, sur lesquels il n'est pas ou il est difficilement possible d'agir et qui sont la source de variations résiduelles, fréquemment considérées comme aléatoires.

4° Facteurs essentiels et accessoires

Les variations résiduelles dues aux facteurs non contrôlés peuvent malgré tout être maîtrisées dans une certaine mesure par la définition de blocs, de dispositifs expérimentaux constitués de lignes et de colonnes, etc.

Dans cette optique, on parle fréquemment de facteurs essentiels ou principaux à propos des facteurs qui constituent la raison d'être de l'expérience, et de facteurs accessoires ou auxiliaires à propos de ceux qui sont introduits en vue de maîtriser les variations résiduelles (facteur blocs par exemple).

4.3. Les expériences à un facteur

4.3.1. Le choix des modalités

1° Cas d'un facteur qualitatif

Dans le cas d'un facteur qualitatif unique, le problème du choix des déférentes modalités ou variantes ne se pose généralement pas, celles-ci étant définies en même temps que le but de l'expérience (comparaison de quelques variétés données de blé par exemple).

2° Cas d'un facteur quantitatif

Dans le cas d'un facteur quantitatif par contre, le problème du choix des modalités ou niveaux reste entier. Le plus souvent, les niveaux sont choisis, dans l'ensemble du domaine de variation qu'on désire étudier, selon une progression arithmétique ou selon une progression géométrique au moins approximative. Le nombre de niveaux choisis doit toujours être aussi élevé que possible, même si, en conséquence, le nombre de répétitions pour chacun des niveaux doit être considérablement réduit.

4.3.2. Notion de témoin

Le témoin en matière agronomique, ceux-ci peuvent être, par exemple, une ou quelques variétés largement utilisées dans la région considérée, un ensemble de parcelles qui ne sont soumises à aucun des traitements étudiés (parcelles sans engrais), un ensemble de parcelles qui sont soumises à un traitement classique, considéré comme point de comparaison (parcelles traitées avec un herbicide bien connu), etc.

4.4. Les expériences factorielles et factorielles fractionnaires

- ***Nombres de facteurs et de modalités***

Dans le cas des expériences qui font intervenir deux ou plusieurs facteurs, se posent non seulement le problème du choix des modalités de chacun des facteurs, y compris éventuellement un ou plusieurs témoins, mais aussi la question du choix du nombre de facteurs et du mode d'agencement des différentes modalités de chacun des facteurs avec les différentes modalités des autres facteurs.

Sur un plan théorique, on pourrait affirmer tout d'abord qu'il y a toujours intérêt à augmenter au maximum le nombre de facteurs, au même titre que le nombre de modalités de chacun des facteurs. Mais l'application de ce principe conduit très rapidement à prendre en considération un nombre considérable d'objets, alors que les moyens disponibles pour réaliser quelque expérience que ce soit sont toujours limités.

La figure suivante présente deux possibilités relatives au cas de deux facteurs à trois modalités chacun, avec un total de cinq objets. Dans cette figure, les symboles A1, A2 et A3 désignent les trois modalités du premier facteur, et B1, B2 et B3 les trois modalités du deuxième facteur.

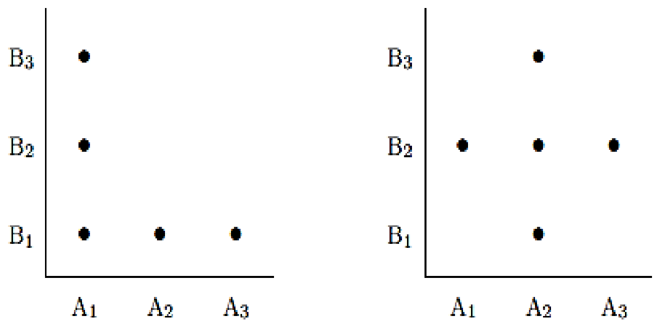


Figure 05: Représentation schématique de deux possibilités d'expériences à deux facteurs du type «un facteur à la fois ».

4.5. Les expériences factorielles complètes

Les expériences factorielles fractionnaires ou incomplètes sont en quelque sorte un intermédiaire entre les expériences de type « un facteur à la fois » et les expériences factorielles complètes. Elles ont pour principe de recourir à des sous-ensembles d'objets (ou de points expérimentaux) des expériences factorielles complètes, choisis en général de telle façon qu'il soit possible d'estimer l'effet individuel de chacun des facteurs et, éventuellement, leurs interactions d'ordre inférieur (interactions des facteurs deux à deux par exemple).

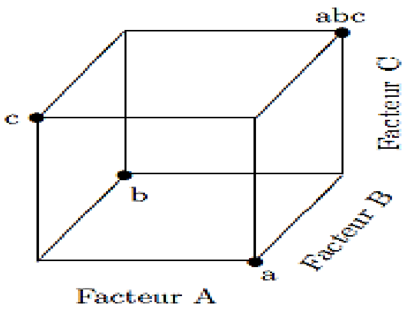


Figure 06: Représentation graphique des objets pris en considération dans une expérience factorielle fractionnaire 2^3

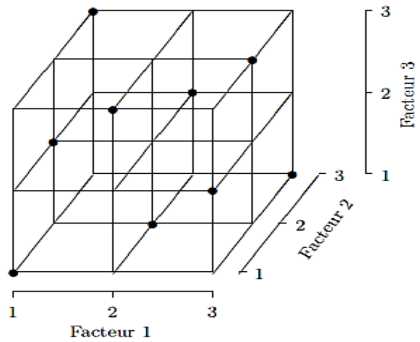


Figure 07: Représentation graphique des objets pris en considération dans une expérience factorielle fractionnaire 3^{3-1}

4.6. Notion d'unité expérimentale

L'unité expérimentale est l'´élément de base de l'expérience, qui est considère individuellement durant tout le processus expérimental. Chacune des unités est notamment soumise au départ à un objet ou un traitement particulier et conduit à une ou plusieurs observations en fin d'expérience, quand ce n'est pas déjà au début ou au cours de l'expérience.

Nous considérons à titre d'exemple le cas de lots d'animaux dont on pourrait supposer que les différents individus sont indépendants les uns des autres. Pour *p* lots de *n* animaux, en désignant par$\sigma_A{}^2$la variance entre lots et par σ la variance entre animaux dans les lots, on peut d´montrer que la variance de la moyenne générale X de tout ensemble de *pn* observations est :

$$\sigma^2_{\bar{X}} = \sigma^2_A/p + \sigma^2/(p\,n)\,.$$

4.7. Bordures et les périodes tampon

Dans les expériences relatives aux productions végétales, des interférences plus ou moins importantes peuvent exister entre parcelles voisines, soit parce que les traitements appliqués aux différentes parcelles (épandages d'engrais, pulvérisations, etc.) ne s'arrêtent pas exactement aux limites des parcelles, soit parce que le système radiculaire des plantes franchit tout naturellement ces limites, soit encore parce que se produisent des phénomènes particuliers de compétition ou de contagion entre plantes de parcelles contiguës.
Cette indique ce que peuvent être la partie observée ou utile de quatre parcelles et leurs bordures, par exemple pour quatre variétés d'une même espèce végétale, les différentes variétés étant désignées par les chiffres 1, 2, 3 et 4.

1	1	1	1	1	2	2	2	2	2
1	1	1	1	1	2	2	2	2	2
1	1	1	1	1	2	2	2	2	2
1	1	1	1	1	2	2	2	2	2
4	4	4	4	4	3	3	3	3	3
4	4	4	4	4	3	3	3	3	3
4	4	4	4	4	3	3	3	3	3
4	4	4	4	4	3	3	3	3	3

0	0	0	0	0	0	0	0	0
0	1	1	1	0	2	2	2	0
0	1	1	1	0	2	2	2	0
0	0	0	0	0	0	0	0	0
0	4	4	4	0	3	3	3	0
0	4	4	4	0	3	3	3	0
0	0	0	0	0	0	0	0	0

Figure 08: Exemples de parcelles entourées de bordures de deux types différents.

Exemple : expérience d'uniformité

Le tableau 14 contient les rendements obtenus pour des parcelles unitaires et pour l'ensemble des deux *tippings*, en tonnes par hectare.
La figure 10 fournit une représentation graphique de ces observations sous forme de « boxplot ». Elle met en évidence une certaine dissymétrie de la distribution des rendements, caractérisée notamment par la présence de quelques valeurs très élevées.
De plus, le calcul montre que le rendement moyen général est égal à 3,64 t/ha, avec une variance estimée égale à 3,49 $(t/ha)^2$, un écart-type égal à 1,87 t/ha et un coefficient de variation égal à 51,4 %.

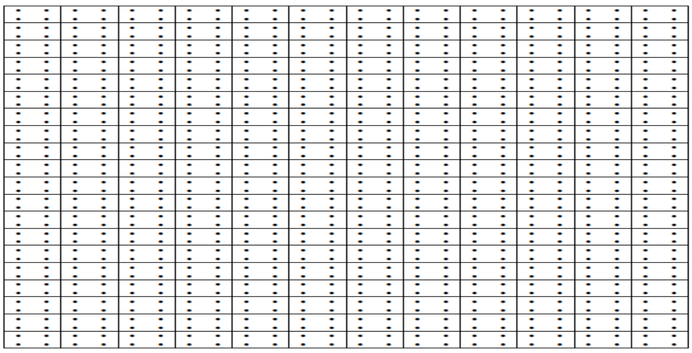


Figure 09: Expérience d'uniformité sur théier : représentation schématique de la plantation (40 lignes ⇥ 24 colonnes) et des parcelles unitaires de quatre plantes.

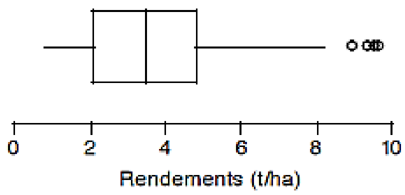


Figure 10: Présentation des rendements par parcelle sous forme de « *boxplot* ».

Tableau 14: Rendements totaux des deux *tippings*, pour 240 parcelles unitaires de quatre théiers, en tonnes par hectare

2,12	1,56	1,06	2,31	0,89	2,84	3,13	3,29	1,28	2,56	3,89	4,52
1,67	1,46	0,88	1,63	3,95	3,02	2,57	2,89	3,27	4,38	3,66	4,68
1,04	2,40	1,18	2,60	3,75	3,69	5,10	4,49	3,18	5,00	5,45	3,84
1,32	1,66	1,27	2,43	2,61	2,78	3,30	3,75	3,86	4,84	6,00	4,10
1,22	1,72	2,17	3,33	4,51	5,14	2,86	4,88	4,38	5,04	4,36	4,40
1,18	2,19	1,55	1,16	1,46	2,66	4,25	3,91	4,23	2,53	5,98	4,47
1,60	1,86	2,01	2,67	2,32	2,18	4,89	3,62	4,49	5,41	4,46	6,63
1,61	1,54	1,52	0,80	2,10	4,20	2,32	4,74	7,33	7,64	5,80	7,11
1,15	1,29	1,11	2,12	2,34	2,92	4,88	3,98	4,21	3,78	3,70	7,97
1,02	1,66	1,45	1,57	2,40	2,92	5,08	3,43	8,93	5,23	7,41	9,36
1,42	1,46	1,13	2,29	4,18	3,23	2,53	4,95	4,40	4,91	7,24	5,88
1,28	1,32	1,29	2,53	2,29	5,24	4,41	4,53	4,39	5,60	5,34	4,40
2,52	1,28	1,78	1,13	3,19	3,78	5,73	5,00	5,95	5,97	5,80	9,53
1,61	1,46	2,12	3,90	3,75	2,66	4,17	4,60	6,27	4,42	4,57	8,19
0,99	2,09	2,02	2,74	1,32	3,65	4,13	3,61	4,98	5,17	6,45	7,05
1,49	1,82	1,48	1,15	4,31	4,29	5,78	4,53	5,51	5,07	4,50	5,94
2,47	2,52	2,19	3,11	3,39	3,70	2,64	6,31	4,36	6,18	5,30	8,19
1,53	3,16	2,34	3,42	1,88	4,45	4,85	2,94	5,45	4,57	6,09	9,66
2,60	1,94	1,81	3,33	4,69	3,19	4,62	2,09	5,59	4,77	5,28	7,60
3,58	3,08	2,26	3,16	2,41	3,73	5,21	3,29	3,26	7,08	4,72	5,66

(Huit théiers), les rendements moyens suivants :

1,84	1,68	1,86	3,21	1,92	4,20
1,56	1,26	3,48	2,73	3,82	4,17

et pour des groupes de trois parcelles unitaires (12 théiers) **:**

1,58	2,01	2,57	3,66
1,34	2,87	2,91	4,24

Tableau 15: Rendements moyens par colonne et par ligne (t/ha).

Nos des col. ou lignes	Moyennes par	
1	1,67	2,45
2	1.87	2.84
3	1.63	3.48
4	2.37	3.16
5	2.89	3.67
6	3.51	2.96
7	4.12	3.51
8	4.04	3.89
9	4.77	3.29
10	5.01	4.21
11	5.30	3.64
12	6.46	3.55
13	–	4.30
14	–	3.98
15	–	3.68
16	–	3.82
17	–	4.20
18	–	4.20
19	–	3.96
20	–	3,95

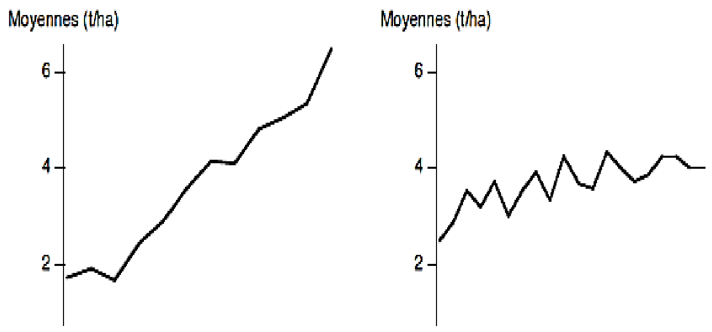


Figure 11: évolution des rendements moyens par colonne et par ligne.

Colonnes Lignes

Les variances et les écarts-types correspondant à ces regroupements sont respectivement :

	2,94 $(t/ha)^2$	et	1,71 t/ha,
ainsi que:	2,67 $(t/ha)^2$	et	1,63 t/ha

- **Interprétation des résultats**

Le tableau 16 donne les valeurs des coefficients de corrélation et des demi-moyennes des carrés des différences entre les parcelles qui occupent des lignes ou des colonnes plus ou moins éloignées les unes des autres.

Tableau 16: Coefficients de corrélation et demi-moyennes des carrés des différences entre parcelles, en fonction des distances qui les séparent.

Distances	Corrélations entre colonnes	lignes	Différences entre colonnes	lignes
1	0,681	0,70	1,07	1,05
2	0,663	0,69	1,32	1,12
3	0,618	0,75	1,81	0,90
4	0,484	0,71	2,67	1,09
5	0,471	0,71	3,40	1,09
6	0,306	0,68	4,71	1,27
7	0,209	0,64	5,95	1,46
8	0,158	0,69	7,48	1,23
9	0,250	0,67	8,62	1,36
10	0,021	0,62	9,93	1,44
11	0,179	0,64	13,24	1,36
12	–	0,62	–	1,49
13	–	0,61	–	1,40
14	–	0,60	–	1,45
15	–	0,53	–	1,62
16	–	0,56	–	1,73
17	–	0,64	–	1,55
18	–	0,52	–	1,77
19	–	0,65	–	1,68

Les tendances observées doivent en conséquence être éliminées par régression, en exprimant les rendements en fonction des coordonnées des parcelles, et l'étude géostatistique peut alors être envisagée pour les résidus de la régression.

$$y = 0{,}409 + 0{,}179\, x_h + 0{,}0536\, x_v$$

y représentant les rendements (en t/ha), et *xh* et xv les coordonnées horizontales et verticales des centres des parcelles (en m), ces coordonnées étant mesurées à partir du coin supérieur gauche du dispositif. La variance résiduelle relative à cette ´équation de régression est ´égale `a 1,14 (t/ha), alors que la variance initiale était égale à 3,49 (t/ha).

4.8. Les expériences complètement aléatoires

1° Notion d’expérience complètement aléatoire

Une première manière de se prémunir contre d’éventuelles fluctuations des facteurs non contrôlés, dans l’espace et/ou dans le temps, est de procéder à une affectation complètement aléatoire des différents objets aux différentes unités expérimentales. Cette opération de répartition « au hasard » est appelée randomisation ou, plus rarement, casualisation.
Les dispositifs expérimentaux conçus de cette manière sont dits aléatoires ou complètement aléatoires, ou encore randomisés ou complètement randomisés 2. La figure 12 en présente deux exemples.

4	1	5	2	4	3	4	4
3	5	2	6	2	7	4	8
1	9	5	10	5	11	3	12
1	13	2	14	2	15	3	16
3	17	1	18	5	19	1	20

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
22	23	23	11	13	22	21	33	13	32	12	31	33	11	12	21	32	31

Exemples d’expériences complètement aléatoires.

La première partie de cette figure illustre ce que peut être une expérience en champ qui comporterait cinq objets, numérotés de 1 à 5, et quatre répétitions de chacun des objets. Les numéros des objets sont indiqués au centre des parcelles, de forme rectangulaire, et les numéros des parcelles dans leur coin inférieur droit. L’objet 1 est ainsi affecté aux parcelles 9, 13, 18 et 20, etc.

2° Répartition « au hasard »

La répartition aléatoire ou « au hasard » des objets peut être réalisée à l’aide de tables de nombres aléatoires ou de permutations aléatoires, telles qu’il en existe dans de nombreux livres de statistique et dans certains recueils de tables.

Exemple : expérience complètement aléatoire à deux facteurs

Le facteur dimension des fragments de bois présente trois niveaux, correspondant à des cubes de 2 cm, 4 cm et 8 cm de côté. Le facteur humidité a au contraire quatre niveaux, à savoir une humidité nulle (bois anhydre) et des humidités égales à (ou proches de) 10 %, 20 % et 40 % (en pourcentages de la masse du bois anhydre). L’expérience est une expérience factorielle complète et comporte donc 12 objets.
Les résultats obtenus figurent dans le tableau 17 et, sous une autre forme, dans le tableau 18. Dans ce deuxième tableau, les colonnes D (dimensions des fragments de bois) pourraient tout aussi bien contenir un premier indice (i, variant de 1 à 3), les colonnes H (humidités des fragments de bois) pourraient contenir un deuxième indice (j variant de 1 à 4), et les colonnes Rdt pourraient alors être intitulées xijk , l’indice k étant relatif aux répétitions.

Tableau 17: Étude de la carbonisation du bois de hêtre: rendements en charbon de bois, en pourcentages du bois anhydre (première présentation).

Dim.(cm)	Humidité (%)			
	0	10	20	40
2	30,00	29,82	29,27	33,11
	29,67	29,71	30,11	30,18
	29,78	29,87	30,58	29,16
4	29,38	29,11	29,98	29,31
	28,98	29,18	30,02	29,22
	29,82	30,22	29,49	29,93
8	29,11	28,93	28,67	29,13
	29,78	29,78	29,44	29,42
	29,11	28,84	30,33	29,73

La deuxième présentation, constituée de quatre colonnes (D, H, *k*et Rdt, ou *i*, *j*, *k*et $x_{ijk)}$ et 36 lignes, est la plus couramment utilisée en vue du traitement des données par ordinateur.

Tableau 18: Étude de la carbonisation du bois de hêtre : rendements en charbon de bois, en pourcentages du bois anhydre (deuxième présentation)

D	H	*k*	Rdt	D	H	*k*	Rdt	D	H	*k*	Rdt
2	0	1	30,00	4	0	1	29,38	8	0	1	29,11
2	0	2	29,67	4	0	2	28,98	8	0	2	29,78
2	0	3	29,78	4	0	3	29,82	8	0	3	29,11
2	10	1	29,82	4	10	1	29,11	8	10	1	28,93
2	10	2	29,71	4	10	2	29,18	8	10	2	29,78
2	10	3	29,87	4	10	3	30,22	8	10	3	28,84
2	20	1	29,27	4	20	1	29,98	8	20	1	28,67
2	20	2	30,11	4	20	2	30,02	8	20	2	29,44
2	20	3	30,58	4	20	3	29,49	8	20	3	30,33
2	40	1	33,11	4	40	1	29,31	8	40	1	29,13
2	40	2	30,18	4	40	2	29,22	8	40	2	29,42
2	40	3	29,16	4	40	3	29,93	8	40	3	29,73

Analyse des résultats

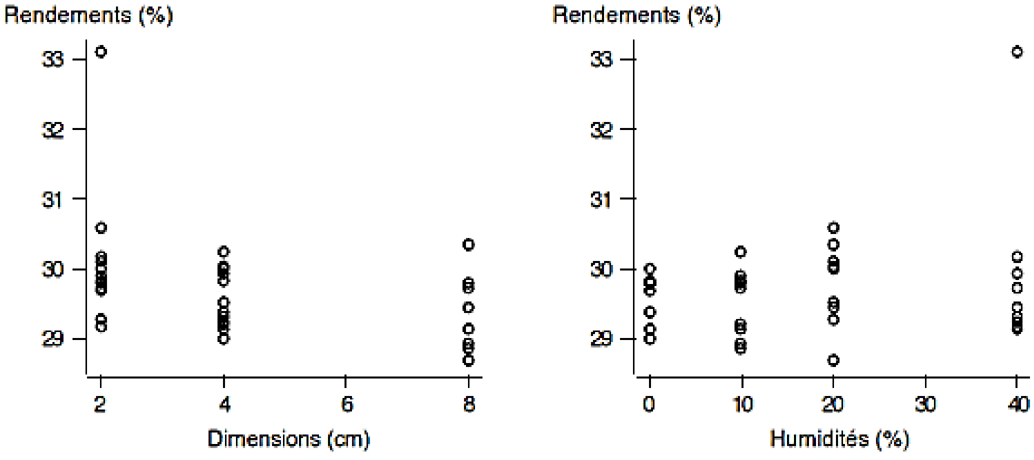


Figure 13: Étude de la carbonisation du bois de hêtre : représentation graphique de l'influence des deux facteurs considérés.

Analyse de la variance:

Sources de variation	Degrés de liberté	Sommes des carrés écarts	Carrés moyens	*F*	P
Dimension	2	1,2926	0,6463	2,72	0,087
Humidité 3		0,4367	0,1456	0,61	0,61
Dimension-humidité 6		0,1947	0,0325	0,14	0,99
Variation résiduelle	23	5,4675	0,2377		
Totaux	34	7,3915			

L'analyse de la variance des données initiales ne fait apparaître aucune déférence significative, dans les limites du domaine considéré, ce qui irait à l'encontre de l'impression donnée par la figure 14.

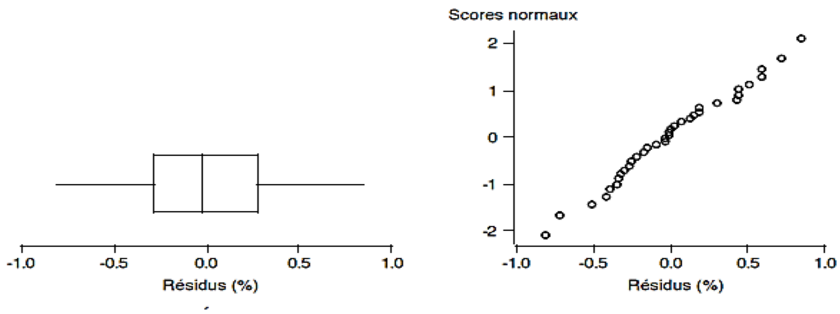


Figure 14: Étude de la carbonisation du bois de hêtre : représentation graphique des résidus de l'analyse de la variance sous la forme de « *boxplot* » et de diagramme de probabilité.

Il existe bien une relation significative entre la dimension des fragments de bois et le rendement. La somme des c5rrés des écarts relative à la relation linéaire entre ces deux variables est en effet égale à 1,1978, ce qui donne, par comparaison avec le carré moyen résiduel de l'analyse de la variance, une valeur de la variable *F* égale à 5,04 et une probabilité égale à 0,035.
L'équation de régression linéaire, qui est valable uniquement pour des dimensions allant de 2 à 8 cm, est :

$$Rdt = 29{,}917 - 0{,}0731\ D,$$

si on utilise les notations du tableau 18.

Exemple 2: expérience complètement aléatoire à quatre facteurs

Le tableau 19 présente les facteurs étudiés et leurs différents niveaux. Il comprend également les lettres A, B, C et D, que nous utiliserons dans la suite pour désigner les quatre facteurs :

Tableau 19: Étude des conditions de production d'une enzyme: facteurs pris en considération et niveaux de ces facteurs.

Facteurs	**Symboles**	**Niveaux (%)** -1	0	+1
Glucose	A	0,25	0,5	1
Lactosérum	B	1	2	3
Sulfate ammonique	C	0	0,4	0,8
Corn-steep	D	0,25	0,5	1

Le premier élément qui apparaît, à la vue du tableau 20, est la très grande dispersion des 40 observations individuelles, qui s'étendent de 0 à 1.243 U/ml. Leur moyenne générale est égale à 490 U/ml et leur écart-type estimé à 375 U/ml.

La figure 15 présente également les valeurs observées en fonction des niveaux des différents facteurs. Les quatre diagrammes de cette figure, qui ne tiennent aucun compte des éventuelles interactions, semblent indiquer que les facteurs A (glucose), B (lactosérum) et D (corn-steep) pourraient avoir chacun une influence significative.

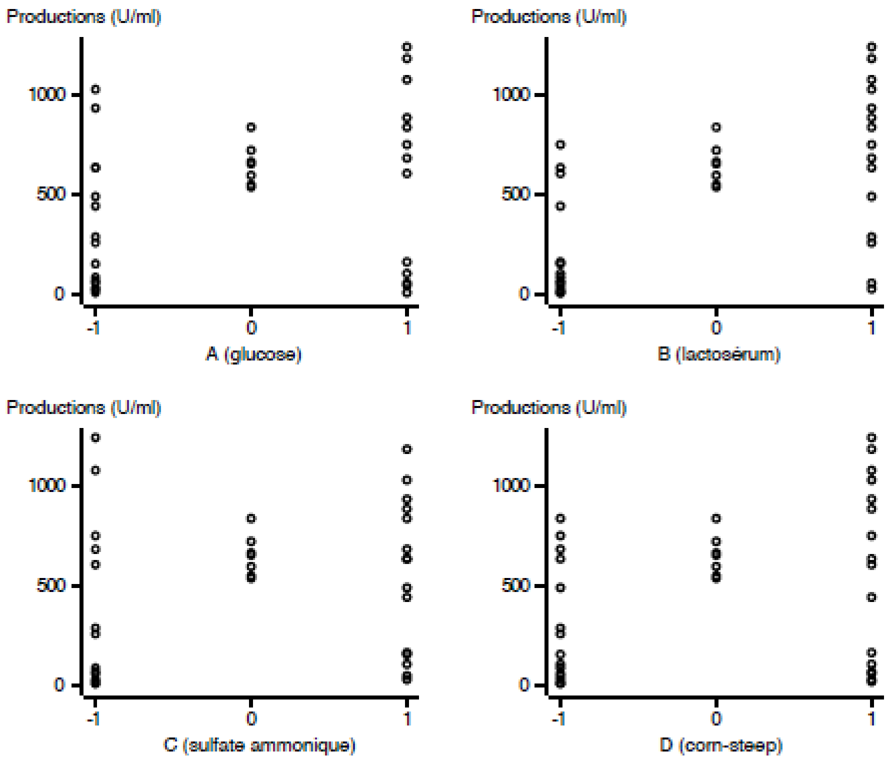


Figure 15: Etude des conditions de production d'une enzyme : représentation graphique de l'influence des différents facteurs.

Tableau 20: Étude des conditions de production d'une enzyme : niveaux des facteurs, symboles des points factoriels, valeurs observées et moyennes de ces valeurs, en unités par millilitre (U/ml).

A	B	C	D	Symb.	Valeurs obs.		Moy.
—1	—1	—1	—1	(1)	77	0	38,5
—1	—1	—1	+1	d	61	15	38,0
—1	—1	+1	—1	c	144	24	84,0
—1	—1	+1	+1	cd	442	634	538,0
—1	+1	—1	—1	b	252	279	265,5
—1	+1	—1	+1	bd	53	24	38,5
—1	+1	+1	—1	bc	482	633	557,5
—1	+1	+1	+1	bcd	1.028	936	982,0
+1	—1	—1	—1	a	54	0	27,0
+1	—1	—1	+1	ad	601	745	673,0
+1	—1	+1	—1	ac	102	44	73,0
+1	—1	+1	+1	acd	156	101	128,5
+1	+1	—1	—1	ab	750	680	715,0
+1	+1	—1	+1	abd	1.243	1.077	1.160,0
+1	+1	+1	—1	abc	682	831	756,5
+1	+1	+1	+1	abcd	883	1.184	1.033,5
0	0	0	0	–	837	658	747,5
0	0	0	0	–	590	546	568,0
0	0	0	0	–	839	538	688,5
0	0	0	0	–	717	655	686,0

Analyse de la variance

Tableau 21: Étude des conditions de production d'une enzyme : tableau d'analyse de la variance.

Sources de var.	Degrés de lib.	S. des carrés des ´écarts	Carrés moyens	*F*	P	*F*23	P23
A	1	512.325	512.325	45,2**	0,0067	50,8***	0,0000
B	1	1.909.546	1.909.546	169***	0,0010	189***	0,0000
C	1	179.251	179.251	15,8*	0,028	17,8***	0,0003
D	1	537.944	537.944	47,5**	0,0063	53,4***	0,0000
AB	1	327.443	327.443	28,9*	0,013	32,5***	0,0000
AC	1	698.857	698.857	61,7**	0,0043	69,3***	0,0000
AD	1	74.594	74.594	6,59	0,083	7,40*	0,012
BC	1	152.214	152.214	13,4*	0,035	15,1***	0,0007
BD	1	6.933	6.933	0,61	0,49	0,69	0,42
CD	1	15.095	15.095	1,33	0,33	1,50	0,23
ABC	1	9.557	9.557	0,84	0,43	0,95	0,34
ABD	1	9.557	9.557	0,84	0,43	0,95	0,34
ACD	1	434.545	434.545	38,4**	0,0085	43,1***	0,0000
BCD	1	47.972	47.972	4,24	0,13	4,76*	0,040
ABCD	1	6.356	6.356	0,56	0,51	0,63	0,44
Non-lin.	1	333.336	333.336	29,4*	0,012	33,1***	0,0000
Var. r´es.	3	33.967	11.322	1,14	0,36		
Fioles	19	5.289.492					
Mesures	20	197.933	9.897				
Totaux	39	5.487.425					

L'interprétation des résultats de l'analyse de variance est rendue difficile par la présence d'une interaction importante de trois facteurs (interaction ACD). Cette interaction, à laquelle il n'est pas possible de donner une explication simple, met en cause la validité des tests relatifs aux trois facteurs qui y interviennent. Les moyennes observées pour les niveaux -1 et +1 de ce facteur sont respectivement égales à 200 et à 689 U/ml, avec donc une différence égale à 489 U/ml, et la moyenne intermédiaire relative au niveau 0 est égale à 672 U/ml.

On peut remarquer en outre que la composante de non-linéarité est, elle également, significative. Ce fait est certainement lié en partie à la non-linéarité de l'influence du facteur B (valeur intermédiaire égale à 672, qui est loin de se situer à mi-chemin entre les valeurs extrêmes 200 et 689).

Comme la variance résiduelle n'est pas significativement supérieure aux différences entre mesures, et comme les deux carrés moyens correspondants sont du même ordre de grandeur

(11.322 et 9.897), on peut envisager aussi de regrouper ces carrés moyens de manière à définir une nouvelle composante résiduelle, avec 23 degrés de liberté:

$$(33.967 + 197.933)/(3 + 20) = 10.083 .$$

Les résultats des tests réalisés en fonction de cette composante sont donnés dans les colonnes F23 et P23 du tableau 21.

Cette façon de procéder rend les tests sensiblement plus puissants, en raison de la présence de 23 degrés de liberté, au lieu de 3, au dénominateur des variables de FISHER-SNEDECOR.

Dans le cas présent, cette procédure ne fait cependant que compliquer encore l'interprétation des résultats.

4.9. Les expériences en blocs aléatoires complets

Dans le cas des expériences en champ, au sens large (champ, verger, forêt, etc.), on entend classiquement par bloc un ensemble de parcelles voisines et très semblables les unes aux autres, quant aux conditions de croissance et de développement de la végétation. Ces blocs sont dits complets quand tous les objets mis en expérience sont présents dans chacun d'eux, le nombre de parcelles par bloc étant au moins égal au nombre d'objets.

La répartition des objets au sein des déférents blocs se fait normalement de façon complètement aléatoire et indépendamment d'un bloc à l'autre, d'où la notion de blocs aléatoires complets, aussi appelés blocs randomisés et parfois blocs FISHER.

La forme des blocs doit toujours être définie en même temps que la forme des parcelles, de manière à garantir une similitude aussi grande que possible des conditions de croissance et de développement à l'intérieur de chacun des blocs.

3	4	4	6
6	3	5	3
1	5	2	1
2	1	7	5
7	7	3	2
4	2	1	4
5	6	6	7

Bloc 1 Bloc 2 Bloc 3 Bloc 4

Gradient

Figure 16: Exemple de forme et d'orientation des parcelles et des blocs en présence d'un gradient (de fertilité ou de toute autre nature).

Analyse des résultats

- **Cas d'un seul facteur**

Le facteur blocs est considéré le plus souvent comme aléatoire, dans la mesure où on ne cherche pas à obtenir des conclusions relatives aux seuls blocs étudiés, mais bien à une population ou à un ensemble plus vaste, dont les quelques blocs étudiés sont représentatifs. Il en résulte que le modèle d'analyse de la variance pris en considération est en général un modèle croisé mixte ou aléatoire.

- **Cas de Deux ou plusieurs facteurs**

Pour deux ou plus de deux facteurs, l'analyse de la variance est aussi la méthode d'analyse qui est sans doute la plus utilisée. Dans le cas des expériences factorielles complètes par, il s'agit normalement de modèles croisés mixtes ou aléatoires d'analyse à trois ou plus de trois critères de classification.

Avantages

Les expériences en blocs aléatoires complets présentent de multiples avantages. Elles sont en premier lieu souvent aussi simples, voire même plus simples à organiser que les expériences complètement aléatoires.

Elles peuvent en outre être réalisées pour n'importe quel nombre d'objets et n'importe quel nombre de répétitions, la seule restriction étant que le nombre de répétitions doit être, d'une façon générale, identique pour les déférents objets. Il est toutefois possible de ne pas respecter entièrement cette dernière contrainte en attribuant à certains objets (témoin par exemple) deux ou plusieurs unités expérimentales par bloc, ce qui conduit à adopter pour ces objets un nombre de répétitions qui est un multiple du nombre de blocs.

- **Efficacité relative**

Considérons une expérience en blocs aléatoires complets qui comporterait p objets et q blocs, et désignons par SCE_b et $SCEa_b$ respectivement la somme des carrés des écarts relative au facteur blocs et la somme des carrés des écarts relative à l'interaction objets-blocs. Le carré moyen qui sert de base de comparaison dans l'analyse de la variance à deux critères de classification est dans ces conditions:

$$CM_{ab} = SCE_{ab} /[(p - 1)(q - 1)],$$

Tandis que le carré moyen résiduel de l'analyse de la variance à un critère de classification, qui ne tiendrait compte que du seul facteur objets, en négligeant le facteur blocs, serait :

$$CM_r = (SCE_b + SCE_{ab}) /[(q - 1) + (p - 1)(q - 1)] .$$

Le rapport des carrés moyens, et donc l'efficacité relative, est en conséquence :

$$CM_r/CM_{ab} = (p - 1)(SCE_b / SCE_{ab} + 1)/p.$$

Cette formulation montre bien que le gain de précision ou de puissance du dispositif en blocs aléatoires complets, par rapport au dispositif complètement aléatoire, est d'autant plus important que le facteur blocs, matérialisé par la somme des carrés des écarts SCE_b, est lui-même important.

Cette présentation de l'efficacité relative n'est cependant qu'approchée et conduit à une surestimation, qui ne peut pas être négligée quand le nombre d'objets et le nombre de blocs sont réduits.

Exemple : expérience en blocs aléatoires complets (planification) :

Nous supposons que la plantation de théiers présentée à la figure 09 est destinée à l'installation d'une expérience qui comporte cinq objets (quatre fumures et un témoin, par exemple), et nous nous basons sur les données étudiées antérieurement et relatives à deux tippings

La figure 17 présente neuf découpages déférents du terrain d'expérience en blocs de cinq parcelles, l'étendue des parcelles allant de 2,88 m^2 (2,4 m sur 1,2 m, soit quatre théiers par parcelle) à 46,08 m^2 (9,6 m sur 4,8 m, soit 64 théiers par parcelle), et corrélativement, le nombre de blocs variant de 48 à 3.

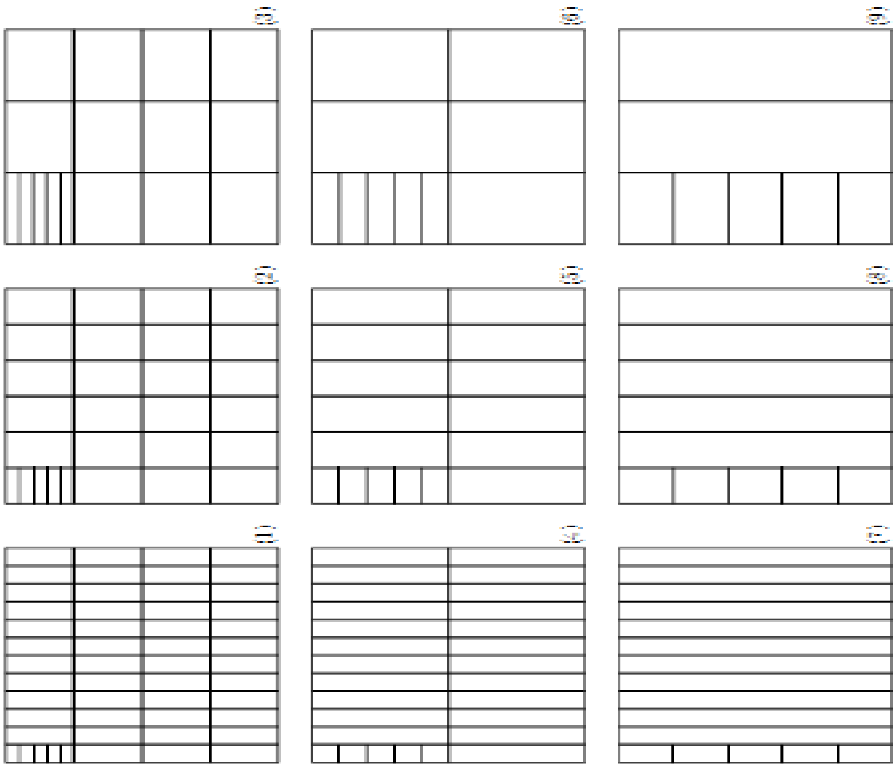


Figure 17: Planification d'une expérience sur théier : déférents découpages du terrain en blocs de 05 parcelles

Dans ce cas de la figure 17 (parcelles expérimentales confondues avec les parcelles unitaires initiales), la somme des carrés des écarts totale et la variance ou le carré moyen correspondant sont :

SCE = 834,2238 et CM = 834,2238/239 = 3,49 ; la somme des carrés des écarts et le carré moyen entre blocs sont :

SCE_b = 630,4548 et CM_b= 630,4548/47 = 13,41; tandis que la somme des carrés des écarts et le carré moyen dans les blocs ou résiduels sont :

SCE_r = 203,7690 et CM_r = 203,7690/192 = 1,0613.

La valeur 3,4905 donne une idée de ce que serait la variabilité résiduelle d'une expérience complètement aléatoire dont les parcelles auraient une étendue de 2,88 m2 (quatre théiers par parcelle). Il y correspond un écart-type égal à 1,87 t/ha, un coefficient de variation égal à 51,4 %, une erreur standard des différences de moyennes entre objets égale à 1,87//24 ou 0,38 t/ha, et un coefficient de variation de ces différences égal à 10,5 %.

De même, la valeur 1,0613 donne une idée de ce que pourrait être, dans les mêmes conditions, la variabilité résiduelle d'une expérience en blocs aléatoires complets. On peut en déduire, entre autres choses, le coefficient de variation résiduelle, égal à 28,3 % au lieu de 51,4 %, et le coefficient de variation des différences de moyennes, égal à 5,8 % au lieu de 10,5 %.
Le rapport :

3,4905/1,0613 = 3,29 ou 329 %,

est une estimation de ce que serait, dans ces conditions, l'efficacité relative de l'expérience en blocs, par comparaison avec l'expérience complètement aléatoire.

Tableau 22 : résultats de différents découpages du terrain dans le cas de parcelles sans bordures

Numéros de la fig. 17	Dimensions des parcelles (m)	Nombres de théiers par parc.	Nombres de blocs	Coeff. de variation résiduelle	C. de var. des différe. de moy.	Efficacité relative (%)	Différences significat. (%)
1	2,4 x 1,2	4	48	28,3	5,8	329	12 à 19
2	4,8 x 1,2	8	24	21,1	6,1	500	12 à 20
4	2,4 x 2,4	8	24	23,4	6,8	413	14 à 22
3	9,6 x 1,2	16	12	17,6	7,2	641	14 à 23
5	4,8 x 2,4	16	12	17,7	7,2	642	14 à 23
7	2,4 x 4,8	16	12	18,5	7,6	593	15 à 24
6	9,6 x 2,4	32	6	16,2	9,4	714	19 à 30
8	4,8 x 4,8	32	6	15,6	9,0	764	18 à 29
–	7,2 x 4,8	48	4	14,9	10,6	730	21 à 34
9	9,6 x 4,8	64	3	15,3	12,5	792	25 à 40

Bien sûr, ces considérations théoriques ne doivent pas faire oublier les contingences pratiques et, dans cette optique, il serait sans doute excessif d'envisager la possibilité de réaliser une expérience qui comporterait 24 ou 48 blocs.
Dans la pratique courante, le choix de l'expérimentateur « traditionnel » se porterait vraisemblablement, a priori, sur un dispositif comportant quatre ou six blocs, mais il nous paraît opportun de voir néanmoins quel serait le gain de précision qui pourrait être attendu en faisant l'effort, s'il est réalisable, de passer à 12 blocs.

L'efficacité relative du dispositif n° 3 de la figure 18 (12 blocs) par rapport au dispositif n° 6 (six blocs), par exemple, peut être déterminée par une simple comparaison des carrés des coefficients de variation des différences de moyennes (tableau 22) :

9,42/7,22 = 1,70 ou 170 %,

ou avec plus de précision :

(16,22/6)/(17,62/12) = 1,69 ou 169 %.

Des résultats quasi identiques peuvent être obtenus en comparant le dispositif n° 5 (12 blocs également) et le dispositif n° 6 (efficacité relative égale à 168 %, au lieu de 169 %).
Ces différentes valeurs indiquent qu'il faudrait augmenter de près de 70 % le nombre de répétitions du dispositif n°6, pour atteindre la précision du dispositif N°3 ou du dispositif n°5, c'est-à-dire aussi qu'il y a pratiquement équivalence entre les dispositifs 3 et 5 (60 parcelles de 16 théiers, soit 960 théiers), et 10 répétitions (6×1,70 = 10,2) semblables à celles du dispositif n°6 (50 parcelles de 32 théiers, soit 1.600 théiers).
Par rapport aux dispositifs 3 et 5, le choix de l'expérimentateur qui ne souhaiterait pas dépasser six répétitions est donc : adopter le dispositif n°6, avec une perte d'efficacité de l'ordre de 40 % (1/1,70 = 0,59), ou vouloir conserver la même efficacité et organiser alors, si c'est possible, une expérience sensiblement plus étendue (1.150 m^2 au lieu de 690 m^2).

Ce type de comparaison conduirait à une conclusion plus nette encore pour le dispositif qui comporte quatre blocs (non représenté dans la figure 17), l'effica-citérelative par rapport aux dispositifs 3 et 5 étant alors (tableau 22) :

$$10{,}6^2/7{,}2^2 \text{ ou } (14{,}9^2/4)/(17{,}6^2/12) = 2{,}15 \text{ ou } 215\ \%.$$

Ces considérations, relatives évidemment à un cas bien particulier, mais loin d'être exceptionnel (des coefficients de variation résiduelle de l'ordre de 15 à 20 % ne sont en effet pas exceptionnels), montrent combien il peut être dangereux de s'en tenir de façon systématique à de petits nombres de répétitions (quatre, cinq ou six blocs), comme il est de pratique assez courante en matière agronomique.

Chapitre 05: Les expériences en parcelles divisées (split-plot) et en bandes croisées (Slpit-Block)

5.1 Principes

5.1.1 Les expériences en blocs aléatoires complets et parcelles divisées

1° Deux facteurs

Nous introduisons le principe des *parcelles divisées «en anglais: Split Plot »* en considérant tout d'abord le cas le plus classique des expériences factorielles organisées en blocs aléatoires complets.

Le premier facteur possède six modalités et le deuxième facteur en possède trois. L'expérience comporte quatre blocs, correspondant aux quatre quadrants, et la répartition des 18 objets au sein des blocs a été réalisée en deux temps, correspondant aux deux facteurs.

Dans un premier temps, on a défini six parcelles au sein de chacun des quatre blocs, et on leur a attribué les différentes modalités du premier facteur (différentes variétés, dans l'exemple considéré), comme dans toute expérience en blocs aléatoires complets qui ne ferait intervenir qu'un seul facteur. Il s'agit des premiers chiffres (1 à 6) des nombres de deux chiffres qui identifient les différents objets. A ce stade, on ne tient donc aucun compte de l'existence du deuxième facteur.

Dans un deuxième temps, on a divisé chacune des 24 premières parcelles en trois parcelles plus petites, et on a réparti les trois modalités du deuxième facteur (différentes fumures, dans l'exemple considéré) au sein de chacun des 24 groupes de trois parcelles, comme s'il s'agissait de 24 blocs (deuxièmes chiffres, à savoir 1 à 3, des nombres qui désignent les différents objets). A ce stade de la randomisation, on néglige donc tout à fait le premier facteur.

D'une manière générale, pour deux facteurs présentant respectivement p et *q* modalités et pour r blocs, la première étape est une répartition classique des p modalités du premier facteur au sein des r blocs, dans p r parcelles de relativement grande taille. La deuxième étape consiste ensuite en une répartition aléatoire et indépendante des *q* modalités du deuxième facteur à l'intérieur de ces p r parcelles, en constituant un total de *p q* r parcelles de plus petite dimension.

Les premières parcelles, de plus grande taille, sont appelées *grandes parcelles* et parfois *sous-blocs*, tandis que les secondes, de plus petite taille, sont qualifiées de *petites parcelles* ou *sous-parcelles*.

D'une manière plus générale, en dehors des expériences en champ, en forêt, etc., on parle aussi d'unités du premier degré et d'unités du deuxième degré.

Dans une expérience qui serait relative par exemple à la fabrication de papier, les quatre blocs pourraient consister en quatre journées de travail, les six « grandes parcelles » pourraient correspondre à six types de matière première, pour chacun desquels une fabrication serait réalisée au cours de chacun de des quatre journées, et les « petites parcelles » pourraient correspondre à trois modalités de séchage du papier.

2° Trois et plus de trois facteurs

Le principe des blocs aléatoires complets avec parcelles divisées peut être étendu de trois façons différentes au cas de trois facteurs.

Pour une expérience factorielle comportant trois facteurs, respectivement à p, p' et p'' modalités, et r blocs, on peut tout d'abord associer les différentes modalités d'un seul facteur aux grandes parcelles (ou aux unités du premier degré), puis toutes les combinaisons des deux autres facteurs aux petites parcelles (ou aux unités du deuxième degré), en définissant ainsi p r grandes parcelles et $p\,p'\,p''$ r petites parcelles.

On peut aussi associer toutes les combinaisons des modalités de deux facteurs aux grandes parcelles, puis les différentes modalités du troisième facteur aux petites parcelles, en définissant pp' r grandes parcelles et, aussi, $pp'\,p''$ r petites parcelles. On notera que les grandes parcelles de la première option sont en fait plus étendues que celles de la deuxième option, puisqu'elles réunissent chacune $p'\,p''$ petites parcelles, au lieu de p'' petites parcelles.

Mais, toujours pour une expérience qui comporterait trois facteurs, on peut envisager en outre d'effectuer la répartition des objets en trois temps, en définissant des parcelles divisées à deux niveaux 6.

Les différentes modalités d'un premier facteur sont alors réparties tout d'abord à l'intérieur des r blocs, dans p r grandes parcelles (ou unités du premier degré). Les différentes modalités d'un deuxième facteur sont ensuite réparties à l'intérieur des p r grandes parcelles, en pp' r parcelles intermédiaires ou sous-parcelles (ou unités du deuxième degré). Et les différentes modalités du troisième facteur sont enfin réparties à l'intérieur des pp' r parcelles intermédiaires, en pp´ p˝ r petites parcelles ou sous-sous-parcelles (ou unités du troisième degré) 7, toutes les répartitions étant effectuées « au hasard » et indépendamment les unes des autres.

5.1.2 Autres dispositifs expérimentaux en parcelles divisées

1° Dispositif complètement aléatoire

Le principe des parcelles divisées peut être associé aussi au cas des expériences complètement aléatoires. Comme première illustration, considérons par exemple une expérience d'alimentation de truies et de porcelets.

Dans cette optique, nous supposons qu'on souhaite étudier la croissance de porcelets en fonction, à la fois, de l'alimentation qui est donnée à leurs mères durant leurs gestations ou leurs périodes d'allaitement, et de l'alimentation qui est donnée ensuite aux porcelets eux-mêmes, après sevrage.

Si on considère deux alimentations différentes des truies et quatre alimentations différentes des porcelets, et si on dispose de 10 truies, on peut répartir « au hasard » dans un premier temps les deux alimentations entre les 10 truies, puis choisir quatre porcelets dans chacune des portées, et répartir entre eux au hasard et indépendamment, dans chacune des portées, les quatre alimentations destinées aux porcelets. Il s'agirait ainsi d'une expérience factorielle 2× 4, avec cinq répétitions et utilisant le principe des parcelles divisées.

2° Carré latin et blocs incomplets

Le principe des expériences en parcelles divisées peut être adapté à de nombreux autres dispositifs expérimentaux, dont les dispositifs en carré latin et en blocs incomplets.

Dans le cas du carré latin 4 × 4 qui est présenté à la figure 19 et qui fait intervenir un facteur objets, un facteur lignes et un facteur colonnes, on pourrait subdiviser par exemple chacune des 16 parcelles en trois sous-parcelles, en vue d'introduire dans l'expérience un facteur supplémentaire à trois modalités. En dehors des facteurs lignes et colonnes, l'expérience serait alors une expérience factorielle 4 × 3 en carré latin et parcelles divisées, avec quatre répétitions.

3° Expériences factorielles fractionnaires, surfaces de réponse

Les différentes applications du principe des parcelles divisées que nous avons envisagées jusqu'à présent concernent toutes des expériences factorielles complètes. Mais le même principe peut être appliqué aussi aux expériences non factorielles (au sens strict) : expériences factorielles fractionnaires, étude des surfaces de réponse, etc.

En matière d'expérience factorielle fractionnaire, la répartition des objets pourrait être :

221 232 213 312 331 323 133 122 111,

une première répartition au hasard ayant été réalisée tout d'abord pour le premier facteur (premiers chiffres), et une deuxième répartition au hasard étant réalisée ensuite globalement pour les deux autres facteurs (deux derniers chiffres).

De même, en ce qui concerne l'étude des surfaces de réponse, dans le cas du plan de DOEHLERT à deux facteurs, on pourrait envisager de procéder en un premier temps à une répartition aléatoire des trois niveaux du facteur $x2$, puis à une répartition aléatoire, pour chacun de ces niveaux, des deux ou des trois niveaux du facteur $x1$.

Le résultat pourrait être par exemple :

(-1; 0) (1; 0) (0; 0) (0,5; -0.866) (0,5; 0,866) (-0.5; 0,866) (-0,5; -0,866).

De telles situations, essentiellement liées aux expériences organisées de façon consécutive, se présentent principalement dans le domaine industriel.

5.1.3 Les expériences en bandes croisées

1° Blocs aléatoires complets

Aux expériences en blocs aléatoires complets et parcelles divisées, on peut associer les expériences en blocs aléatoires complets et bandes croisées ou blocs divisés 8. Il s'agit d'expériences factorielles qui font généralement intervenir deux facteurs et qui sont conçues de telle sorte qu'à l'intérieur de chacun des blocs, les différents traitements soient appliqués à des bandes de terrain allongées dans une première direction pour un des facteurs et dans une deuxième direction, perpendiculaire à la première, pour l'autre facteur.

La figure 20 donne une illustration de ce type d'expérience. Il s'agit de deux blocs, dans chacun desquels apparaissent, d'une part, huit bandes verticales relativement étroites correspondant aux huit modalités d'un premier facteur (premiers chiffres, 1 à 8, des nombres qui désignent les différents objets), et d'autre part, deux

bandes horizontales plus larges correspondant aux deux modalités d'un deuxième facteur (deuxièmes chiffres, 1 ou 2, des nombres qui désignent les différents objets). Dans une direction comme dans l'autre, la répartition des différentes modalités est réalisée « au hasard » et indépendamment dans chacun des blocs.
Cette façon de procéder permet de disposer de grandes parcelles, qui correspondent aux bandes verticales et horizontales, pour chacun des deux facteurs, et de petites parcelles, qui correspondent aux intersections des deux réseaux de bandes, en ce qui concerne l'interaction des deux facteurs.

2° Carré latin, blocs incomplets, etc.

Le même principe peut être appliqué à d'autres dispositifs expérimentaux, dont le carrélatin et les blocs incomplets. Il s'agit toutefois de situations peu courantes.
De nombreuses variantes des dispositifs expérimentaux en parcelles divisées et en bandes croisées sont également présentées par Federer et King (2007).

5.2 Analyse des résultats

5.2.1 Les expériences en blocs aléatoires complets et parcelles divisées

1° Analyse de la variance

Dans le cas de deux facteurs, auquel nous nous limitons, l'interprétation des résultats des expériences en blocs aléatoires complets et parcelles divisées consiste généralement en une analyse de la variance à trois critères de classification, le plus souvent avec une seule observation par petite parcelle et par variable étudiée.

Les différentes sources de variation d'une telle analyse sont, classiquement, le premier facteur considéré au départ, le deuxième facteur considéré au départ, le facteur blocs, l'interaction des deux facteurs initiaux, et les interactions facteur 1 – blocs, facteur 2 – blocs et facteur 1 – facteur 2 – blocs. Ces trois dernières interactions ne peuvent toutefois pas être regroupées en une seule composante « résiduelle », comme dans le cas général des blocs aléatoires complets, car on doit s'attendre à ce que la première soit plus importante, et parfois même beaucoup plus importante, que les deux autres.

L'interaction facteur 1 – blocs se mesure en effet au sein des blocs, au niveau des grandes parcelles, tandis que les interactions facteur 2 – blocs et facteur 1 – facteur 2 – blocs résultent essentiellement des disparités qui peuvent exister entre les petites parcelles, au sein des grandes parcelles. Or, l'hétérogénéité observée au premier niveau (grandes parcelles) est normalement supérieure à celle observée au deuxième niveau (petites parcelles).

En conséquence, l'interaction facteur 1 – blocs est généralement considérée individuellement et utilisée comme base de comparaison pour tester l'influence du premier facteur. Cette composante est souvent appelée aussi *variation résiduelle* 1 ou *erreur* 1. Par contre, les deux autres interactions sont généralement regroupées en une composante unique, appelée *variation résiduelle* 2 ou *erreur* 2, qui sert de base de comparaison à la fois pour le deuxième facteur et pour l'interaction des deux facteurs.

En outre, on présente souvent séparément l'ensemble des éléments relatifs aux grandes parcelles et l'ensemble des éléments qui concernent les petites parcelles. Le tableau 23 donne le schéma d'une telle analyse, pour p modalités du premier facteur, associé aux grandes parcelles, q modalités du deuxième facteur, associé au petites parcelles, et r blocs.

Tableau 23: Schéma de l'analyse de la variance relative aux expériences en blocs aléatoires complets et parcelles divisées, dans le cas général p× q ×r dans le cas particulier 6 x 3 x 4.

Sources de variation	Degrés de liberté	
Facteur 1	p - 1	5
Blocs	r - 1	3
Variation résiduelle 1	(p - 1) (r - 1)	15
Facteur 2	q - 1	2
Facteur 1 – facteur 2	(p - 1) (q - 1)	10
Variation résiduelle 2	p (q - 1) (r - 1)	36
Totaux	pq r - 1	71

2° Comparaisons de moyennes

Dans les comparaisons de moyennes, au sens des comparaisons particulières ou multiples, il faut être attentif au fait que la manière de calculer les erreurs standards des différences de moyennes dépend du ou des facteurs qui sont pris en considération.

Pour les différences $\bar{x}_i$ — $\bar{x}_i'$ .. , relatives au premier facteur, et pour les différences $\bar{x}_j$ — $\bar{x}_j'$., relatives au deuxième facteur, les erreurs standards sont respectivement:

$$\sqrt{2CM_{(1)}/(q\,r)} \quad \text{et} \quad \sqrt{2CM_{(2)}/(p\,r)}$$

si on désigne par $CM_{(1)}$ et $CM_{(2)}$ les carrés moyens qui concernent les deux sources de variation « résiduelle ». De même, pour les différences $\bar{x}_{ij}$ — $\bar{x}_{ij}'$, relatives également au deuxième facteur, mais en se limitant à une modalité donnée *i* du premier facteur, l'erreur standard est :

$$\sqrt{2CM_{(2)}/(r)}$$

La situation se complique quelque peu quand on considère les différences $\bar{x}_{ij}$ — $\bar{x}_{ij}'$, relatives au premier facteur en se limitant à une modalité donnée *j* du deuxième facteur. L'erreur standard est alors :

$$\sqrt{2\left[CM_{(1)} + (q-1)\ CM_{(2)}\right]}/(qr)$$

Le cas échéant, le nombre de degrés de liberté correspondant peut être déterminé comme suit, de façon approchée, selon la méthode de SATTERTHWAITE :

$$p(p-1)(r-1)\left[CM_{(1)} + (q-1)\ CM_{(2)}\right]\ ^2/\left[pCM_{(1)}^2 + (p-1)(q-1)\ CM_{(2)}^2\right]$$

5.2.2 Autres dispositifs expérimentaux en parcelles divisées

1° Dispositif complètement aléatoire

Des principes semblables peuvent être appliqués au cas des expériences complètement aléatoires organisées en parcelles divisées.

En particulier, pour deux facteurs, l'analyse de la variance est une analyse à deux critères de classification. Le tableau 24 donne le schéma de cette analyse, pour p et q modalités des facteurs et pour r répétitions des p q objets, ainsi que les nombres de degrés de liberté relatifs au premier exemple (alimentation de truies et de porcelets).

2° Autres situations

La méthode qui vient d'être présentée ne s'applique évidemment pas quand on ne dispose que d'une seule répétition, les deux composantes résiduelles disparaissant alors des analyses de la variance. D'autres approches doivent donc être envisagées, et il en est de même pour des effectifs inégaux et pour les expériences non factorielles au sens strict.

Tableau 24: Schéma de l'analyse de la variance relative aux expériences complètement aléatoires organisées en parcelles divisées, dans le cas général p X q X r et dans le cas particulier 2 X 4 X 5.

Sources de variation	Degrés de liberté	
Facteur 1	p - 1	1
Variation résiduelle 1	p (r - 1)	8
Facteur 2	q - 1	3
Facteur 1 – facteur 2	(p - 1) (q - 1)	3
Variation résiduelle 2	p (q - 1) (r - 1)	24
Totaux	pq r - 1	39

5.2.3 Les expériences en bandes croisées

1° Analyse de la variance

Pour les expériences en blocs aléatoires complets et bandes croisées, l'interprétation des résultats consiste normalement en une analyse de la variance à trois critères de classification, dans laquelle n'intervient aucun regroupement a priori des interactions liées au facteur blocs. On doit s'attendre en effet, dans ce cas, à ce que les interactions facteur 1 – blocs et facteur 2 – blocs soient toutes deux plus importantes que l'interaction facteur 1 – facteur 2 – blocs, sans qu'elles ne soient cependant nécessairement du même ordre de grandeur.

Chacune des trois sources de variation auxquelles on s'intéresse principalement, à savoir les deux facteurs initiaux et leur interaction, doit alors être comparée à son interaction avec le facteur blocs. Certains regroupements peuvent néanmoins être envisagés au cas par cas, moyennant la réalisation de tests préliminaires.

Les interactions qui servent de bases de comparaison sont parfois appelées variation résiduelle 1, variation résiduelle 2 et variation résiduelle 3, ou erreur 1, erreur 2 et erreur 3, comme le montre le tableau 25, pour p q objets et r blocs.

Tableau 25: Schéma de l'analyse de la variance relative aux expériences en blocs aléatoires complets et bandes croisées, dans le cas général p ×q ×r et dans le cas particulier 8 ×2 ×2.

Sources de variation	Degrés de liberté	
Facteur 1	$p - 1$	7
Blocs	$r - 1$	1
Variation résiduelle 1	$(p - 1)(r - 1)$	7
Facteur 2	$q - 1$	1
Variation résiduelle 2	$(q - 1)(r - 1)$	1
Facteur 1 – facteur 2	$(p - 1)(q - 1)$	7
Variation résiduelle 3	$(p - 1)(q - 1)(r - 1)$	7
Totaux	$p\,q\,r - 1$	31

5.3. Discussion

1° Avantages et inconvénients

Les expériences en parcelles divisées et en bandes croisées ont l'avantage de permettre l'utilisation de parcelles ou, d'une manière plus générale, d'unités expérimentales de plus grandes dimensions, pour un facteur dans le premier cas et pour les deux facteurs dans le deuxième cas.

Mais il faut savoir que cet avantage est compensé par une perte, parfois importante, de précision ou de puissance pour les comparaisons qui concernent le ou, éventuellement, les facteurs associés aux grandes parcelles. Cette perte de précision ou de puissance résulte non seulement des différences d'ordre de grandeur entre les composantes « résiduelles » qui servent de bases de comparaison, mais aussi des différences qui existent entre les nombres de degrés de liberté de ces composantes. Ce dernier élément intervient de façon d'autant plus marquée que le ou les facteurs associés aux grandes parcelles comportent un petit nombre de modalités.

Par contre, on peut souvent observer un gain de précision ou de puissance pour le facteur qui est associé aux petites parcelles et pour l'interaction des deux facteurs.

2° Applications agronomiques

Dans le domaine agronomique, l'utilisation de parcelles divisées ou de bandes croisées peut donc se justifier en premier lieu quand, pour l'une ou l'autre raison (mécanisation, facilité de réalisation de l'expérience, etc.), un des deux ou les deux facteurs considérés nécessitent des parcelles ou des unités expérimentales de relativement grandes dimensions. Au moment de prendre une décision allant éventuellement dans ce sens, il ne faut cependant pas perdre de vue les réductions de précision ou de puissance auxquelles nous venons de faire allusion.

D'autre part, le recours au principe des parcelles divisées peut aussi se justifier quand on souhaite mettre l'accent plus sur un des facteurs que sur l'autre ou, à la limite, quand on s'intéresse uniquement à un des deux facteurs et à l'interaction des deux facteurs. Tel peut être le cas par exemple si on souhaite comparer différents produits phytosanitaires et étudier leur interaction éventuelle avec un certain nombre de variétés d'une culture donnée, sans s'intéresser particulièrement aux différences pouvant exister entre les variétés.

Enfin, un troisième cas d'application du principe des parcelles divisées concerne l'introduction, en cours d'expérience, d'un facteur supplémentaire non prévu initialement. Une telle situation peut se présenter en particulier dans les expériences de longue durée (expériences forestières notamment), où il est parfois utile ou même nécessaire de subdiviser les parcelles, si leur dimension le permet, de manière à pouvoir tenir compte d'un facteur supplémentaire qui n'avait pas été prévu ou jugé suffisamment important au départ (lutte contre une maladie ou un parasite qui a pris de l'extension en cours d'expérience, par exemple).

3° Parcelles divisées et observations successives

Les observations qui sont répétées à différents moments sur les mêmes individus sont parfois traitées comme s'il s'agissait de parcelles divisées, en assimilant les différents individus observés à de grandes parcelles et les différentes observations successives à de petites parcelles.

Dans cette optique, pour une expérience complètement aléatoire qui ferait intervenir p modalités d'un seul facteur et q observations successives (croissance d'animaux soumis à différentes alimentations et observés à différentes dates, par exemple), l'analyse serait du type de celle du tableau 24, le « facteur 2 » étant en fait le facteur temps ou dates. De même, dans le cas d'une expérience en blocs aléatoires complets, pour p modalités d'un seul facteur, q répétitions successives des mêmes mesures et r blocs, l'analyse de la variance serait celle du tableau 23.

Aucune difficulté particulière ne se présente en ce qui concerne la première partie de telles analyses, c'est-à-dire pour le seul facteur considéré au départ. Il n'en est pas de même par contre pour la deuxième partie de l'analyse, c'est-à-dire pour le facteur temps ou dates et pour l'interaction des deux facteurs.

Les différentes observations successives étant réalisées sur les mêmes individus, il faut en effet s'attendre à ce que les résidus relatifs à chacun des individus ne soient pas indépendants les uns des autres, mais au contraire plus ou moins fortement corrélés entre eux.

5.4 Exemple 1 : Expérience en blocs aléatoires complets et parcelles divisées

5.4.1 Présentation et données

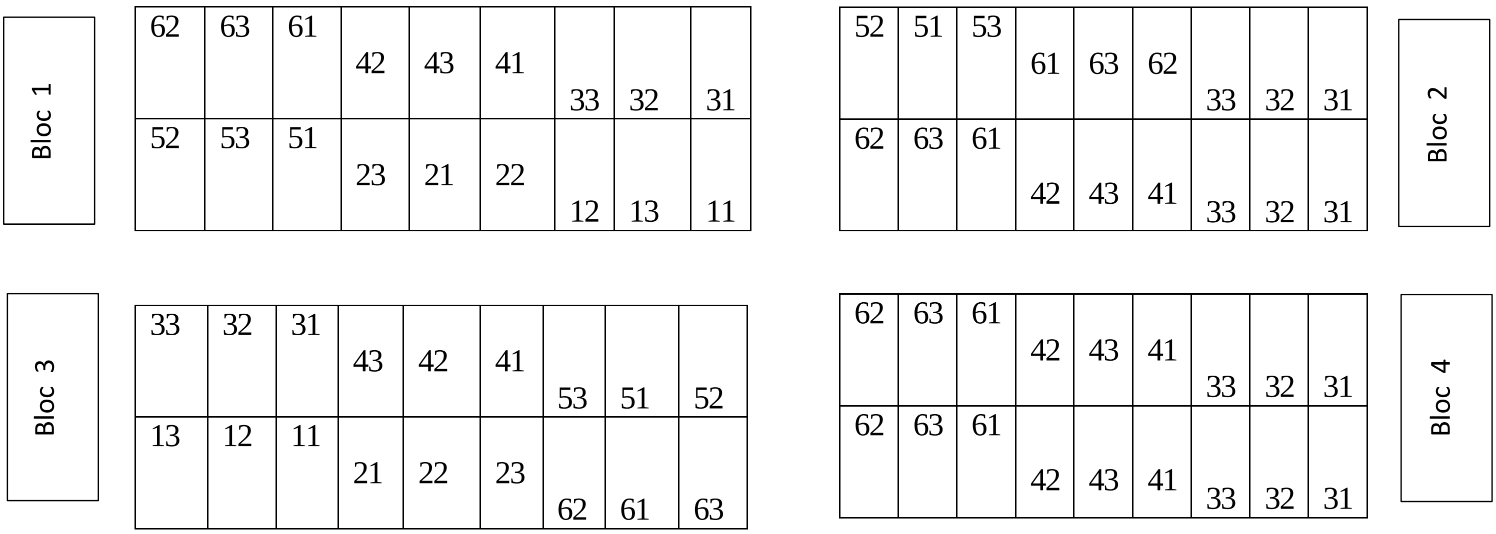
1° Présentation générale

Nous illustrons le principe des blocs aléatoires complets avec parcelles divisées en considérant le cas d'une expérience réalisée sur haricot en vue d'étudier, pour plusieurs variétés, l'influence d'une inoculation bactérienne par Rhizobium.

Les variétés considérées sont au nombre de six et les « traitements » appliqués à ces variétés au nombre de trois, l'ensemble constituant une expérience factorielle complète comportant 18 objets. Les « traitements » sont, dans chaque cas, un témoin, une inoculation et un apport d'azote minéral (50 unités d'azote par ha), cet apport étant destiné à chiffrer autant que possible l'équivalent en azote de l'inoculation bactérienne.

2° Dispositif expérimental

La figure 21 représente, à l'échelle de 1/400, le plan exact de l'expérience. Les parcelles unitaires ont 2 m de largeur sur 6 m de longueur et sont réunies en huit groupes de neuf parcelles, les différents groupes étant séparés et entourés de tous côtés par des chemins d'accès de 2 m de largeur.



Les 72 parcelles ont été divisées en quatre blocs, constitués chacun de deux groupes de neuf parcelles. En l'absence de toute information relative à l'hétérogénéité du terrain, ces blocs ont été définis de manière à être aussi compacts que possible. Ils correspondent aux quatre quadrants de la figure 21.

L'objectif poursuivi en ayant recours au principe des parcelles divisées était, à la fois, de simplifier la réalisation de l'expérience (semis de 24 grandes parcelles au lieu de 72 parcelles unitaires), et de concentrer l'attention sur les différences entre traitements et sur leur interaction avec le facteur variétés, éventuellement aux dépens de la précision des comparaisons entre variétés, considérées comme relativement secondaires.

3° Données

Chacune des 72 petites parcelles comportait cinq lignes de haricots, semées à 40 cm d'écartement. Seules les trois lignes centrales de chaque parcelle ont été l'objet de mesures, de telle sorte que la surface utile par parcelle est égale à 7,2 m^2.

Les rendements obtenus de cette manière et exprimés en kg de gousses fraîches par parcelle, ainsi que les moyennes correspondantes, sont présentés dans le tableau 26.

Tableau 26: Expérience d'inoculation bactérienne sur haricot : rendements observés et moyennes, en kg de gousses fraîches par parcelle.

Variétés et traitements		Blocs 1	2	3	4	Moyennes
1	1	7,20	6,70	5,87	5,35	6,28
1	2	7,38	8,20	7,60	6,08	7,32
1	3	7,82	8,98	7,64	8,20	8,16
2	1	6,95	7,60	7,22	5,60	6,84
2	2	7,55	9,15	8,10	5,30	7,52
2	3	7,42	8,89	8,40	7,37	8,02
3	1	5,37	4,75	3,08	5,30	4,62
3	2	6,95	7,41	4,80	6,00	6,29
3	3	6,00	5,20	4,80	4,60	5,15
4	1	3,87	6,15	4,75	4,50	4,82
4	2	5,35	6,50	4,79	5,50	5,54
4	3	5,35	6,50	6,28	6,10	6,06
5	1	4,50	4,80	3,35	4,11	4,19
5	2	5,14	4,45	4,85	2,95	4,35
5	3	4,50	4,91	3,28	5,02	4,43
6	1	3,10	4,65	2,20	3,50	3,36
6	2	3,00	4,59	2,46	4,30	3,59
6	3	2,95	4,90	1,75	4,48	3,52

5.4.2 Analyse des résultats

1° Examen préliminaire

La figure 18 présente les données sous forme graphique. Dans cette figure, comme dans le tableau 26, on observe tout d'abord que les données sont extrêmement variables (valeur minimum égale à 1,75 et valeur maximum égale à 9,15, soit un rapport de plus de 1 à 5, avec une moyenne générale égale à 5,56 kg par parcelle). On observe aussi que les différences concernent essentiellement les variétés et, à première vue, relativement peu les traitements.

En raison notamment de l'existence de ces différences importantes, les rendements ont été soumis à une transformation logarithmique (logarithmes décimaux), destinée à stabiliser les variances.

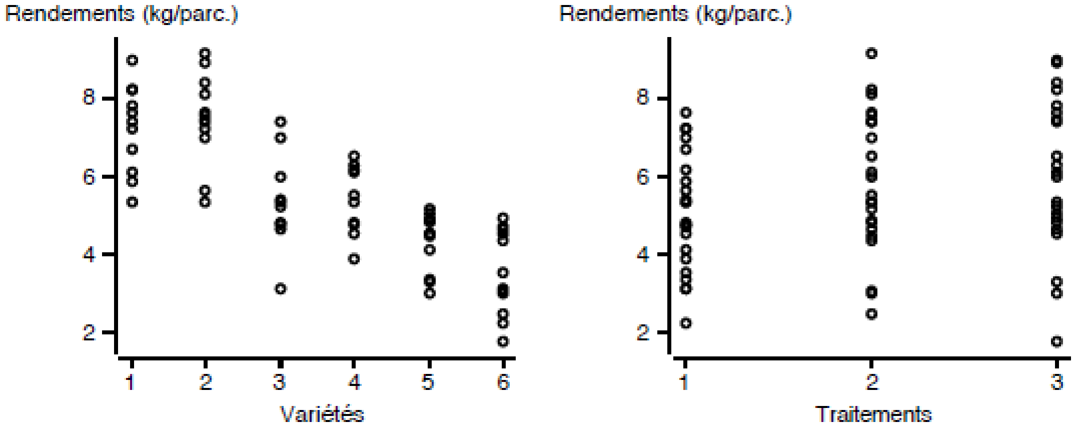


Figure 18: Expérience d'inoculation bactérienne sur haricot : représentation graphique de l'influence des deux facteurs considérés.

2° Analyse de la variance

Le tableau 27 donne les résultats de l'analyse de la variance, réalisée selon le schéma du tableau 23. L'analyse détaillée à trois critères de classification montrerait que la ligne « Variation résiduelle 1 » du tableau 27 correspond bien à l'interaction variétés-blocs, et que le regroupement des interactions traitements-blocs et variétés-traitements-blocs, qui constitue la « Variation résiduelle 2 », se justifie pleinement. Les carrés moyens de ces deux interactions sont en effet égaux respectivement à 0,003854 et 0,002951, avec 6 et 30 degrés de liberté.

Le tableau 27 montre qu'il existe des différences très hautement significatives entre variétés et entre traitements, et que l'interaction des deux facteurs n'est pas significative.

Tableau 27: Expérience d'inoculation bactérienne sur haricot : tableau d'analyse de la variance.

Sources de variation	Degrés de liberté des écarts	Sommes des carrés	Carrés moyens	*F*	P
Variétés	5	1,055289	0,211058	15,0 ***	0,0000
Blocs	3	0,147833	0,049278		
Variation résiduelle 1	15	0,211102	0,014073		
Traitements	2	0,056757	0,028378	9,15 ***	0,0006
Variétés-traitements	10	0,044062	0,004406	1,42	0,21
Variation résiduelle 2	36	0,111646	0,003101		
Totaux	71	1,626689			

Ces conclusions peuvent paraître surprenantes, par comparaison avec ce que semblent indiquer les diagrammes de la figure 22. Elles résultent du fait que les tests relatifs au facteur traitements et à l'interaction variétés-traitements sont beaucoup plus puissants que le test relatif au facteur variétés, en raison, d'une part, de la très grande différence entre les carrés moyens des deux composantes résiduelles (rapport de 1 à 4,5 environ), et d'autre part, de la différence entre les nombres de degrés de liberté qui sont associés à ces deux composantes (respectivement 15 et 36).

3° Différences de moyennes

Les différences entre variétés ne constituent pas, dans le cas présent, un des éléments essentiels de l'expérimentation. Par contre, les différences entre traitements doivent être prises en considération de manière plus précise.

Les moyennes des 24 observations relatives à chacun des traitements sont, après transformation logarithmique :

Témoin: 0,68144,
Inoculation : 0,73875,
Fumure azotée: 0,74301.

On peut en déduire les intervalles de confiance suivants, d'une part pour l'influence de l'inoculation :

$$0,73875 - 0,68144 \pm 2,028 \sqrt{2 (0,003101)/24} = 0,05731 \pm 0,03260$$
$$= 0,02471 \text{ et } 0,08991,$$

et d’autre part pour l’influence de la fumure azotée :

$$0{,}74301 - 0{,}68144 \pm 2{,}028 \sqrt{2\,(0{,}003101)/24} = 0{,}06157 \pm 0{,}03260$$
$$= 0{,}02897 \text{ et } 0{,}09417.$$

La valeur 0,05731 est en fait une estimation de la différence des logarithmes des rendements moyens et, donc aussi, du logarithme du quotient des rendements moyens. Cette différence indique que les rendements obtenus après inoculation sont égaux, dans l’ensemble, à 114 % des rendements des parcelles témoins, puisque :

$$100{,}05731 = 1{,}141 \text{ ou } {}_{\log 10}\, 1{,}141 = 0{,}0573.$$

De même, aux valeurs 0,02471 et 0,08991, correspondent des limites égales à 106 et 123 %.
Ces différents résultats montrent que l’accroissement de rendement lié à l’inoculation peut être estimé à 14 %, avec des limites de confiance égales à 6 et 23 %.
Quant à l’accroissement de rendement lié à la fumure azotée, il peut être estime d’une manière identique à 15 %, avec des limites de confiance égales à 7 et 24 %.
L’accroissement de rendement dû à l’inoculation étant égal à environ 93 % de l’accroissement lié à l’apport de 50 unités d’azote par ha, on peut en conclure que l’inoculation bactérienne équivaut dans ce cas à l’apport de 46 ou 47 unités d’azote.

4° Efficacité relative

Le tableau 27 montre aussi que, pour une répartition supposée complètement aléatoire des 18 objets au sein des quatre blocs, c’est-à-dire en négligeant la structure propre aux parcelles divisées, la variation « résiduelle » relative aux deux facteurs étudiés et à leur interaction aurait été :

$$CM_r = (0{,}211102 + 0{,}111646)/(15 + 36) = 0{,}006328.$$

On peut en déduire qu’en première approximation, par rapport à une expérience classique en blocs aléatoires complets, l’efficacité du test relatif aux variétés est, dans le cas des parcelles divisées :

$$CMr/CM(1) = 0{,}006328/0{,}014073 = 0{,}45 \text{ ou } 45\ \%.$$

On obtient de la même façon, pour les tests relatifs aux traitements et à l’interaction des deux facteurs :

$$CMr/CM(2) = 0{,}006328/0{,}003101 = 2{,}04 \text{ ou } 204\ \%.$$

On peut en conclure que le quasi-doublement de l’efficacité des tests relatifs aux traitements et à l’interaction, qui constituent les principaux objectifs de l’exépérience, est contre balancé par une perte de plus de 50 % de l’efficacité du test relatif aux variétés.
En raison de la réduction du nombre de degrés de liberté, la perte d’efficacité est sans doute sensiblement plus importante que 55 %, et elle serait plus importante encore si le nombre de variétés considérées avait été plus réduit.

5° À propos de la transformation logarithmique

L’emploi de la transformation logarithmique soulève quelques questions complémentaires, qui méritent d’être discutées ici et qui s’appliquent également à d’autres exemples présentés plus loin.
En ce qui concerne les moyennes, il faut se souvenir du fait que le retour à la variable initiale, à partir des moyennes des logarithmes, donne naissance à des moyennes géométriques, qui sont toujours inférieures ou égales aux moyennes arithmétiques correspondantes. A titre indicatif, les moyennes relatives aux trois traitements, obtenues par l’intermédiaire des logarithmes, sont :

$$100{,}68144 = 4{,}80,\ 100{,}73875 = 5{,}48 \text{ et } 100{,}74301 = 5{,}53 \text{ kg/parcelle,}$$

alors que les moyennes arithmétiques, déduites directement des données du tableau 26, sont :

$$5{,}02,\ 5{,}77 \text{ et } 5{,}89 \text{ kg/parcelle.}$$

On peut noter aussi que, globalement, les conclusions auxquelles on aboutit dans le cas présent, avec ou sans transformation de variable, sont fort semblables. Mais l’emploi de la transformation logarithmique, qui complique quelque peu les calculs, présente cependant deux avantages.
D’une part, cette transformation permet d’exprimer toutes les différences entre objets en valeur relative, c’est-à-dire indépendamment notamment des unités de mesure utilisées, de l’étendue des parcelles, etc. Cet avantage n’est pas propre au cas présenté ici, mais est, au contraire, tout à fait général.

D'autre part, l'emploi de la transformation logarithmique rend non significative l'interaction variétés-traitements, qui sans transformation, serait juste significative ($F_{obs} = 2,36$ et $P = 0,029$). Cette absence d'interaction, en termes de logarithmes, signifie que les effets de l'inoculation et de la fumure azotée, exprimés en valeur relative, ne dépendent pas de la variété considérée. Les pourcentages d'augmentation de rendement qui ont été calculés ci-dessus globalement, pour l'ensemble des variétés, peuvent donc être appliqués indifféremment à chacune d'entre elles.
En outre, en ce qui concerne la variabilité« résiduelle », la relation existant entre le coefficient de variation d'une variable aléatoire et l'écart-type de son logarithme permet de calculer aisément des estimations des deux coefficients de variation résiduelle relatifs aux observations initiales. On obtient pour les grandes parcelles (variétés) :

/

0,014073/0,4343 = 0,273 ou 27,3 %,

et pour les petites parcelles (traitements et interaction des deux facteurs) :

6° Codification du modèle d'analyse de la variance

Le contenu du tableau 27 peut être obtenu à l'aide du code suivant, associé aux commandes ou procédures « anova » ou « glm » :

L10R = V B V*B T V*T,

et cela tant avec le logiciel Minitab qu'avec le logiciel SAS. Dans ce modèle, L10R désigne les logarithmes décimaux des rendements, B, T et V les facteurs blocs, traitements et variétés, V*B l'interaction variétés-blocs, qui correspond à la « variation résiduelle 1 », et V*T l'interaction variétés-traitements.
On peut remarquer que la « variation résiduelle 2 » n'apparaît pas dans le code, cette source de variation étant en fait considérée comme un résidu obtenu par différence. En outre, il y a lieu de préciser chaque fois comment les tests doivent être réalisés, ce qui peut être fait à l'aide d'une instruction « random » pour Minitab et « test » pour SAS.
Quant au logiciel R, la formulation peut être:

L10R ← V*T + Error(B + B:V),

en relation avec la fonction « aov ». L'écriture V*T désigne ici à la fois les facteurs V et T et leur interaction, tandis que B:V représente la seule interaction de B et V, le terme « Error » étant relatif à la procédure de réalisation des tests.
Il faut noter aussi que, pour les logiciels R et SAS, les symboles B, T et V doivent être définis comme étant des facteurs, et non pas des variables (respectivement par des instructions « factor » et « class »).
De plus, pour les trois logiciels, d'autres commandes, procédures ou fonctions, liées notamment au modèle linéaire mixte, pourraient aussi être utilisées.
Ces dernières remarques, de même que les remarques relatives à la réalisation des tests, s'appliquent tout autant à d'autres exemples présentés plus loin.

Références bibliographiques

- Ayache antoine; Hamonier Julien. 2019. Cours de statistique descriptive. Université de Bordeaux. 39p.
- Azouzi Belel. 2006. L'outil statistique en expérimentation. Ed OPU, 163P.
- Belazreg salah. 2021. PASS biostatistiques, probabilités, mathématiques. Ed Ediscience.320P.
- Bertrand Frédéric , Maumy-Bertrand Myriam. 2011. Statistique en 80 fiches pour les scientifiques. Ed Dunod, Université de Strasbourg. Institut de Recherche Mathématique Avancée,220P.
- Brakel J., Manil P. 1965. La fifixation symbiotique de l'azote chez le haricot (Phaseolus vulgaris L.) : essais de bact´erisation par Rhizobium phaseoli. Bull. Inst. Agron. Stat. Rech. Gembloux 33 (1), 3-25pp.
- Chekroun abdennasser. 2018. Statistiques descritives et exercices. Rappels de cours et exercices corrigés sur la statistique descriptive. Polycopié de cours. Université Aboubekr Belkaid. Tlemcen. 89P.
- Dagnelie P. [1969-1970]. Th´eorie et m´ethodes statistiques : applications agronomiques (2 vol.). Gembloux, Presses agronomiques de Gembloux, 378 + 451 p.
- Dagnelie P. [1975]. Analyse statistique `a plusieurs variables. Gembloux, Presses agronomiques de Gembloux, 362 p.
- Dagnelie P. [1981]. Théorie et m´ethodes statistiques : exercices (en collaboration avec J.J. Claustriaux et C. Debouche). Gembloux, Presses agronomiques de Gembloux, 186 p.
- Dagnilier pierre. 2012. Principes d'expérimentation planification des expériences et analyse de leurs résultats. Ed Les presses agronomiques de gembloux, Belgique.
- Fisher R.A., Yates F. 1982. Statistical tables for biological, agricultural and medical research. Harlow (UK), Longman, 146 p.
- ognès xavier ; Garenne andré ; Bouteiller xavier; Fiévet virgil. 2018. Le cours de biostatistique. Ed Dunod, France. 358P.
- Oubejja mohamed, 2020. La moyenne arithmétique, 4P. Enseignant des sciences économiques et de gestion.
- Pearson E.S., Hartley H.O. 1972. Biometrika tables for statisticians (2 vol.). Cambridge (UK), Cambridge University Press, 264 + 385 p.
- Tillé yves. 2010. Résumédu cours de statistique descriptive. 196P.
- Rakotomalala ricco.2013. Comparaison de populations. Tests paramétriques. Université Lumière Lyon 2. 109P.