République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université 8 mai 1945 Guelma

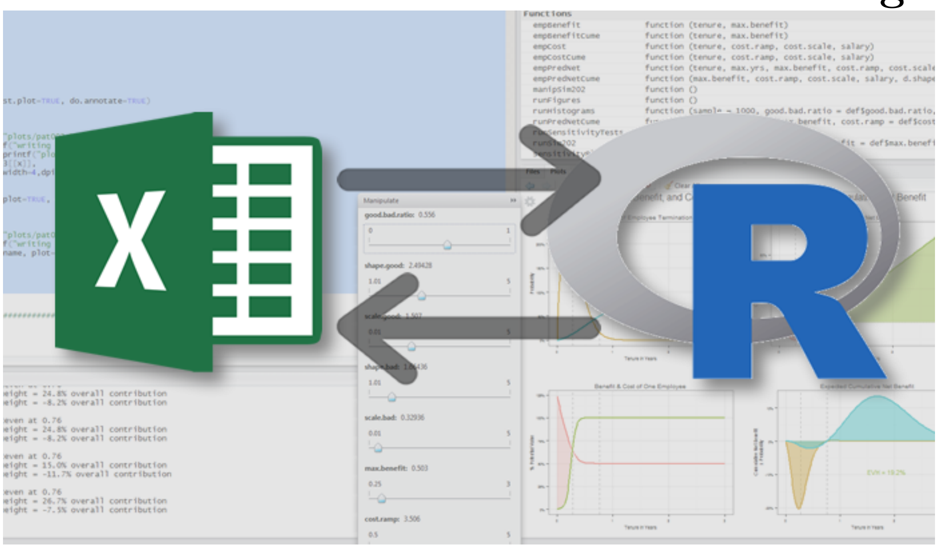
Sciences de la Nature et de la Vie et Science de la Terre de l'Univers

Département de la Biologie

Polycopié de Cours

Bio-statistique réaliser avec Le logiciel R / Excel

1ère Année Master - Pharmacotoxicologie



Enseignant : Dr. SEGNI Sami

Année universitaire : 2025-2026

TABLE DES MATIÈRES

INTRODUCTION GÉNÉRALE

Les biostatistiques occupent une place centrale dans la formation méthodologique des étudiants en Pharmacotoxicologie. Elles constituent un ensemble d'outils scientifiques destinés à recueillir, organiser, analyser et interpréter des données biologiques, expérimentales et pharmacologiques. Dans un contexte où la recherche biomédicale produit des volumes dinformations de plus en plus importants, la maîtrise des méthodes statistiques devient indispensable pour conduire des études fiables, objectiver les résultats, et prendre des décisions éclairées dans la pratique professionnelle comme dans la recherche académique.

L'objectif principal de ce polycopié est d'offrir aux étudiants de Master 1 un support clair, méthodique et complet permettant de comprendre les bases théoriques des statistiques, tout en privilégiant l'aspect pratique à travers l'utilisation des logiciels R et Excel. Chaque chapitre est conçu de manière progressive : définitions fondamentales, exemples illustratifs, exercices corrigés et applications numériques. Ce document constitue ainsi un pont entre la théorie statistique et les exigences concrètes de la pharmacologie expérimentale, de la toxicologie et des sciences biomédicales en général.

Grands axes du polycopié

Ce polycopié est structuré autour de plusieurs axes majeurs, correspondant chacun à un chapitre du cours. Ils couvrent l'ensemble des notions essentielles que tout étudiant en Pharmacotoxicologie doit maîtriser.

1. **Statistique descriptive** Ce premier axe introduit les concepts fondamentaux de la statistique : types de variables, échelles de mesure, représentations graphiques, indicateurs de tendance centrale (moyenne, médiane, mode) et de dispersion (variance, écart-type). Ces outils constituent la base indispensable pour explorer et résumer un jeu de données.
2. **Probabilités et lois de probabilité** Dans ce chapitre, les notions de probabilités sont formalisées : événements, indépendance, conditionnement. Les principales lois discrètes (binomiale, Poisson) et continues (loi normale) sont étudiées, avec une attention particulière portée à leurs applications biomédicales.
3. **Estimation et intervalles de confiance** Une grande partie de l'analyse statistique repose sur l'estimation de paramètres inconnus à partir d'un échantillon. Sont présentées ici les notions d'estimateurs ponctuels, d'intervalle de confiance et de précision des estimations. De nombreux exemples montrent comment interpréter correctement ces résultats dans un cadre expérimental.

4. **Tests d'hypothèses** Ce chapitre expose la logique des tests statistiques : formulation des hypothèses, risques d'erreur, p-valeur, puissance d'un test. Différents tests usuels sont illustrés (test t de Student, test du Khi-deux, tests non paramétriques). L'accent est mis sur la compréhension et la bonne interprétation des conclusions.
5. **Analyse de la variance (ANOVA)** L'ANOVA permet de comparer plusieurs groupes simultanément, ce qui en fait un outil incontournable dans les études expérimentales. Le chapitre détaille l'ANOVA à un facteur, l'ANOVA à deux facteurs et les tests post-hoc tels que Tukey. Des applications sous R permettent d'analyser des données réelles utilisées en pharmacotoxicologie.
6. **Régression linéaire et corrélation** Ici sont abordées les relations entre variables : corrélation, modèles de régression simple et multiple, interprétation des coefficients, validité du modèle. Ces méthodes sont essentielles dans l'analyse de données pharmacologiques (relations doseréponse, effets combinés, prédiction de paramètres biologiques).
7. **Méthodes avancées et notions complémentaires** Ce dernier axe introduit des outils complémentaires : tests non paramétriques, estimation de densité, transformations de données, méthodes de bootstrap, ainsi que quelques notions utiles en toxicologie analytique.

En complément de ces chapitres, une **annexe de tableaux statistiques** est fournie : loi normale centrée réduite, test t de Student, test du Khi-deux, loi de Fisher-Snedecor, ainsi que des tableaux pour les lois binomiale et de Poisson. Ces ressources permettent un accès rapide aux valeurs critiques nécessaires aux calculs statistiques.

Conclusion générale

Ce polycopié a pour vocation de guider progressivement l'étudiant vers une compréhension solide et opérationnelle des méthodes statistiques. Il met l'accent sur l'interprétation, la rigueur méthodologique et la capacité à appliquer les concepts dans le cadre d'expériences biomédicales et pharmacotoxicologiques. En alliant théorie, exemples concrets, exercices et applications logicielles, il constitue un outil complet et adapté aux besoins actuels de la recherche et de la pratique scientifique.

CHAPITRE 1

VARIABLES ALÉATOIRES, ESTIMATEURS STATISTIQUES ET LOIS DE DISTRIBUTION

Objectifs

Ce chapitre a pour objectifs :
- de définir les variables aléatoires et leurs types ;
- de présenter les lois de probabilité usuelles (binomiale, normale) ;
- d'introduire les principaux estimateurs statistiques (moyenne, variance, écart-type) ;
- de décrire les propriétés fondamentales des estimateurs.

1.1 Variables aléatoires

Une variable aléatoire est une fonction qui associe une valeur réelle à chaque résultat possible d'une expérience aléatoire.

Types de variables aléatoires

- **Variable discrète** : prend un nombre fini ou dénombrable de valeurs.
  Exemple : X = nombre de patients infectés.
- **Variable continue** : prend une infinité de valeurs dans un intervalle.
  Exemple : Y = concentration plasmatique d'un médicament.

1.2 Lois de distribution

1.2.1 Loi Binomiale $\mathcal{B}(n, p)$

- Modèle : n essais de Bernoulli avec probabilité de succès p.
- Fonction de probabilité :

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

— Espérance : $\mathbb{E}(X) = np$
— Variance : $\mathrm{Var}(X) = np(1-p)$
Exemple : Si $X \sim \mathcal{B}(10, 0.9)$, alors :

$$P(X = 9) = \binom{10}{9}(0.9)^9(0.1)^1 \approx 0.387$$

1.2.2 Loi Normale $\mathcal{N}(\mu, \sigma^2)$

— Densité :

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

— Espérance : μ ; Variance : σ^2
Exemple : $X \sim \mathcal{N}(300, 20^2)$, alors :

$$P(280 < X < 320) \approx 68\% \quad \text{(règle empirique)}$$

1.3 Estimateurs Statistiques

1.3.1 Moyenne empirique

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

1.3.2 Médiane

Valeur centrale d'une série ordonnée.

1.3.3 Variance corrigée

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

1.3.4 Écart-type

$$S = \sqrt{S^2}$$

Exemple : Soit $X = \{85, 90, 88, 92, 95\}$
— $\bar{X} = 90$
— $S^2 = 14.5$
— $S = \sqrt{14.5} \approx 3.81$

1.4 Propriétés des Estimateurs

— **Non biaisé** : $\mathbb{E}[\hat{\theta}] = \theta$
— **Consistant** : $\hat{\theta}_n \xrightarrow[n\to\infty]{P} \theta$
— **Efficace** : variance minimale parmi les estimateurs non biaisés

Exemple : Estimation de la proportion de patients positifs :

$$\hat{p} = \frac{25}{100} = 0.25, \quad \mathrm{Var}(\hat{p}) = \frac{0.25 \times 0.75}{100} = 0.001875$$

1.5 Applications

1.5.1 En R

```
echantillon <- c(85, 90, 88, 92, 95)
mean(echantillon)
median(echantillon)
var(echantillon)
sd(echantillon)
```

1.5.2 En Excel

— Moyenne : `=MOYENNE(A1:A5)`
— Médiane : `=MEDIANE(A1:A5)`
— Variance : `=VAR.S(A1:A5)`
— Écart-type : `=ECARTYPE.S(A1:A5)`

1.6 À retenir

— Variable aléatoire = modèle d'un phénomène aléatoire.
— Loi binomiale = nombre de succès dans n essais.
— Loi normale = distribution continue en cloche.
— Moyenne, variance, écart-type : outils de synthèse des données.
— Estimateur = outil pour estimer un paramètre inconnu.

Exercice 1 Identifier les types de variables

Un chercheur collecte les informations suivantes sur des patients :
— Âge (en années)
— Sexe (Homme/Femme)
— Groupe sanguin (A, B, AB, O)
— Température corporelle (řC)
— Présence d'hypertension (Oui/Non)

Questions :

1. Identifier le type de chaque variable.

Solution :

Variable	**Type**	**Nature**
Âge	Quantitative	Discrète (entiers)
Sexe	Qualitative	Nominale
Groupe sanguin	Qualitative	Nominale
Température	Quantitative	Continue
Hypertension	Qualitative	Binaire

Application en R :

```
age <- c(23, 45, 36, 29)
sexe <- c("Homme", "Femme", "Femme", "Homme")
groupe <- c("A", "O", "B", "AB")
temperature <- c(36.7, 38.2, 37.0, 36.5)
hypertension <- c("Oui", "Non", "Non", "Oui")

str(data.frame(age, sexe, groupe, temperature, hypertension))
```

Application en Excel : Créer un tableau avec colonnes, utiliser les filtres et validation des données.

Exercice 2 Tableau de fréquences

Données de poids (kg) : `60, 65, 65, 70, 75, 60, 80, 85, 90, 75`
Questions :

1. Construire le tableau des fréquences.
2. Calculer les fréquences relatives et cumulées.

Solution :

Poids	Fréquence	Fréquence relative	Cumulée
60	2	0.2	0.2
65	2	0.2	0.4
70	1	0.1	0.5
75	2	0.2	0.7
80	1	0.1	0.8
85	1	0.1	0.9
90	1	0.1	1.0

Application en R :

```
poids <- c(60, 65, 65, 70, 75, 60, 80, 85, 90, 75)
freq_abs <- table(poids)
freq_rel <- prop.table(freq_abs)
freq_cum <- cumsum(freq_rel)

data.frame(Freq_Abs = freq_abs, Freq_Rel = round(freq_rel, 2), Freq_Cum
   = round(freq_cum, 2))
```

Application en Excel : utiliser `NB.SI`, `=cellule/n`, puis `=SOMME(...)` pour cumuler.

Exercice 3 Histogramme et boxplot

Âges : `20, 22, 25, 30, 28, 32, 40, 45, 35, 38`

Questions :

1. Tracer L'histogramme.
2. Faire le boxplot.
3. Interpréter la distribution.

Application en R :

```
ages <- c(20, 22, 25, 30, 28, 32, 40, 45, 35, 38)

hist(ages, col = "skyblue", main = "Histogramme des âges", xlab = "Âge")
boxplot(ages, main = "Boxplot des âges", col = "orange")
```

Application en Excel : Insertion → Histogramme, puis Boîte à moustaches.

Exercice 4 Moyenne, médiane, variance

Scores : `12, 14, 10, 18, 20, 16, 12, 14, 15, 17`

Calculs :

— Moyenne : $\bar{x} = 14.8$
— Médiane : $(14 + 15)/2 = 14.5$
— Variance : $\sigma^2 = 8.16$

Application en R :

```
scores <- c(12, 14, 10, 18, 20, 16, 12, 14, 15, 17)

mean(scores)
median(scores)
var(scores)
sd(scores)
```

Application en Excel :

— Moyenne : `=MOYENNE()`
— Médiane : `=MEDIANE()`
— Variance : `=VAR.P()` ou `VAR.S()`

Exercice 5 Tri à plat d'une base de données

ID	Sexe	Âge	Tabac	Cholestérol
1	H	45	Oui	2.3
2	F	34	Non	1.9
3	F	39	Oui	2.0
4	H	50	Non	2.7
5	H	41	Oui	2.4

Questions :

- Tri à plat : Sexe et Tabac.
- Moyennes d'âge et cholestérol selon le sexe.

Application en R :

```r
df <- data.frame(
  ID = 1:5,
  Sexe = c("H", "F", "F", "H", "H"),
  Age = c(45, 34, 39, 50, 41),
  Tabac = c("Oui", "Non", "Oui", "Non", "Oui"),
  Cholesterol = c(2.3, 1.9, 2.0, 2.7, 2.4)
)

table(df$Sexe)
table(df$Tabac)

aggregate(Age ~ Sexe, data = df, mean)
aggregate(Cholesterol ~ Sexe, data = df, mean)
```

Application en Excel :

- Tableau croisé dynamique pour comptage (Sexe, Tabac)
- Moyennes avec `=MOYENNE.SI()`

CHAPITRE 2

ORGANISATION ET REPRÉSENTATION DES DONNÉES

2.1 Introduction

L'analyse statistique commence toujours par une phase de description des données. Cette étape vise à organiser l'information brute pour en faciliter l'interprétation. Deux outils principaux sont utilisés :

— La présentation tabulaire

— Les représentations graphiques

Ces outils permettent de résumer visuellement la distribution des données, d'identifier des tendances générales, des valeurs atypiques ou encore des relations entre variables.

2.2 Présentation tabulaire des données

2.2.1 Tableaux de fréquences

Soit une variable X observée sur n individus. On peut construire un tableau de fréquences comme suit :

Valeur	Fréquence absolue (n_i)	Fréquence relative (f_i)	Fréquence cumulée
x_1	n_1	$f_1 = \frac{n_1}{n}$	$F_1 = f_1$
x_2	n_2	$f_2 = \frac{n_2}{n}$	$F_2 = f_1 + f_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_k	$f_k = \frac{n_k}{n}$	$F_k = \sum_{i=1}^{k} f_i$

Exemple :

Pour les âges suivants : 20, 25, 25, 30, 30, 30, 35, 40

— Valeurs : 20, 25, 30, 35, 40

— Fréquences absolues : 1, 2, 3, 1, 1

— Fréquences relatives : 0.125, 0.25, 0.375, 0.125, 0.125

2.2.2 Tableaux croisés (ou de contingence)

Ils permettent d'étudier la relation entre deux variables qualitatives. Par exemple :

	Fumeur	Non-fumeur
Homme	20	15
Femme	10	25

On peut ensuite calculer les fréquences marginales, conditionnelles ou totales.

2.3 Représentations graphiques

2.3.1 Diagramme en bâtons

Utilisé pour les variables qualitatives ou discrètes. Chaque valeur est représentée par une barre dont la hauteur est proportionnelle à sa fréquence.

2.3.2 Histogramme

Utilisé pour les variables quantitatives continues, avec regroupement en classes. Les aires des barres représentent les fréquences.

2.3.3 Diagramme circulaire

Partage un cercle en secteurs proportionnels aux fréquences relatives. Convient aux variables qualitatives à peu de modalités.

2.3.4 Boxplot (boîte à moustaches)

Représente la médiane, les quartiles et les valeurs extrêmes. Il permet de visualiser la dispersion et les valeurs aberrantes.

2.3.5 Nuage de points (scatter plot)

Utilisé pour représenter graphiquement deux variables quantitatives et étudier leur relation.

2.4 Réalisation avec R

Exemple : Histogramme et Boxplot

```
donnees <- c(20, 25, 25, 30, 30, 30, 35, 40)

hist(donnees, main = "Histogramme des âges", col = "lightblue", xlab = "
   Âge")
boxplot(donnees, main = "Boxplot des âges", col = "orange")
```

Exemple : Diagramme circulaire

```
groupes <- c("A", "B", "A", "O", "AB", "B", "A")
table_groupes <- table(groupes)
pie(table_groupes, main = "Répartition des groupes sanguins")
```

2.5 Réalisation avec Excel

- **Histogramme** : Insertion → Graphiques → Histogramme
- **Diagramme circulaire** : Insertion → Camembert
- **Boxplot (boîte à moustaches)** : possible à partir d'Excel 2016+
- **Nuage de points** : Insertion → Graphique → Nuage
- **Tableaux croisés dynamiques** : très utiles pour croiser deux variables

2.6 Interprétation

Une bonne représentation doit permettre de :

- Synthétiser les informations importantes
- Repérer les valeurs extrêmes
- Identifier des tendances ou associations

L'analyse descriptive ne permet pas de conclure, mais prépare à l'analyse statistique inférentielle.

2.7 Exercice

2.7.1 Exercice 1 Tableau de fréquences

On relève les âges suivants chez 12 patients :

```
23, 25, 25, 27, 27, 27, 29, 30, 30, 32, 35, 35
```

Questions :

1. Construire le tableau des fréquences.
2. Interpréter les résultats.

Solution :

Âge	Fréquence	Fréquence relative	Cumulée
23	1	0.083	0.083
25	2	0.167	0.25
27	3	0.25	0.5
29	1	0.083	0.583
30	2	0.167	0.75
32	1	0.083	0.833
35	2	0.167	1.0

Application en R :

```
ages <- c(23, 25, 25, 27, 27, 27, 29, 30, 30, 32, 35, 35)

freq_abs <- table(ages)
freq_rel <- prop.table(freq_abs)
freq_cum <- cumsum(freq_rel)

data.frame(Freq_Abs = freq_abs,
           Freq_Rel = round(freq_rel, 3),
           Freq_Cum = round(freq_cum, 3))
```

Application en Excel : Utiliser les fonctions `NB.SI`, `=Fréquence/Total`, et `=SOMME()` pour les cumuls.

2.7.2 Exercice 2 Diagramme en bâtons

Voici les groupes sanguins de 10 patients :

`A, B, AB, O, A, B, A, O, AB, A`

Questions :

1. Construire un diagramme en bâtons.
2. Identifier le groupe le plus fréquent.

Solution : Le groupe A est le plus fréquent (4 fois).
Application en R :

```
groupes <- c("A", "B", "AB", "O", "A", "B", "A", "O", "AB", "A")

barplot(table(groupes), col = "lightblue", main = "Diagramme des groupes
    sanguins")
```

Application en Excel : Utiliser `NB.SI` pour les fréquences, puis Insertion Graphique à barres.

2.7.3 Exercice 3 Histogramme de glycémie

Valeurs : `0.95, 1.02, 0.88, 1.10, 1.25, 1.05, 0.99, 1.15, 1.30, 1.12, 1.00, 0.92, 1.28, 1.05, 1.08`
Questions :

1. Tracer un histogramme.
2. Commenter la distribution.

Solution : Distribution légèrement asymétrique à droite.
Application en R :

```
glycemie <- c(0.95, 1.02, 0.88, 1.10, 1.25, 1.05, 0.99,
              1.15, 1.30, 1.12, 1.00, 0.92, 1.28, 1.05, 1.08)

hist(glycemie, col = "lightgreen", main = "Histogramme de la glycémie",
   xlab = "Glycémie (g/L)")
```

Application en Excel : Créer classes manuelles (ex. 0.850.95...) puis insérer un histogramme.

2.7.4 Exercice 4 Boxplot des scores

Scores : `12, 15, 14, 18, 20, 21, 16, 15, 17, 19`

Questions :

1. Tracer un boxplot.
2. Identifier médiane et quartiles.

Application en R :

```
scores <- c(12, 15, 14, 18, 20, 21, 16, 15, 17, 19)

boxplot(scores, col = "orange", main = "Boxplot des scores", ylab = "
   Score")
```

Application en Excel : Utiliser Insertion Boîte à moustaches (Excel 2016+).

2.7.5 Exercice 5 Tableau croisé et camembert

Sexe	Tabac
H	Oui
F	Non
F	Oui
H	Non
H	Oui
F	Non
H	Non
F	Oui

Questions :

1. Construire un tableau croisé Sexe Œ Tabac.
2. Faire un camembert des fumeurs par sexe.

Solution :

	Oui	Non
Homme	2	2
Femme	2	1

Application en R :

```
sexe <- c("H", "F", "F", "H", "H", "F", "H", "F")
tabac <- c("Oui", "Non", "Oui", "Non", "Oui", "Non", "Non", "Oui")

tableau <- table(Sexe = sexe, Tabac = tabac)
tableau

fumeurs <- sexe[tabac == "Oui"]
pie(table(fumeurs), main = "Répartition des fumeurs par sexe")
```

Application en Excel : Créer un tableau croisé dynamique puis un graphique camembert.

1. Introduction

Les mesures de tendance centrale et de dispersion permettent de résumer une série de données numériques. Elles sont essentielles pour comprendre la structure d'un jeu de données, en biostatistique comme en pharmacotoxicologie.

2.8 Mesures de tendance centrale

2.8.1 Moyenne arithmétique

La moyenne arithmétique est définie par :

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

Exemple : Soit les valeurs : 10, 12, 14, 15

$$\bar{x} = \frac{10 + 12 + 14 + 15}{4} = 12.75$$

R :

```
x <- c(10, 12, 14, 15)
mean(x)
```

Excel : `=MOYENNE(A1:A4)`

—

2.8.2 Médiane

La médiane est la valeur qui partage l'échantillon en deux parties égales.

— Si n est impair, c'est la valeur du milieu.

— Si n est pair, c'est la moyenne des deux valeurs centrales.

Exemple : 12, 13, 15, 17 $\Rightarrow$ Médiane $= (13 + 15)/2 = 14$

R : `median(x)` **Excel :** `=MEDIANE(A1:A4)`

—

2.8.3 Mode

Le mode est la valeur la plus fréquente.

Exemple : 1, 2, 2, 2, 3, 4, 5 $\Rightarrow$ Mode $= 2$

R :

```
x <- c(1, 2, 2, 2, 3, 4, 5)
table(x)
```

Excel : `=MODE.SNGL(A1:A7)`

2.9 Mesures de dispersion

2.9.1 Étendue

$$\text{Étendue} = x_{\max} - x_{\min}$$

Exemple : Max = 20, Min = 10 $\Rightarrow$ Étendue = 10

2.9.2 Variance

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

Exemple : Pour $x = \{10, 12, 14\}$, $\bar{x} = 12$

$$s^2 = \frac{(10-12)^2 + (12-12)^2 + (14-12)^2}{2} = 4$$

2.9.3 Écart-type

$$s = \sqrt{s^2}$$

L'écart-type mesure la dispersion autour de la moyenne.

2.9.4 Coefficient de variation (CV)

$$CV = \frac{s}{\bar{x}} \times 100\%$$

Permet de comparer la variabilité relative entre plusieurs séries.

2.9.5 Intervalle interquartile (IQR)

$$IQR = Q_3 - Q_1$$

Avec Q_1 le premier quartile et Q_3 le troisième quartile.

2.10 Comparaison des mesures

- La moyenne est sensible aux valeurs extrêmes, contrairement à la médiane.
- L'écart-type est souvent utilisé pour résumer la variabilité dans les études biomédicales.

2.11 Application sur R

```
x <- c(12, 14, 15, 18, 20)

mean(x)         # Moyenne
median(x)       # Médiane
var(x)          # Variance
sd(x)           # Écart-type
quantile(x)     # Quartiles
IQR(x)          # Intervalle interquartile
sd(x)/mean(x) * 100  # Coefficient de variation (%)
```

2.12 Application sur Excel

- Moyenne : `=MOYENNE()`
- Médiane : `=MEDIANE()`
- Variance : `=VAR.S()` ou `=VAR.P()`
- Écart-type : `=ECARTYPE.S()` ou `=ECARTYPE.P()`
- Quartiles : `=QUARTILE.INF()` et `=QUARTILE.SUP()`
- CV : `=ECARTYPE()/MOYENNE()*100`

2.13 Conclusion

Ces indicateurs sont fondamentaux pour toute étude statistique, notamment en pharmacotoxicologie, où il est crucial de connaître la variabilité des mesures (doses, concentrations, réponses biologiques, etc...).

2.14 Exercices

2.14.1 Exercice 1 Moyenne, Médiane, Mode

Scores à un test : `12, 14, 13, 15, 14, 16, 18`

Questions :

1. Calculer la moyenne, la médiane et le mode.
2. Interpréter les résultats.

Solution :

- Moyenne : $\bar{x} = \frac{102}{7} \approx 14.57$
- Médiane = 14 (valeur centrale)
- Mode = 14 (valeur la plus fréquente)

Application R :

```
x <- c(12, 14, 13, 15, 14, 16, 18)
mean(x)
median(x)
table(x)
```

Application Excel :

- Moyenne : `=MOYENNE(A1:A7)`

— Médiane : `=MEDIANE(A1:A7)`
— Mode : `=MODE.SNGL(A1:A7)`

2.14.2 Exercice 2 Variance et écart-type

Poids de 5 échantillons (kg) : `2.5, 2.8, 3.0, 2.7, 2.9`
Solution :
— Moyenne : 2.78
— Variance : $s^2 \approx 0.037$
— Écart-type : $s \approx 0.192$
Application R :

```
x <- c(2.5, 2.8, 3.0, 2.7, 2.9)
var(x)
sd(x)
```

Application Excel :
— Variance : `=VAR.S(A1:A5)`
— Écart-type : `=ECARTYPE.S(A1:A5)`

2.14.3 Exercice 3 Coefficient de variation

Concentration (mg/L) : `15, 18, 17, 16, 14`
Solution :
— Moyenne : 16
— Écart-type : ≈ 1.58
— CV : $1.58/16 \times 100 \approx 9.88\%$
Application R :

```
x <- c(15, 18, 17, 16, 14)
sd(x) / mean(x) * 100
```

Application Excel : `=ECARTYPE.S(A1:A5)/MOYENNE(A1:A5)*100`

2.14.4 Exercice 4 Étendue et IQR

Durées d'effet (h) : `3.2, 3.5, 4.0, 4.3, 4.1, 3.9, 3.8, 4.5`
Solution :
— Étendue : $4.5 - 3.2 = 1.3$
— IQR : $Q_3 - Q_1 = 4.2 - 3.5 = 0.7$
Application R :

```
x <- c(3.2, 3.5, 4.0, 4.3, 4.1, 3.9, 3.8, 4.5)
range(x)
IQR(x)
```

Application Excel :
— Étendue : `=MAX(A1:A8)-MIN(A1:A8)`
— IQR : `=QUARTILE.INF(A1:A8,3)-QUARTILE.INF(A1:A8,1)`

2.14.5 Exercice 5 Synthèse descriptive

Concentrations (ţg/mL) : `102, 105, 98, 100, 110, 115, 108, 103`

Solution :

— Moyenne 105.13
— Médiane = 104
— Écart-type 5.37
— CV 5.1%

Les données sont centrées et peu dispersées.

Application R :

```
x <- c(102, 105, 98, 100, 110, 115, 108, 103)
mean(x)
median(x)
sd(x)
sd(x)/mean(x)*100
```

Application Excel :

— Moyenne : `=MOYENNE(A1:A8)`
— Médiane : `=MEDIANE(A1:A8)`
— Écart-type : `=ECARTYPE.S(A1:A8)`
— CV : `=ECARTYPE.S(A1:A8)/MOYENNE(A1:A8)*100`

CHAPITRE 3

PROBABILITÉ CONDITIONNELLE

3.1 Définitions de base

Une **expérience aléatoire** est une expérience dont le résultat n'est pas prévisible avec certitude.

Exemples :

— Tirer une carte au hasard dans un jeu.

— Réaliser un test de dépistage chez un patient.

Espace probabilisable : Ensemble Ω des issues possibles.

Un **événement** est un sous-ensemble de Ω. Si $A \subset \Omega$, on appelle A un événement.

Probabilité : Une fonction P telle que :

$$P : \mathcal{P}(\Omega) \to [0, 1] \quad \text{avec } P(\Omega) = 1$$

3.2 Règles de calcul des probabilités

3.2.1 Union et intersection

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Si A et B sont **incompatibles**, alors $P(A \cap B) = 0$, donc :

$$P(A \cup B) = P(A) + P(B)$$

3.2.2 Complémentaire

$$P(A^c) = 1 - P(A)$$

3.3 Probabilité conditionnelle et indépendance

3.3.1 Probabilité conditionnelle

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \text{si } P(B) > 0$$

3.3.2 Événements indépendants

A et B sont indépendants si :

$$P(A \cap B) = P(A) \cdot P(B)$$

3.4 Théorème de Bayes

Soit une partition $\{A_1, A_2, ..., A_n\}$ de Ω, alors :

$$P(A_i \mid B) = \frac{P(B \mid A_i) \cdot P(A_i)}{\sum_{j=1}^{n} P(B \mid A_j) \cdot P(A_j)}$$

Application : interprétation d'un test médical (sensibilité/spécificité).

3.5 Application en pharmacotoxicologie

- **Test de dépistage :** Probabilité qu'un individu soit malade sachant que le test est positif.
- **Effet indésirable :** Probabilité qu'un effet secondaire se produise après exposition à une molécule.
- **Études cliniques :** Probabilité d'observer une réponse thérapeutique.

3.6 Application sur R

Exemple : Un médicament a 60% d'efficacité, on traite 3 patients. Quelle est la probabilité que 2 réussites ?

Modélisation par loi binomiale : $n = 3$, $p = 0.6$, $k = 2$

```
dbinom(2, size = 3, prob = 0.6)  # Résultat : 0.432
```

Probabilité cumulative (2 réussites) :

```
pbinom(2, size = 3, prob = 0.6)
```

3.7 Application sur Excel

- `=LOI.BINOMIALE(2;3;0,6;FAUX)` P(X=2)
- `=LOI.BINOMIALE(2;3;0,6;VRAI)` P(X 2)
- `=1 - LOI.BINOMIALE(2;3;0,6;VRAI)` P(X > 2)

3.8 Conclusion

Le calcul des probabilités est essentiel pour interpréter les résultats issus de données biomédicales, notamment dans les contextes de dépistage, de diagnostic et d'études cliniques. Il prépare également l'étude des lois de probabilité vues au prochain chapitre.

3.9 Exercices

3.9.1 Exercice 1 Probabilité simple

Une étude rapporte que 20% des patients développent un effet secondaire après un médicament.

Questions :

1. Quelle est la probabilité qu'un patient ne développe pas d'effet secondaire ?
2. Et que deux patients consécutifs en développent un ?

Solutions :
— $P(A^c) = 1 - 0.2 = 0.8$
— $P(A \cap A) = 0.2^2 = 0.04$
R :

```
p <- 0.2
1 - p
p^2
```

Excel :
— `=1 - 0.2`$\to$ 0.8
— $= 0.2^2 \to 0.04$

3.9.2 Exercice 2 Union et intersection

On donne : $P(A) = 0.3$, $P(B) = 0.5$, $P(A \cap B) = 0.2$
Questions :

1. Calculer $P(A \cup B)$
2. Calculer $P(A^c \cap B)$

Solutions :
— $P(A \cup B) = 0.3 + 0.5 - 0.2 = 0.6$
— $P(B) - P(A \cap B) = 0.5 - 0.2 = 0.3$
R :

```
a <- 0.3; b <- 0.5; ab <- 0.2
a + b - ab
b - ab
```

Excel :
— `=0.3 + 0.5 - 0.2` 0.6
— `=0.5 - 0.2` 0.3

3.9.3 Exercice 3 Théorème de Bayes

Un test médical a une sensibilité de 0.95, une spécificité de 0.90, et la prévalence est de 5%.

Question : Quelle est la probabilité qu'un patient soit malade sachant que le test est positif ?

Solution :

$$P(T^+) = 0.95 \cdot 0.05 + 0.1 \cdot 0.95 = 0.1425$$

$$P(M \mid T^+) = \frac{0.0475}{0.1425} \approx 0.333$$

R :

```
sens <- 0.95; spec <- 0.90; prev <- 0.05
ppv <- (sens * prev) / (sens * prev + (1 - spec) * (1 - prev))
ppv
```

Excel : `=(0.95*0.05)/(0.95*0.05 + 0.1*0.95)` 0.333

3.9.4 Exercice 4 Indépendance

On observe : $P(A) = 0.3$, $P(B) = 0.4$, $P(A \cap B) = 0.12$

Question : A et B sont-ils indépendants ?

Solution :

$$P(A) \cdot P(B) = 0.3 \cdot 0.4 = 0.12 = P(A \cap B)$$

Donc A et B sont indépendants.

R :

```
pA <- 0.3; pB <- 0.4; pAB <- 0.12
pA * pB == pAB
```

Excel : `=0.3*0.4` 0.12

3.9.5 Exercice 5 Loi binomiale

Un médicament a 70% d'efficacité. On le donne à 4 patients. Quelle est la probabilité que 3 répondent ?

Solution :

$$P(X = 3) = \binom{4}{3} \cdot 0.7^3 \cdot 0.3 = 4 \cdot 0.343 \cdot 0.3 \approx 0.4116$$

R :

```
dbinom(3, size = 4, prob = 0.7)
```

Excel : `=LOI.BINOMIALE(3;4;0.7;FAUX)` 0.4116

CHAPITRE 4

LOIS DE PROBABILITÉ USUELLES

4.1 Loi uniforme

Une variable X suit une loi uniforme discrète si toutes les issues sont équiprobables.

$$P(X = x_i) = \frac{1}{n} \quad \text{pour } x_i \in \{x_1, ..., x_n\}$$

Exemple : Une boîte contient 5 comprimés numérotés de 1 à 5, un seul est choisi au hasard.

Espérance : $\mathbb{E}(X) = \frac{n+1}{2}$ **Variance :** $Var(X) = \frac{(n^2-1)}{12}$

R :

```
x <- 1:5
mean(x)
var(x)
```

Excel : '=MOYENNE(A1 :A5)', '=VAR.P(A1 :A5)'

4.2 Loi binomiale

Modélise le nombre de succès dans n essais indépendants, avec probabilité p de succès.

$$X \sim \mathcal{B}(n, p) \quad \text{et} \quad P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Espérance : $\mathbb{E}(X) = np$ **Variance :** $np(1-p)$

Exemple : On administre un vaccin à 10 patients, efficacité de 80%.

```
dbinom(7, size = 10, prob = 0.8)   # P(7 succès)
pbinom(7, size = 10, prob = 0.8)   # P( 7 succès)
```

Excel :

— '=LOI.BINOMIALE(7;10;0.8;FAUX)' P(X=7)

— '=LOI.BINOMIALE(7;10;0.8;VRAI)' P(X 7)

4.3 Loi de Poisson

Modélise le nombre d'événements rares dans un intervalle fixe.

$$X \sim \mathcal{P}(\lambda) \quad \text{et} \quad P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

Espérance et variance : $\mathbb{E}(X) = Var(X) = \lambda$
Exemple : Une réaction allergique survient en moyenne 2 fois par semaine.

```
dpois(3, lambda = 2)  # P(3 cas)
ppois(3, lambda = 2)  # P( 3 cas)
```

Excel :
- '=LOI.POISSON(3 ;2 ;FAUX)' P(X = 3)
- '=LOI.POISSON(3 ;2 ;VRAI)' P(X 3)

4.4 Loi normale

Loi continue en forme de cloche (distribution gaussienne).

$$X \sim \mathcal{N}(\mu, \sigma^2) \quad \text{densité} : f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Propriétés :
- Symétrique autour de μ
- Environ 68% des valeurs dans $[\mu - \sigma, \mu + \sigma]$
- 95% dans $[\mu - 2\sigma, \mu + 2\sigma]$

Exemple : Concentrations plasmatiques $\sim \mathcal{N}(100, 15^2)$

```
pnorm(120, mean = 100, sd = 15)     # P(X  120)
1 - pnorm(120, mean = 100, sd = 15) # P(X > 120)
```

Excel :
- '=LOI.NORMALE(120 ;100 ;15 ;VRAI)' P(X 120)
- '=1 - LOI.NORMALE(120 ;100 ;15 ;VRAI)' P(X > 120)

4.5 Comparaison des lois

- **Binomiale** : succès/échec sur n essais.
- **Poisson** : événements rares (surtout si n grand et p petit).
- **Normale** : phénomène continu, nombreuses observations.

4.6 Applications en pharmacotoxicologie

- Loi binomiale : efficacité d'un médicament, réponse binaire.
- Loi de Poisson : fréquence d'effets indésirables.
- Loi normale : distribution des concentrations, tension artérielle.

4.7 Application sur R et Excel Récapitulatif

R :
— Binomiale : `dbinom()`, `pbinom()`
— Poisson : `dpois()`, `ppois()`
— Normale : `dnorm()`, `pnorm()`
Excel :
— `LOI.BINOMIALE()`
— `LOI.POISSON()`
— `LOI.NORMALE()`

4.8 Conclusion

Les lois de probabilité sont essentielles en biostatistique pour modéliser l'incertitude et la variabilité des phénomènes biologiques. Leur bonne utilisation permet d'évaluer objectivement les effets d'un traitement ou la fréquence d'un événement.

4.9 Exercices

4.9.1 Exercice 1 Loi uniforme discrète

Une infirmière choisit au hasard une seringue parmi 4 lots numérotés de 1 à 4.
Questions :

1. Quelle est la probabilité de tirer le lot 3 ?
2. Quelle est l'espérance et la variance ?

Solution :

$$P(X = 3) = \frac{1}{4} = 0{,}25$$

$$\mathbb{E}(X) = \frac{1+2+3+4}{4} = 2{,}5 \quad ; \quad Var(X) = \frac{4^2 - 1}{12} = 1{,}25$$

R :

```
x <- 1:4
mean(x)
((4^2 - 1)/12)
```

Excel :
— Moyenne : `=MOYENNE(A1:A4)`
— Variance théorique : `=(`4^2` - 1)/12`

4.9.2 Exercice 2 Loi binomiale

Un médicament a une efficacité de 70%. Il est administré à 5 patients.
Questions :

1. Probabilité que 4 patients répondent ?
2. Probabilité que 3 ou moins répondent ?

Solution :

$$X \sim \mathcal{B}(5, 0.7)$$

$$P(X = 4) = \binom{5}{4} \cdot 0.7^4 \cdot 0.3 = 5 \cdot 0.2401 \cdot 0.3 = 0.36015$$

$$P(X \leq 3) = \sum_{k=0}^{3} \binom{5}{k} \cdot 0.7^k \cdot 0.3^{5-k} \approx 0.4718$$

R :

```
dbinom(4, size = 5, prob = 0.7)
pbinom(3, size = 5, prob = 0.7)
```

Excel :

— `=LOI.BINOMIALE(4;5;0.7;FAUX)`
— `=LOI.BINOMIALE(3;5;0.7;VRAI)`

4.9.3 Exercice 3 Loi de Poisson

Une unité hospitalière observe 3 effets indésirables graves par semaine en moyenne.
Questions :

1. Probabilité d'observer exactement 2 cas ?
2. Probabilité d'en observer au plus 3 ?

Solution :

$$X \sim \mathcal{P}(3)$$

$$P(X = 2) = \frac{e^{-3} \cdot 3^2}{2!} \approx 0.2240 \quad ; \quad P(X \leq 3) \approx 0.647$$

R :

```
dpois(2, lambda = 3)
ppois(3, lambda = 3)
```

Excel :

— `=LOI.POISSON(2;3;FAUX)`
— `=LOI.POISSON(3;3;VRAI)`

4.9.4 Exercice 4 Loi normale

La concentration plasmatique d'un médicament suit une loi $\mathcal{N}(100, 15^2)$.
Questions :

1. Probabilité que la concentration soit supérieure à 120 ?
2. Probabilité qu'elle soit entre 90 et 110 ?

Solution :

$$P(X > 120) = 1 - P(X \leq 120) = 1 - 0.9088 = 0.0912$$

$$P(90 \leq X \leq 110) = P(X \leq 110) - P(X \leq 90) = 0.7475 - 0.2525 = 0.495$$

R :

```r
1 - pnorm(120, mean = 100, sd = 15)
pnorm(110, 100, 15) - pnorm(90, 100, 15)
```

Excel :

- `=1 - LOI.NORMALE(120;100;15;VRAI)`
- `=LOI.NORMALE(110;100;15;VRAI) - LOI.NORMALE(90;100;15;VRAI)`

4.9.5 Exercice 5 Approximation binomiale/Poisson

On administre un vaccin à 1000 patients. La probabilité d'effet secondaire grave est 0.1%.

Question : Comparer $P(X = 0)$ sous :

- $X_1 \sim \mathcal{B}(1000, 0.001)$
- $X_2 \sim \mathcal{P}(1)$

Solution :

$$P_{\text{Binomiale}}(X = 0) = (1 - 0.001)^{1000} \approx 0.3677 \quad ; \quad P_{\text{Poisson}}(X = 0) = e^{-1} \approx 0.3679$$

R :

```r
dbinom(0, size = 1000, prob = 0.001)
dpois(0, lambda = 1)
```

Excel :

- =$(1 - 0.001)^1000$
- `=LOI.POISSON(0;1;FAUX)`

CHAPITRE 5

ESTIMATION STATISTIQUE

5.1 Introduction

L'estimation statistique permet d'évaluer, à partir d'un échantillon, une caractéristique inconnue d'une population (moyenne, proportion, variance).

5.2 Types d'estimation

5.2.1 Estimation ponctuelle

Elle fournit une valeur unique comme approximation du paramètre.
Exemples :
- Moyenne observée $\bar{x}$ comme estimation de la moyenne μ
- Proportion $f = \frac{x}{n}$ comme estimation de p

5.2.2 Estimation par intervalle (de confiance)

Elle fournit un intervalle dans lequel le paramètre a une forte probabilité de se trouver (ex. 95%).

5.3 Propriétés d'un bon estimateur

- **Non biaisé :** $\mathbb{E}(\hat{\theta}) = \theta$
- **Convergent :** $\hat{\theta}_n \to \theta$ quand $n \to \infty$
- **Efficace :** variance minimale parmi les estimateurs sans biais

5.4 Estimation de paramètres usuels

5.4.1 Moyenne d'une population

Estimation ponctuelle :

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Intervalle de confiance à 95% (si σ connue) :

$$IC_{95\%} = \bar{x} \pm z_{0.975} \cdot \frac{\sigma}{\sqrt{n}}$$

Si σ inconnue (cas habituel) :

$$IC_{95\%} = \bar{x} \pm t_{n-1,0.975} \cdot \frac{s}{\sqrt{n}}$$

5.4.2 Proportion (cas binaire)

$$f = \frac{x}{n} \quad ; \quad IC_{95\%} = f \pm z_{0.975} \cdot \sqrt{\frac{f(1-f)}{n}}$$

5.4.3 Variance

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

Intervalle de confiance pour la variance :

$$\left[\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{\alpha/2}} \right]$$

5.5 Application en pharmacotoxicologie

- Estimer la moyenne des concentrations plasmatiques d'un médicament
- Estimer la proportion d'effets indésirables observés
- Calculer des intervalles de confiance pour comparer des lots

5.6 Implémentation sur R et Excel

5.6.1 Moyenne et IC sur R

```
x <- c(105, 98, 112, 101, 100)
mean(x)
sd(x)/sqrt(length(x))     # Erreur standard
t.test(x)                 # IC 95%
```

5.6.2 Moyenne et IC sur Excel

- Moyenne : `=MOYENNE(A1:A5)`
- E.T. : `=ECART.TYPE.STANDARD(A1:A5)`
- IC à 95% : `=MOYENNE(A1:A5)` ± `1.96*E.T./RACINE(N)`

5.6.3 Proportion sur R

```
prop.test(x = 30, n = 50, correct = FALSE)
```

5.6.4 Proportion sur Excel

- $f = x/n$: `=30/50`
- IC95% : `=f ± 1.96*RACINE(f*(1-f)/n)`

5.7 Conclusion

L'estimation statistique est un outil central pour déduire des informations sur une population à partir d'un échantillon. Elle permet de résumer les données tout en tenant compte de l'incertitude inhérente à l'échantillonnage.

5.8 Exercices

5.8.1 Exercice 1 Estimation de la moyenne

Un dosage plasmatique d'un médicament a été mesuré chez 5 patients :
`[105, 98, 112, 101, 100]`

Questions :

1. Estimer la moyenne.
2. Calculer l'erreur standard.

Solution :

$$\bar{x} = \frac{105 + 98 + 112 + 101 + 100}{5} = 103{,}2$$

$$s = \sqrt{\frac{1}{4}\sum (x_i - \bar{x})^2} \approx 5{,}63 \quad \Rightarrow \quad ES = \frac{s}{\sqrt{5}} \approx 2{,}52$$

R :

```
x <- c(105, 98, 112, 101, 100)
mean(x)
sd(x)/sqrt(length(x))  # erreur standard
```

Excel :

- Moyenne : `=MOYENNE(A1:A5)`
- ES : `=ECART.TYPE.STANDARD(A1:A5)/RACINE(5)`

5.8.2 Exercice 2 IC à 95% pour la moyenne

On reprend les données de l'exercice 1.

Solution :

$$t_{0.975,4} \approx 2.776 \quad \Rightarrow \quad IC = 103{,}2 \pm 2{,}776 \times 2{,}52 \Rightarrow [96{,}2; 110{,}2]$$

R :

```
t.test(x)
```

Excel :

- Limite inférieure : `=MOYENNE(A1:A5) - 2.776 * ES`
- Limite supérieure : `=MOYENNE(A1:A5) + 2.776 * ES`

5.8.3 Exercice 3 Estimation d'une proportion

Dans une étude, 18 patients sur 40 ont présenté un effet secondaire.
Solution :

$$f = \frac{18}{40} = 0{,}45 \quad ; \quad ES = \sqrt{\frac{0{,}45 \cdot 0{,}55}{40}} \approx 0{,}078$$
$$IC_{95\%} = 0{,}45 \pm 1{,}96 \times 0{,}078 \Rightarrow [0{,}297; 0{,}603]$$

R :

```
prop.test(18, 40, correct = FALSE)
```

Excel :
— Proportion : `=18/40`
— IC inf. : `=f - 1.96*RACINE(f*(1-f)/40)`
— IC sup. : `=f + 1.96*RACINE(f*(1-f)/40)`

5.8.4 Exercice 4 IC pour la variance

Tensions systoliques (mmHg) : `[115, 122, 118, 121, 117, 119]`
Solution :

$$n = 6 \quad ; \quad s^2 = 6{,}17$$
$$\chi^2_{0{,}025;5} \approx 12{,}833 \quad ; \quad \chi^2_{0{,}975;5} \approx 0{,}831$$
$$IC = \left[\frac{5 \cdot 6{,}17}{12{,}833}; \frac{5 \cdot 6{,}17}{0{,}831}\right] \approx [2{,}4; 37{,}1]$$

R :

```
x <- c(115, 122, 118, 121, 117, 119)
n <- length(x)
s2 <- var(x)
c((n-1)*s2/qchisq(0.975, df=n-1), (n-1)*s2/qchisq(0.025, df=n-1))
```

Excel :
— `=VAR.P(A1:A6)`
— `=5*VAR.P()/CHI.INVERSE(0.975;5)`
— `=5*VAR.P()/CHI.INVERSE(0.025;5)`

5.8.5 Exercice 5 Interprétation clinique d'un IC

Une étude estime le taux de réponse à 65% [IC95% : 55% ; 75%].
Interprétation :
— Le taux réel se situe avec 95% de confiance entre 55% et 75%.
— Comme l'IC ne contient pas 50%, l'efficacité est significativement supérieure au hasard.

CHAPITRE 6

TESTS D'HYPOTHÈSES

6.1 Introduction

La statistique inférentielle ne se limite pas à estimer les paramètres : elle permet également de prendre des décisions à l'aide de tests d'hypothèses. On souhaite souvent savoir si une affirmation sur une population est compatible avec les données observées.

6.2 Hypothèses nulle et alternative

- Hypothèse nulle H_0 : l'hypothèse de référence (ex. : *la moyenne est égale à 100*)
- Hypothèse alternative H_1 : ce que l'on souhaite démontrer (ex. : *la moyenne est différente de 100*)

6.3 Erreurs de décision

- Erreur de type I : rejeter H_0 alors qu'elle est vraie risque noté α
- Erreur de type II : ne pas rejeter H_0 alors qu'elle est fausse risque noté β

6.4 Valeur p et niveau de signification

- La **valeur** p est la probabilité d'obtenir une statistique aussi extrême (ou plus) que celle observée, si H_0 est vraie.
- On rejette H_0 si p-valeur $< \alpha$ (souvent $\alpha = 0{,}05$).

6.5 Test unilatéral vs bilatéral

- Test bilatéral : $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$
- Test unilatéral à droite : $H_1 : \mu > \mu_0$
- Test unilatéral à gauche : $H_1 : \mu < \mu_0$

6.6 Tests usuels

6.6.1 Test sur une moyenne (σ connue)

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

6.6.2 Test sur une moyenne (σ inconnue)

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim \mathcal{T}_{n-1}$$

6.6.3 Test sur une proportion

$$Z = \frac{f - p_0}{\sqrt{p_0(1 - p_0)/n}} \sim \mathcal{N}(0, 1)$$

6.6.4 Comparaison de deux moyennes

Cas indépendant, variances égales :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{où } s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

6.6.5 Comparaison de deux proportions

$$Z = \frac{f_1 - f_2}{\sqrt{f(1 - f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{où } f = \frac{x_1 + x_2}{n_1 + n_2}$$

6.7 Application en pharmacotoxicologie

- Comparer l'effet de deux traitements (moyennes ou proportions)
- Déterminer si un médicament augmente une valeur biologique (ex. : enzymes hépatiques)
- Évaluer si un groupe présente une valeur moyenne anormale

6.8 Implémentation sur R

6.8.1 Test sur une moyenne :

```
x <- c(103, 101, 106, 97, 99)
t.test(x, mu = 100)
```

6.8.2 Comparaison de moyennes indépendantes :

```
g1 <- c(100, 102, 101)
g2 <- c(95, 96, 94)
t.test(g1, g2, var.equal = TRUE)
```

6.8.3 Comparaison de proportions :

```
prop.test(c(18, 22), c(40, 50), correct = FALSE)
```

6.9 Implémentation sur Excel

- Test sur une moyenne : `=TEST.T(A1:A5;100;2;1)`
- Test de Student (deux groupes) : `=TEST.T(A1:A3;B1:B3;2;1)`
- Test de proportion : pas natif calcul via formule Z

6.10 Conclusion

Les tests d'hypothèses permettent d'évaluer statistiquement une affirmation sur une population. Ils reposent sur une démarche rigoureuse où le risque d'erreur est contrôlé, et permettent des prises de décision objectives à partir des données échantillonnées.

6.11 Exercices

6.11.1 Exercice 1 Test sur une moyenne (bilatéral)

On a mesuré la concentration sanguine moyenne d'un médicament chez 8 patients :

$$x = [101, 98, 100, 103, 99, 102, 97, 100]$$

Test : $H_0 : \mu = 100$ contre $H_1 : \mu \neq 100$, au seuil 5%.

Solution :

- $\bar{x} = 100$, $s \approx 2$, $n = 8$
- $t = \frac{100-100}{2/\sqrt{8}} = 0$
- $|t| < t_{0.975,7} = 2.365$ donc on ne rejette pas H_0

R :

```
x <- c(101, 98, 100, 103, 99, 102, 97, 100)
t.test(x, mu = 100)
```

Excel : `=TEST.T(A1:A8;100;2;1)`

6.11.2 Exercice 2 Test unilatéral à droite sur la moyenne

Un chercheur soupçonne qu'un traitement augmente le taux d'une enzyme :

$$x = [125, 130, 128, 127, 129] \quad (\mu_0 = 120)$$

Test : $H_0 : \mu = 120$ vs $H_1 : \mu > 120$

Solution :

$$\bar{x} = 127.8, \quad s \approx 1.92, \quad t = \frac{127.8 - 120}{1.92/\sqrt{5}} \approx 9.06$$

$$t > t_{0.95,4} = 2.132 \Rightarrow \text{Rejet de } H_0$$

R :

```
x <- c(125, 130, 128, 127, 129)
t.test(x, mu = 120, alternative = "greater")
```

Excel : Calcul de la moyenne et de t, puis :
`=LOI.T.DROITE(t;4)`

6.11.3 Exercice 3 Test sur une proportion (bilatéral)

Parmi 50 patients, 18 ont eu un effet indésirable. Test : $H_0 : p = 0.3$, $H_1 : p \neq 0.3$.

Solution :

$$f = 0.36, \quad SE = \sqrt{\frac{0.3 \cdot 0.7}{50}} \approx 0.065, \quad Z \approx 0.975$$

$$|Z| < 1.96 \Rightarrow \text{non significatif}$$

R :

```
prop.test(18, 50, p = 0.3, correct = FALSE)
```

Excel :

— $f = 18/50$

— $SE = \sqrt{0.3 \cdot 0.7/50}$

— $Z = (f - 0.3)/SE$

— p-valeur $= 2 * (1 - LOI.NORMALE.STANDARD.N(Z))$

6.11.4 Exercice 4 Comparaison de deux moyennes

Traitement A : [102, 104, 98], Traitement B : [95, 93, 96]

Test : $H_0 : \mu_1 = \mu_2$

Solution :

$$\bar{x}_1 = 101.3, \quad \bar{x}_2 = 94.7, \quad s_p \approx 3.2, \quad t \approx 3.06$$

$$df = 4, \quad t_{0.975} = 2.776 \Rightarrow \text{Rejet de } H_0$$

R :

```
a <- c(102, 104, 98)
b <- c(95, 93, 96)
t.test(a, b, var.equal = TRUE)
```

Excel : `=TEST.T(A1:A3;B1:B3;2;1)`

6.11.5 Exercice 5 Comparaison de deux proportions

Effets indésirables :
- Groupe 1 : 18/40
- Groupe 2 : 10/30

Solution :

$$f_1 = 0.45, \quad f_2 = 0.33, \quad f = 0.4, \quad SE = \sqrt{0.4 \cdot 0.6 \cdot \left(\frac{1}{40} + \frac{1}{30}\right)} \approx 0.118$$

$$Z = \frac{0.45 - 0.33}{0.118} \approx 1.02 \Rightarrow \text{non significatif}$$

R :

```
prop.test(c(18, 10), c(40, 30), correct = FALSE)
```

Excel :
- $f_i = x_i/n_i$
- SE calculé
- Z, puis : `=2*(1 - LOI.NORMALE.STANDARD.N(Z))` pour la p-valeur

CHAPITRE 7

RÉGRESSION LINÉAIRE SIMPLE

7.1 Objectif

La régression linéaire simple permet de modéliser la relation entre une variable dépendante quantitative Y et une variable explicative X, à l'aide d'une droite :

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

où ε est une erreur aléatoire supposée centrée.

7.2 Hypothèses du modèle

— Linéarité : la relation entre Y et X est linéaire.
— Indépendance des observations.
— Homoscédasticité : la variance des erreurs est constante.
— Normalité des résidus.

7.3 Estimation des coefficients

Les moindres carrés ordinaires (MCO) consistent à minimiser la somme des carrés des résidus :

$$S = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Les estimateurs sont :

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

7.4 Interprétation

— $\hat{\beta}_0$: ordonnée à l'origine
— $\hat{\beta}_1$: variation de Y pour une unité de X

7.5 Qualité de l'ajustement

7.5.1 Coefficient de détermination R^2

$$R^2 = \frac{\text{variance expliquée}}{\text{variance totale}} = 1 - \frac{SCR}{SCT}$$

- $R^2 \in [0, 1]$ mesure la qualité de l'ajustement

7.5.2 Analyse des résidus

On vérifie :
- la normalité des résidus (qqplot, histogramme),
- l'homoscédasticité (résidus vs. prédictions),
- l'absence de structure (autocorrélation, points influents).

7.6 Test sur le coefficient de régression

On teste :

$$H_0 : \beta_1 = 0 \quad \text{vs} \quad H_1 : \beta_1 \neq 0$$

avec une statistique t :

$$t = \frac{\hat{\beta}_1}{\mathrm{SE}(\hat{\beta}_1)} \sim \mathcal{T}_{n-2}$$

Si la valeur-p est < 0.05, on conclut à une relation significative entre X et Y.

7.7 Application en pharmacotoxicologie

- Relation entre dose administrée et concentration sanguine.
- Relation entre âge et clairance rénale.
- Évaluation d'un biomarqueur prédictif.

7.8 Implémentation sous R

7.8.1 Exemple de jeu de données

```
dose <- c(0, 1, 2, 3, 4)
concentration <- c(0, 2, 4.2, 6.1, 8.3)
mod <- lm(concentration ~ dose)
summary(mod)
```

7.8.2 Graphique

```
plot(dose, concentration, pch=19)
abline(mod, col="blue")
```

7.9 Implémentation sous Excel

- Utiliser l'outil : **Données** > **Analyse de régression** (dans le complément *Analyse de données*).
- Insérer la colonne de X (ex. dose) et Y (ex. concentration).
- Excel fournit $\hat{\beta}_0$, $\hat{\beta}_1$, le test de Student et R^2.

7.10 Conclusion

La régression linéaire simple est un outil essentiel pour évaluer l'impact d'une variable quantitative explicative sur une réponse quantitative, en fournissant à la fois un modèle prédictif et un cadre statistique de validation.

7.11 Exercices

7.11.1 Exercice 1 Estimation manuelle des coefficients

Un chercheur étudie la relation entre la dose d'un médicament X et la concentration plasmatique Y :

Dose (mg)	Concentration (mg/L)
0	0.0
1	2.0
2	4.1
3	6.0
4	8.2

Solution :

$$\bar{x} = 2, \quad \bar{y} = 4.06$$

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{20.2}{10} = 2.02$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x} = 0.02$$

R :

```
dose <- c(0,1,2,3,4)
conc <- c(0,2,4.1,6,8.2)
mod <- lm(conc ~ dose)
summary(mod)
```

Excel :

- Formule pente : `=PENTE(B2:B6;A2:A6)`
- Ordonnée à l'origine : `=ORDOREE.ORIGINE(B2:B6;A2:A6)`

7.11.2 Exercice 2 Interprétation des coefficients

On considère le modèle estimé :

$$\hat{Y} = 0.5 + 2.5X$$

Interprétation :

— $\hat{\beta}_0 = 0.5$: concentration initiale en absence de dose

— $\hat{\beta}_1 = 2.5$: augmentation de la concentration de 2.5 mg/L par mg de dose administrée

7.11.3 Exercice 3 Qualité d'ajustement (R^2)

X	Y
1	3
2	5.1
3	7
4	9.2
5	11.1

Solution :

— Régression estimée : $\hat{Y} = 1.04 + 2.00X$

— Coefficient de détermination : $R^2 \approx 0.998$

— Interprétation : modèle très bien ajusté aux données

R :

```r
x <- 1:5
y <- c(3, 5.1, 7, 9.2, 11.1)
mod <- lm(y ~ x)
summary(mod)$r.squared
```

7.11.4 Exercice 4 Test de significativité

On considère le modèle :

$$\hat{Y} = 2 + 1.8X, \quad \text{SE}(\hat{\beta}_1) = 0.3, \quad n = 10$$

Hypothèse : $H_0 : \beta_1 = 0$

Statistique :

$$t = \frac{1.8}{0.3} = 6.0, \quad df = 8, \quad t_{0.975,8} \approx 2.31$$

Conclusion : $|t| > 2.31 \Rightarrow$ rejet de H_0. La relation est significative.

7.11.5 Exercice 5 Analyse des résidus

On estime un modèle linéaire. Le graphique des résidus en fonction des prédictions montre une forme en cône.

Question : Quelle conclusion ?

Réponse :

— Il y a **hétéroscédasticité** : la variance des erreurs augmente avec X

— Le modèle viole l'hypothèse d'homoscédasticité

— Solution : transformation des variables (log, racine) ou autre modèle

CHAPITRE 8

ANALYSE DE LA VARIANCE (ANOVA)

8.1 Introduction

L'analyse de la variance (ANOVA) permet de comparer les moyennes de plusieurs groupes (≥ 2) afin de déterminer s'il existe une différence significative. Elle généralise le test t de Student à plus de deux groupes.

8.2 Hypothèses du modèle ANOVA à un facteur

— $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$
— H_1 : au moins une moyenne diffère
— Homoscédasticité : $\sigma_1^2 = \cdots = \sigma_k^2$
— Normalité des résidus
— Indépendance des observations

8.3 Décomposition de la variance

La somme totale des carrés (SCT) est décomposée en :

$$SCT = SCE + SCR$$

— SCE : somme des carrés expliquée (entre groupes)
— SCR : somme des carrés résiduels (intra-groupes)

8.4 Statistique de test

$$F = \frac{\text{Variance expliquée}}{\text{Variance résiduelle}} = \frac{SCE/(k-1)}{SCR/(n-k)} \sim \mathcal{F}_{k-1,n-k}$$

Si $F > F_\alpha$, on rejette H_0 : au moins un groupe diffère significativement.

8.5 Conditions de validité

— Normalité (test de Shapiro-Wilk ou QQ-plot)
— Homoscédasticité (test de Levene ou Bartlett)
— Indépendance des données

8.6 Tests post-hoc (comparaisons multiples)

Si H_0 est rejetée, on identifie les groupes différents à l'aide de tests :
— Tests de Tukey HSD (honestly significant difference)
— Test de Bonferroni
— Comparaisons par paires

8.7 Application en pharmacotoxicologie

— Comparaison de l'effet de 3 doses d'un médicament sur la pression artérielle
— Étude de la variation d'un biomarqueur en fonction du traitement
— Analyse d'efficacité de plusieurs formulations d'un produit

8.8 Implémentation sous R

8.8.1 Exemple

```
dose <- factor(rep(c("Faible", "Moyenne", "Forte"), each=5))
reponse <- c(10,11,12,11,10, 14,15,14,13,14, 18,19,20,19,18)
anova_mod <- aov(reponse ~ dose)
summary(anova_mod)
```

Test de Tukey :

```
TukeyHSD(anova_mod)
```

Diagnostics :

```
plot(anova_mod) # Résidus, QQ-plot, etc.
```

8.9 Implémentation sous Excel

— Menu **Données** > **Analyse de données** > **ANOVA à un facteur**
— Saisir les groupes en colonnes (un groupe par colonne)
— Excel fournit F, p-valeur, SCE, SCR, etc.

8.10 Conclusion

L'ANOVA est une méthode incontournable pour comparer plusieurs groupes dans les essais cliniques, les études biologiques ou toxicologiques. Sa mise en uvre est facilitée par R et Excel, à condition que les hypothèses soient vérifiées.

8.11 Exercices

8.11.1 Exercice 1 ANOVA à un facteur

On mesure la pression artérielle chez des patients traités par trois doses :

Faible	Moyenne	Forte
120	130	140
118	132	138
122	131	142
119	129	141
121	130	139

Hypothèses :

$$H_0 : \mu_{\text{faible}} = \mu_{\text{moyenne}} = \mu_{\text{forte}}, \quad H_1 : \text{au moins une moyenne diffère}$$

R :

```
faible <- c(120,118,122,119,121)
moyenne <- c(130,132,131,129,130)
forte <- c(140,138,142,141,139)
groupe <- factor(rep(c("Faible","Moyenne","Forte"), each=5))
reponse <- c(faible, moyenne, forte)
mod <- aov(reponse ~ groupe)
summary(mod)
```

Excel : Menu **Données > Analyse de données > ANOVA à un facteur**
Conclusion : $p < 0.05$ on rejette H_0

8.11.2 Exercice 2 Vérification des hypothèses

Objectif : vérifier la normalité et l'homoscédasticité des résidus.
R :

```
plot(mod) # Diagrammes des résidus
shapiro.test(mod$residuals)      # Normalité
bartlett.test(reponse ~ groupe) # Homoscédasticité
```

Interprétation : Si $p > 0.05$, les hypothèses sont respectées.

8.11.3 Exercice 3 Comparaisons multiples (Tukey)

Objectif : identifier les groupes différents.
R :

```
TukeyHSD(mod)
```

Excel : Comparaisons manuelles (pas automatique).

8.11.4 Exercice 4 Cas pharmacologique

On étudie l'effet de 4 formulations d'un antibiotique :

A	B	C	D
15	17	18	20
14	16	17	19
15	18	16	21

R :

```
formulation <- factor(rep(c("A","B","C","D"), each=3))
zone <- c(15,14,15,17,16,18,18,17,16,20,19,21)
anova_mod <- aov(zone ~ formulation)
summary(anova_mod)
TukeyHSD(anova_mod)
```

Conclusion : certaines paires de formulations sont significativement différentes.

8.11.5 Exercice 5 Interprétation d'un tableau ANOVA

Données :
- $SCT = 200$, $SCE = 150$, $SCR = 50$
- $k = 3$, $n = 15$

Calcul :

$$ddl_1 = 2, \quad ddl_2 = 12$$

$$F = \frac{SCE/ddl_1}{SCR/ddl_2} = \frac{75}{4.17} \approx 18$$

Conclusion : Si $F > F_{\text{critique}} = 4.10$, alors rejet de H_0

CHAPITRE 9

HYPOTHÈSES DU MODÈLE ANOVA À DEUX FACTEURS

9.1 Introduction

L'analyse de la variance à deux facteurs (ou à double entrée) permet d'évaluer l'effet simultané de deux variables qualitatives (facteurs) sur une variable quantitative. Elle permet aussi d'analyser une éventuelle interaction entre ces deux facteurs.

9.2 Objectifs

— Tester l'effet de chaque facteur principal
— Tester l'existence d'une interaction entre les deux facteurs
— Comparer des groupes croisés (ex. : sexe et traitement)

9.3 Modèle statistique

Soit :
— A un facteur à a niveaux
— B un facteur à b niveaux
— n répétitions par combinaison
Le modèle est :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

où :
— μ est la moyenne générale
— α_i l'effet du niveau i de A
— β_j l'effet du niveau j de B
— $(\alpha\beta)_{ij}$ l'interaction entre A_i et B_j
— $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$

9.4 Hypothèses testées

— H_0^A : pas d'effet du facteur A
— H_0^B : pas d'effet du facteur B
— H_0^{AB} : pas d'interaction entre A et B

9.5 Décomposition de la variance

$$SCT = SCA + SCB + SCAB + SCR$$

avec :
— SCA : somme des carrés due au facteur A
— SCB : somme des carrés due au facteur B
— SCAB : somme des carrés due à l'interaction
— SCR : somme des carrés résiduelle

9.6 Exemple en pharmacotoxicologie

On étudie l'effet de deux formulations (A, B) et deux voies d'administration (orale, IV) sur la concentration sanguine.

Plan factoriel : 2 formulations Œ 2 voies = 4 groupes

9.7 Implémentation sous R

9.7.1 Création des données

```
formulation <- factor(rep(c("A","B"), each=6))
voie <- factor(rep(c("Orale","IV"), times=6))
y <- c(4.2, 4.0, 4.1, 5.5, 5.3, 5.4, 6.0, 6.2, 6.1, 5.0, 5.2, 5.1)
anova2 <- aov(y ~ formulation * voie)
summary(anova2)
```

9.7.2 Résultats

Le tableau ANOVA affichera : - Effet de `formulation` - Effet de `voie` - Interaction `formulation:voie`

9.8 Représentation graphique

```
interaction.plot(voie, formulation, y)
```

9.9 Implémentation sous Excel

— Pas de fonction native pour ANOVA à deux facteurs avec interaction
— Possible via complément d'analyse de données **(ANOVA à deux facteurs sans réplication)**
— Sinon : utiliser **XLSTAT**, **JASP**, ou un script R embarqué dans Excel

9.10 Conclusion

L'ANOVA à deux facteurs permet d'étudier l'effet combiné de plusieurs facteurs, ainsi que leurs interactions. C'est une méthode très utile en expérimentation biomédicale, notamment pour optimiser les conditions d'administration d'un traitement.

9.11 Exercices

9.11.1 Exercice 1 Interaction formulation et voie

On teste l'effet de deux formulations (A, B) et deux voies d'administration (orale, IV) sur la concentration plasmatique (mg/L) :

Formulation	Voie	Valeurs
A	Orale	4.2, 4.1, 4.0
A	IV	5.5, 5.4, 5.3
B	Orale	6.0, 6.2, 6.1
B	IV	5.0, 5.1, 5.2

R :

```
formulation <- factor(rep(c("A", "B"), each=6))
voie <- factor(rep(c("Orale", "IV"), times=6))
concentration <- c(4.2,4.1,4.0,5.5,5.4,5.3,6.0,6.2,6.1,5.0,5.1,5.2)
mod <- aov(concentration ~ formulation * voie)
summary(mod)
```

Conclusion : significativité des effets principaux et interaction selon les p-valeurs.

9.11.2 Exercice 2 Représentation graphique de l'interaction

On souhaite visualiser l'interaction :
R :

```
interaction.plot(voie, formulation, concentration,
                 col=c("red","blue"), lwd=2)
```

Excel : créer un graphique en lignes : une courbe par formulation selon la voie.
Conclusion : non-parallélisme interaction possible.

9.11.3 Exercice 3 Table ANOVA manuelle

On a :

$$\text{SCA} = 10.5, \quad \text{SCB} = 12.8, \quad \text{SCAB} = 5.2, \quad \text{SCR} = 3.0$$
$$ddl_A = ddl_B = ddl_{AB} = 1, \quad ddl_{res} = 8$$

Calculs :

$$F_A = \frac{10.5}{0.375} = 28, \quad F_B = \frac{12.8}{0.375} = 34.1, \quad F_{AB} = \frac{5.2}{0.375} = 13.9$$

Conclusion : tous les effets sont significatifs si $F > F_{\text{critique}}$.

9.11.4 Exercice 4 Effet croisé dose et genre

Étude de l'effet de deux doses sur hommes et femmes :

Sexe	Basse	Haute
H	12,13,14	20,22,21
F	10,11,9	18,19,17

R :

```
dose <- factor(rep(c("Basse", "Haute"), times=6))
genre <- factor(rep(c("H", "F"), each=6))
y <- c(12,13,14,20,22,21,10,11,9,18,19,17)
mod <- aov(y ~ genre * dose)
summary(mod)
```

Conclusion : interpréter les effets principaux et leur interaction.

9.11.5 Exercice 5 Cas pharmacologique complet

Effet de trois formulations (A, B, C) et deux voies (Injection, Spray) sur la réponse immunitaire :

Formulation	Voie	Réponse
A	Inj.	50, 52
A	Spray	47, 48
B	Inj.	55, 56
B	Spray	53, 52
C	Inj.	60, 61
C	Spray	59, 60

R :

```
form <- factor(rep(c("A","B","C"), each=4))
voie <- factor(rep(c("Inj", "Spray"), times=6))
reponse <- c(50,52,47,48,55,56,53,52,60,61,59,60)
mod <- aov(reponse ~ form * voie)
summary(mod)
```

Conclusion : déterminer si les deux facteurs et leur interaction ont un effet significatif.

CHAPITRE 10

EXERCICES SUPPLÉMENTAIRES (ANOVA À UN FACTEUR ET À DEUX FACTEURS)

10.1 ANOVA à un facteur

10.1.1 Exercice 1 Effet de traitements sur la glycémie

Un pharmacologue teste trois traitements sur la glycémie :

Traitement A	Traitement B	Traitement C
100	110	130
102	115	128
98	112	132

a) Compléter le tableau de variation (SCE, SCR, SCT)
b) Calculer les valeurs de CM_E, CM_R, F
c) Conclure à $\alpha = 5\%$

10.1.2 Exercice 2 Étude de 4 régimes alimentaires

Régime 1	Régime 2	Régime 3	Régime 4
70	80	90	100
72	82	91	99
71	78	89	102

Tester $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ à l'aide de l'ANOVA1.

10.1.3 Exercice 3 Données résumées

On donne : $SCT = 300$, $SCE = 200$, $SCR = 100$, $k = 3$, $n = 12$
a) Compléter le tableau de variation
b) Calculer les F
c) Conclure à $\alpha = 5\%$

10.2 ANOVA à deux facteurs

10.2.1 Exercice 4 Effet de dose et de sexe

Sexe	Basse dose	Haute dose
Homme	12, 13	19, 21
Femme	10, 11	16, 17

a) Compléter le tableau ANOVA2
b) Tester les effets principaux et l'interaction

10.2.2 Exercice 5 Résumé chiffré ANOVA2

On donne : $SCA = 30$, $SCB = 25$, $SCAB = 15$, $SCR = 80$
avec $a = 2$, $b = 3$, $n = 4$
a) Compléter le tableau de variation
b) Calculer les F
c) Conclure à 5% pour chaque facteur

10.2.3 Exercice 6 Formulation et Durée

Effet de trois formulations et deux durées sur la toxicité hépatique :

Formulation	7 jours	14 jours
F1	30, 32, 31	28, 29, 27
F2	35, 34, 33	31, 32, 30
F3	38, 37, 36	34, 35, 33

a) Construire le tableau ANOVA2 complet
b) Tester les effets de la formulation, de la durée, et de leur interaction

ANNEXE A TABLES STATISTIQUES

10.3 Loi normale

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998	0.9998

10.4 Table Fisher-Snedecor

Table de Fisher-Snedecor, $\alpha = 5\%$ (95$^{\text{ème}}$ centile)

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10	20	30	40	50	60	80	100	200	500	1000
1	161.45	19.50	10.13	7.71	6.61	5.99	5.59	5.32	5.12	4.96	4.35	4.14	4.05	3.99	3.94	3.87	3.84	3.78	3.75	3.74
2	159.20	19.16	9.28	6.99	5.79	5.41	5.14	4.86	4.63	4.39	3.90	3.72	3.64	3.58	3.54	3.48	3.46	3.41	3.39	3.38
3	133.53	17.06	8.19	6.11	5.05	4.39	4.39	3.86	3.49	3.20	2.77	2.56	2.47	2.42	2.38	2.34	2.32	2.29	2.28	2.28
4	106.90	13.67	6.66	5.08	4.42	4.17	3.79	3.73	3.39	3.20	2.75	2.60	2.54	2.49	2.45	2.41	2.39	2.36	2.35	2.35
5	82.25	11.03	5.59	4.48	4.02	3.88	3.44	3.31	3.08	2.97	2.65	2.54	2.48	2.45	2.41	2.39	2.37	2.35	2.34	2.34
6	66.41	9.50	4.74	4.06	3.77	3.44	3.25	3.05	2.85	2.72	2.46	2.38	2.33	2.30	2.28	2.26	2.25	2.24	2.23	2.23
7	57.27	8.48	4.12	3.73	3.35	3.07	2.93	2.75	2.59	2.46	2.23	2.17	2.13	2.11	2.09	2.08	2.07	2.06	2.05	2.05
8	50.97	7.73	3.70	3.45	3.12	2.88	2.70	2.59	2.44	2.32	2.11	2.06	2.02	2.00	1.99	1.98	1.97	1.96	1.95	1.95
9	46.58	7.15	3.41	3.21	2.91	2.68	2.55	2.41	2.31	2.21	2.02	1.97	1.93	1.92	1.91	1.90	1.89	1.89	1.88	1.88
10	43.17	6.71	3.18	3.02	2.75	2.55	2.42	2.31	2.22	2.13	1.95	1.91	1.87	1.86	1.85	1.85	1.84	1.83	1.83	1.83
20	31.41	4.35	2.77	2.48	2.28	2.14	2.06	1.97	1.92	1.86	1.65	1.59	1.56	1.54	1.53	1.52	1.52	1.51	1.50	1.50
30	25.17	3.72	2.56	2.33	2.14	2.02	1.94	1.88	1.83	1.78	1.59	1.54	1.51	1.49	1.48	1.47	1.47	1.46	1.46	1.45
40	22.25	3.64	2.47	2.26	2.09	1.98	1.91	1.85	1.80	1.75	1.56	1.51	1.48	1.47	1.46	1.45	1.44	1.44	1.44	1.43
50	20.50	3.58	2.42	2.21	2.04	1.94	1.87	1.82	1.77	1.72	1.53	1.49	1.46	1.45	1.44	1.44	1.43	1.43	1.42	1.42
60	19.16	3.54	2.38	2.18	2.02	1.92	1.86	1.80	1.75	1.70	1.52	1.47	1.44	1.43	1.42	1.42	1.41	1.41	1.41	1.40
80	17.52	3.48	2.34	2.15	1.99	1.90	1.83	1.78	1.73	1.68	1.49	1.45	1.42	1.41	1.40	1.39	1.39	1.39	1.39	1.38
100	16.35	3.46	2.32	2.13	1.97	1.88	1.82	1.77	1.72	1.67	1.48	1.44	1.41	1.40	1.39	1.39	1.38	1.38	1.38	1.37
200	14.13	3.41	2.29	2.10	1.94	1.85	1.79	1.75	1.70	1.65	1.46	1.42	1.39	1.38	1.38	1.37	1.37	1.36	1.36	1.36
500	12.11	3.39	2.28	2.08	1.92	1.83	1.77	1.72	1.68	1.63	1.44	1.40	1.38	1.37	1.36	1.36	1.35	1.35	1.35	1.34
1000	11.61	3.38	2.28	2.08	1.91	1.83	1.76	1.72	1.67	1.62	1.43	1.40	1.37	1.36	1.36	1.35	1.35	1.34	1.34	1.34

10.5 Table du test t de Student (bilatéral, $\alpha = 0{,}05$)

ν	0.50	0.80	0.70	0.60	0.40	0.30	0.20	0.10	0.05	0.02	0.01
1	0.000	0.325	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.142	0.289	0.845	0.617	0.816	1.061	1.886	2.920	4.303	6.965	9.925
3	0.137	0.277	0.842	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.134	0.271	0.741	0.741	1.094	1.533	2.132	2.776	3.747	3.747	4.604
5	0.132	0.267	0.686	0.727	1.096	1.476	2.015	2.571	3.365	4.032	4.032
6	0.130	0.265	0.606	0.708	1.030	1.440	1.943	2.447	3.143	3.707	3.707
7	0.130	0.263	0.540	0.701	1.065	1.415	1.895	2.365	3.020	3.499	3.499
8	0.129	0.262	0.519	0.684	1.055	1.397	1.860	2.306	2.896	3.355	3.355
9	0.129	0.261	0.508	0.670	1.043	1.383	1.833	2.262	2.821	3.250	3.250
10	0.129	0.260	0.503	0.660	1.036	1.372	1.812	2.228	2.764	3.169	3.169
11	0.129	0.260	0.500	0.652	1.030	1.363	1.796	2.201	2.718	3.106	3.106
12	0.128	0.259	0.497	0.645	1.026	1.356	1.782	2.179	2.681	3.055	3.055
13	0.128	0.259	0.494	0.638	1.023	1.350	1.771	2.160	2.650	3.012	3.012
14	0.128	0.258	0.492	0.633	1.021	1.345	1.761	2.145	2.624	2.977	2.977
15	0.128	0.258	0.490	0.629	1.020	1.341	1.753	2.131	2.602	2.947	2.947
16	0.128	0.258	0.489	0.625	1.019	1.337	1.746	2.120	2.583	2.921	2.921
17	0.127	0.257	0.488	0.622	1.018	1.333	1.740	2.110	2.567	2.898	2.898
18	0.127	0.257	0.487	0.620	1.017	1.330	1.734	2.101	2.552	2.878	2.878
19	0.127	0.257	0.486	0.617	1.016	1.328	1.729	2.093	2.539	2.861	2.861
20	0.127	0.257	0.485	0.615	1.015	1.325	1.725	2.086	2.528	2.845	2.845
∞	0.126	0.253	0.385	0.524	0.842	1.036	1.282	1.645	1.960	2.326	2.576

Nota. ν est le nombre de degrés de liberté.
Le quantile d'ordre $1 - \frac{\alpha}{2}$ se lit dans la colonne $P = \alpha$.
Le quantile d'ordre $1 - \alpha$ se lit dans la colonne $P = 2\alpha$.

10.6 Table du test du Khi-deux (χ^2) pour $\alpha = 0{,}05$

n	0,995	0,99	0,975	0,95	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,01
1	0,000	0,000	0,001	0,004	0,016	0,064	0,148	0,275	0,455	0,708	1,074	1,642	2,706	3,841	6,635
2	0,010	0,020	0,051	0,103	0,211	0,446	0,713	1,022	1,386	2,408	3,219	4,605	5,991	5,991	9,210
3	0,072	0,115	0,216	0,352	0,584	1,005	1,424	1,669	2,366	3,665	4,642	6,251	7,815	7,815	11,345
4	0,207	0,297	0,484	0,711	1,064	1,649	2,195	2,366	3,357	4,878	6,251	7,779	9,488	9,488	13,277
5	0,412	0,554	0,831	1,145	1,610	2,343	3,000	3,000	4,351	5,989	7,779	9,236	11,070	11,070	15,086
6	0,676	0,872	1,237	1,635	2,204	3,070	3,828	3,645	5,348	7,042	9,236	10,645	12,592	12,592	16,812
7	0,989	1,239	1,690	2,167	2,833	3,822	4,674	4,343	6,346	8,035	10,597	12,017	14,067	14,067	18,475
8	1,344	1,646	2,180	2,733	3,490	4,594	5,527	5,041	7,344	8,945	11,778	13,362	15,507	15,507	20,090
9	1,735	2,088	2,700	3,325	4,168	5,380	6,390	5,899	8,343	9,782	12,899	14,684	16,919	16,919	21,666
10	2,156	2,558	3,247	3,940	4,865	6,179	7,260	6,737	9,342	10,558	13,957	15,987	18,307	18,307	23,209
11	2,603	3,053	3,816	4,575	5,578	6,989	8,136	7,584	10,341	11,275	14,953	17,275	19,675	19,675	24,725
12	3,074	3,571	4,404	5,226	6,304	7,807	9,034	8,520	11,340	11,940	15,812	18,549	21,026	21,026	26,217
13	3,565	4,107	5,009	5,892	7,042	8,634	9,926	9,480	12,339	12,552	16,623	19,812	22,362	22,362	27,688
14	4,074	4,660	5,629	6,571	7,790	9,468	10,819	10,468	13,338	13,107	17,388	21,064	23,685	23,685	29,141
15	4,601	5,229	6,262	7,261	8,547	10,307	11,701	11,478	14,337	13,617	18,108	22,307	24,996	24,996	30,578
16	5,144	5,812	6,908	7,962	9,312	11,146	12,591	12,488	15,336	14,129	18,885	23,542	26,296	26,296	32,000
17	5,703	6,408	7,564	8,672	10,085	11,982	13,480	13,497	16,335	14,643	19,639	24,769	27,587	27,587	33,409
18	6,275	7,015	8,231	9,390	10,865	12,816	14,368	14,507	17,334	15,158	20,370	25,989	28,869	28,869	34,805
19	6,861	7,633	8,907	10,117	11,651	13,648	15,256	15,516	18,333	15,675	21,088	27,204	30,144	30,144	36,191
20	7,459	8,260	9,591	10,851	12,443	14,478	16,142	16,526	19,332	16,193	21,796	28,412	31,410	31,410	37,566
21	8,069	8,897	10,283	11,591	13,240	15,306	17,028	17,535	20,331	16,713	22,492	29,615	32,671	32,671	38,932
22	8,690	9,542	10,982	12,338	14,041	16,132	17,912	18,545	21,330	17,234	23,178	30,813	33,924	33,924	40,289
23	9,319	10,196	11,689	13,091	14,848	16,957	18,796	19,554	22,329	17,755	23,854	32,007	35,172	35,172	41,638
24	9,957	10,856	12,401	13,848	15,659	17,779	19,679	20,563	23,328	18,278	24,520	33,196	36,415	36,415	42,980
25	10,604	11,524	13,120	14,611	16,473	18,600	20,561	21,572	24,327	18,801	25,179	34,382	37,652	37,652	44,314
26	11,260	12,198	13,844	15,379	17,292	19,420	21,443	22,581	25,326	19,324	25,830	35,563	38,885	38,885	45,642
27	11,923	12,879	14,573	16,151	18,114	20,239	22,323	23,590	26,325	19,849	26,473	36,741	40,113	40,113	46,963
28	12,593	13,565	15,308	16,928	18,939	21,057	23,203	24,599	27,324	20,374	27,110	37,916	41,337	41,337	48,278
29	13,270	14,256	16,047	17,708	19,768	21,874	24,083	25,608	28,323	20,899	27,741	39,087	42,557	42,557	49,588
30	13,954	14,953	16,791	18,493	20,599	22,690	24,961	26,617	29,322	21,425	28,366	40,256	43,773	43,773	50,892
40	23,684	24,433	26,509	28,412	30,582	34,764	37,566	36,015	43,773	30,432	36,015	51,805	55,758	55,758	63,691
50	35,478	36,465	38,885	39,997	42,942	47,400	50,481	50,481	55,758	39,364	45,641	64,319	67,505	67,505	76,154
60	47,229	48,260	51,739	51,739	55,672	60,391	63,835	63,835	67,505	47,400	55,758	76,154	79,082	79,082	86,659

BIBLIOGRAPHIE

[1] A. Morineau et Y.-M. Chatelin, *L'analyse statistique des données : Apprendre, comprendre et réaliser avec Excel*, Ellipses, 2005.

[2] Pierre Lafaye de Micheaux, Rémy Drouilhet, et Benoît Liquet. *Le logiciel R : Maîtriser le langage, effectuer des analyses statistiques.* Springer, 2012.

[3] Gilbert Saporta. *Probabilités, analyse des données et statistique.* Technip, 2006.

[4] David Collet. *Modelling Survival Data in Medical Research*, 3e édition. CRC Press, 2015.

[5] Wayne W. Daniel et Chad L. Cross. *Biostatistics : A Foundation for Analysis in the Health Sciences*, 10e éd. Wiley, 2013.

[6] Marcello Pagano et Kimberlee Gauvreau. *Principles of Biostatistics*, 2e éd. CRC Press, 2018.

[7] Peter Dalgaard. *Introductory Statistics with R*, 2e éd. Springer, 2008.

[8] David J. Biau et Stéphane Kernéis. *Statistique appliquée à la recherche clinique.* Elsevier Masson, 2017.

[9] Bernard Rosner. *Fundamentals of Biostatistics*, 8e éd. Cengage Learning, 2015.

[10] Claude Hamonet et Laurent Gerbaud. *Statistique en médecine et biologie.* De Boeck Supérieur, 2007.

[11] Brian S. Everitt et Anders Skrondal. *The Cambridge Dictionary of Statistics*, 4e éd. Cambridge University Press, 2010.

[12] Thomas B. Newman. *Concepts of Biostatistics in Medicine.* University of California Press, 2003.

[13] Betty R. Kirkwood et Jonathan A. C. Sterne. *Essential Medical Statistics*, 2e éd. Wiley-Blackwell, 2003.

[14] Peter Armitage, Geoffrey Berry, et J.N.S. Matthews. *Statistical Methods in Medical Research*, 4e éd. Wiley, 2008.