Université Mohamed Khider de Biskra
Faculté des sciences exactes
Département de Mathématiques

MÉMOIRE DE MASTER

Spécialité : Probabilités et statistique - Réf : P3188/2026

TITRE

On quelques tests non paramétriques et applications

Présenté et soutenu par

Nesrine BOURAS

Date de soutenance : 06/06/2026

JURY DE SOUTENANCE

Nom et Prénom	Grade	Qualité
Sonia TOUBA	MCB	**Président**
Naceur RAHMANI	MCA	**Encadrant**
Ibtissem DJABER	MAA	**Examinateur**

Année universitaire : 2025/2026

Dédicace

⋆ *À mes chers* ***parents****, source inépuisable de tendresse, de patience et de sacrifices, qui n'ont cessé de m'encourager et de me soutenir tout au long de mon parcours académique.*

⋆ *À mes* ***frères et sœurs****, pour leur affection et leur présence bienveillante à chaque instant.*

⋆ *À tous mes* ***amis et collègues*** *de la promotion, pour les moments partagés et les souvenirs précieux.*

⋆ *À tous ceux qui* ***aiment la science*** *et cherchent la connaissance avec passion.*

Remerciements

- Je tiens à exprimer ma profonde gratitude à mon directeur de mémoire, **Mr. RAHMANI Naceur**, pour son encadrement rigoureux, sa disponibilité constante, ses précieux conseils scientifiques et sa patience tout au long de la réalisation de ce travail. Sa rigueur intellectuelle a été une source d'inspiration permanente.

- Mes sincères remerciements vont aux membres du jury, **Mme. TOUBA Sonia** et **Mme. DJABER Ibtissem**, pour l'honneur qu'elles me font d'évaluer ce travail et d'y apporter leur expertise précieuse.

- Je remercie chaleureusement l'ensemble des enseignants du Département de Mathématiques de l'Université Mohamed Khider de Biskra pour la qualité de la formation dispensée durant ces années de master.

- Enfin, je remercie du fond du cœur ma **famille** pour son soutien indéfectible, ses encouragements permanents et sa confiance qui ont été le moteur de ma réussite tout au long de ce parcours.

Résumé / Abstract

Résumé

Ce mémoire porte sur l'étude des tests non paramétriques, considérés comme l'une des principales méthodes statistiques utilisées lorsque les hypothèses des tests paramétriques ne sont pas vérifiées, notamment l'hypothèse de normalité des données. Nous avons présenté quelques notions fondamentales liées à l'inférence statistique, ainsi qu'un ensemble de tests non paramétriques permettant la comparaison des distributions et l'étude de l'homogénéité et de l'indépendance entre les groupes. Des applications pratiques ont été réalisées à l'aide de données simulées en utilisant le logiciel statistique `R`, ce qui a permis de montrer comment appliquer ces tests et interpréter leurs résultats.

Mots-clés : Tests non paramétriques, rangs, Kolmogorov-Smirnov, Mann-Whitney, Wilcoxon, Kruskal-Wallis, Friedman, logiciel `R`.

Abstract

This thesis focuses on the study of non-parametric tests, which are considered one of the main statistical methods used when the assumptions of parametric tests are not satisfied, especially the normality assumption of the data. We presented some fundamental concepts related to statistical inference, as well as a set of non-parametric tests used for comparing distributions and studying homogeneity and independence between groups. Practical applications were carried out using simulated data with the statistical software `R`, which allowed us to illustrate how to apply these tests and interpret their results.

Keywords : Non-parametric tests, ranks, Kolmogorov-Smirnov, Mann-Whitney, Wilcoxon, Kruskal-Wallis, Friedman, `R` software.

Notations et symboles

Symbole	Signification
H_0	Hypothèse nulle
H_1	Hypothèse alternative
α	Risque de première espèce (seuil de signification)
β	Risque de deuxième espèce
$1-\beta$	Puissance du test
X, Y	Variables aléatoires
$x_1, \ldots, x_n$	Réalisation d'un échantillon de taille n
$x_{(i)}$	La i-ème statistique d'ordre
$\overline{X}_n$	Moyenne empirique
S_n^2	Variance empirique corrigée
$\mathbf{E}[\cdot]$	Espérance mathématique
$\mathbf{Var}[\cdot]$	Variance mathématique
$\mathcal{N}(0,1)$	Loi normale centrée réduite
$\mathcal{N}(\mu, \sigma^2)$	Loi normale de paramètres μ et σ^2
$F_n(x)$	Fonction de répartition empirique
$F_0(x)$	Fonction de répartition théorique
R_i	Rang de l'observation i
θ	Paramètre inconnu
$\hat{\theta}$	Estimateur du paramètre θ
μ_r	Moment théorique d'ordre r
m_r	Moment empirique d'ordre r
D_n	Statistique de Kolmogorov-Smirnov
$d_{n,\alpha}$	Valeur critique de Kolmogorov-Smirnov
χ^2_ν	Loi du khi-deux à ν degrés de liberté
Q	Statistique du khi-deux d'ajustement

Symbole	**Signification**
U	Statistique du test de Mann-Whitney
W^+	Statistique du test de Wilcoxon
H	Statistique du test de Kruskal-Wallis
χ^2_F	Statistique du test de Friedman
p-valeur	Probabilité critique du test
z_γ	Quantile d'ordre γ de $\mathcal{N}(0,1)$
$\mathbb{I}_A(x)$	Fonction indicatrice de l'ensemble A
$\xrightarrow{\mathcal{L}}$	Convergence en loi

Table des matières

Dédicace i

Remerciements ii

Résumé / Abstract iii

Notations et symboles iv

Table des matières vi

Table des figures ix

Liste des tableaux x

Introduction générale 1

1 Fondements théoriques des tests statistiques 3

1.1 L'inférence statistique 3

1.1.1 Cadre général 3

1.1.2 Estimation ponctuelle 4

1.1.3 Estimation par intervalle de confiance 7

1.2 Tests d'hypothèses 7

1.2.1 Cadre général et formulation 7

1.2.2 Erreurs de décision et risques associés 8

1.2.3 La p-valeur 9

1.3 Limites des tests paramétriques 10
1.3.1 L'hypothèse de normalité 10
1.3.2 Homogénéité des variances et robustesse 12
1.4 Principes des tests non paramétriques 12
1.4.1 La notion de rang 13
1.4.2 Avantages, inconvénients et tableau comparatif 14

2 Étude théorique des principaux tests non paramétriques 16
2.1 Tests d'ajustement pour un échantillon 16
2.1.1 Test de Kolmogorov-Smirnov 16
2.1.2 Test d'ajustement du khi-deux 19
2.2 Tests pour deux échantillons 20
2.2.1 Test de Mann-Whitney 20
2.2.2 Test des rangs signés de Wilcoxon 20
2.3 Tests pour plusieurs échantillons 21
2.3.1 Test de Kruskal-Wallis 21
2.3.2 Test de Friedman 22

3 Applications pratiques et analyse comparative 23
3.1 Application du test de Kolmogorov-Smirnov 23
3.1.1 Contexte et données 23
3.1.2 Code `R` et résultats 24
3.1.3 Interprétation 25
3.2 Application du test du khi-deux 25
3.2.1 Contexte et données 25
3.2.2 Code `R` et résultats 26
3.2.3 Interprétation 27
3.3 Application du test de Mann-Whitney 27
3.3.1 Contexte et données 27

3.3.2 Code R et résultats 28
3.3.3 Interprétation 29
3.4 Application du test de Wilcoxon 29
3.4.1 Contexte et données 29
3.4.2 Code R et résultats 30
3.4.3 Interprétation 31
3.5 Application du test de Kruskal-Wallis 31
3.5.1 Contexte et données 31
3.5.2 Code R et résultats 32
3.5.3 Interprétation 34
3.6 Application du test de Friedman 34
3.6.1 Contexte et données 34
3.6.2 Code R et résultats 35
3.6.3 Interprétation 36
3.7 Analyse comparative des six tests 37
3.7.1 Tableau de synthèse multicritère 37
3.7.2 Comparaison graphique des p-valeurs 39
3.7.3 Guide pratique de sélection du test 39

Conclusion générale **39**

Bibliographie **40**

Table des figures

1.1 Illustration graphique des erreurs de type I (α) et type II (β) dans un test unilatéral droit. La zone rouge correspond au rejet incorrect de H_0, la zone orange au non-rejet incorrect de H_0. 9

1.2 Comparaison de la loi normale avec des distributions non normales : exponentielle (asymétrique à droite) et Laplace (queues plus épaisses que la normale). 11

1.3 Arbre de décision pour le choix du test non paramétrique adapté selon la structure des données 15

2.1 Illustration de la statistique D_n : distance maximale entre F_n et F_0 . . 18

2.2 Densités de la loi χ^2_ν pour $\nu \in \{1, 3, 5, 9\}$ 19

2.3 Algorithme du test des rangs signés de Wilcoxon 21

3.1 Comparaison des p-valeurs obtenues pour les six applications. Seuil $\alpha = 5\%$ (ligne rouge pointillée) : cinq tests sur six conduisent à rejeter H_0. ($^*p < 0.05$, $^{**}p < 0.01$) 39

Liste des tableaux

1.1 Tableau des décisions et types d'erreurs dans un test statistique 8
1.2 Méthodes de vérification de l'hypothèse de normalité 11
1.3 Avantages et inconvénients des tests non paramétriques 14
1.4 Correspondances entre tests paramétriques et non paramétriques 14

3.1 Résultats du test de Kolmogorov-Smirnov 25
3.2 Nombre de vainqueurs observés par position de départ 26
3.3 Résultats du test du khi-deux et résidus de Pearson 27
3.4 Notes d'examen de mathématiques par genre 28
3.5 Résultats du test de Mann-Whitney 29
3.6 Notes des 8 étudiants selon les deux méthodes d'enseignement 30
3.7 Résultats du test de Wilcoxon (rangs signés) 31
3.8 Croissance des plants (cm) selon le niveau d'éclairage 32
3.9 Résultats du test de Kruskal-Wallis 33
3.10 Comparaisons post-hoc par paires (correction de Bonferroni) 33
3.11 Scores des 8 sujets aux trois tests de mémoire 34
3.12 Résultats du test de Friedman 36
3.13 Comparaisons post-hoc par paires (test de Wilcoxon apparié) 36
3.14 Synthèse complète des six tests non paramétriques étudiés 37
3.15 Comparaison multicritère des six tests 38

Introduction générale

La statistique inférentielle constitue l'un des piliers fondamentaux de l'analyse quantitative des données. Elle permet de tirer des conclusions sur une population à partir de l'observation d'un échantillon, avec un niveau de confiance entièrement quantifié. Au cœur de cette démarche se trouve le **test statistique**, qui formalise la prise de décision entre deux hypothèses concurrentes H_0 et H_1.

Les méthodes statistiques classiques, dites **paramétriques**, reposent sur des hypothèses structurelles fortes : normalité de la distribution des données, homogénéité des variances entre les groupes, indépendance des observations. Lorsque ces conditions sont vérifiées, ces tests offrent une excellente puissance statistique. Cependant, dans de nombreuses situations pratiques — données médicales, scores psychologiques, mesures biologiques, données en sciences sociales — ces hypothèses ne sont pas satisfaites, en particulier lorsque les échantillons sont de petite taille ou que les données présentent des valeurs aberrantes ou une forte asymétrie.

C'est dans ce contexte que les **tests non paramétriques**, développés principalement au cours du XX^{e} siècle par *Kolmogorov* (1933), *Smirnov* (1948), *Wilcoxon* (1945), *Mann et Whitney* (1947), *Kruskal et Wallis* (1952) et *Friedman* (1937), apportent une réponse méthodologique essentielle. Ces méthodes, dites à *distribution libre*, ne font aucune hypothèse sur la forme paramétrique de la loi sous-jacente et s'avèrent donc applicables dans un spectre de situations bien plus large.

Ce mémoire est organisé en trois chapitres :

- **Chapitre 1 :** Rappels des fondements théoriques de l'inférence statistique — estimation, tests d'hypothèses, erreurs de décision — et introduction aux

principes généraux des tests non paramétriques.

- **Chapitre 2 :** Étude théorique rigoureuse des six tests non paramétriques : Kolmogorov-Smirnov, khi-deux, Mann-Whitney, Wilcoxon, Kruskal-Wallis et Friedman, avec leurs propriétés asymptotiques et illustrations graphiques.
- **Chapitre 3 :** Applications pratiques sur données réelles sous `R`, et analyse comparative multicritère des six tests.

Chapitre 1

Fondements théoriques des tests statistiques

Introduction du chapitre

Ce chapitre pose les bases conceptuelles et mathématiques nécessaires à la compréhension des tests statistiques. Dans un premier temps, nous rappelons les notions fondamentales de l'inférence statistique : estimation ponctuelle et par intervalle de confiance. Nous présentons ensuite le cadre général des tests d'hypothèses, les erreurs de décision et la notion de puissance. Nous examinons les hypothèses requises par les tests paramétriques et leurs limites pratiques. Enfin, nous introduisons les concepts fondateurs des tests non paramétriques, notamment la notion de rang, qui constitue le fil conducteur de l'ensemble du mémoire.

1.1 L'inférence statistique

1.1.1 Cadre général

L'inférence statistique est la branche de la statistique qui permet de tirer des conclusions sur une **population** à partir d'un échantillon aléatoire prélevé dans cette population. On suppose que les observations $(X_1, X_2, \ldots, X_n)$ sont issues d'une loi de probabilité $\mathcal{P}_\theta$, où θ est un paramètre inconnu appartenant à un espace paramétrique

Θ. L'objectif est de déduire des informations sur θ à partir des données observées.

Définition 1.1

Un n-**échantillon** est une suite $(X_1, X_2, \ldots, X_n)$ de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) selon une loi de probabilité $\mathcal{P}_\theta$, où $\theta \in \Theta$ est un paramètre inconnu.

Une fois l'échantillon défini, l'inférence statistique poursuit deux grands objectifs complémentaires : l'**estimation** du paramètre inconnu θ, et le **test d'hypothèses** sur sa valeur. Ces deux aspects sont étroitement liés : un estimateur de bonne qualité est souvent à la base de la construction d'un test puissant.

1.1.2 Estimation ponctuelle

L'estimation ponctuelle consiste à proposer une valeur unique, calculée à partir des données, comme approximation du paramètre inconnu.

Définition 1.2

Un **estimateur ponctuel** de θ est une statistique $T_n = T(X_1, \ldots, X_n)$, c'est-à-dire une fonction mesurable de l'échantillon, dont la valeur observée $T_n(x_1, \ldots, x_n)$ est appelée **estimation** de θ.

Il est crucial de distinguer : l'**estimateur** (variable aléatoire, fonction des données) de l'**estimation** (valeur numérique obtenue sur un échantillon observé). Par exemple, la moyenne empirique $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ est un estimateur de la moyenne théorique μ, tandis que $\overline{x}_n = 7{,}3$ est une estimation.

Un bon estimateur doit satisfaire plusieurs propriétés :

Propriétés souhaitables d'un estimateur

- **Sans biais :** $\mathbf{E}[T_n] = \theta$ pour tout $\theta \in \Theta$. L'estimateur ne surestime ni ne sous-estime systématiquement le paramètre.
- **Convergence (consistance) :** $T_n \xrightarrow{P} \theta$ quand $n \to \infty$. Plus l'échantillon est grand, plus l'estimation s'approche de la vraie valeur.
- **Efficacité :** $\mathbf{Var}(T_n)$ est minimale parmi tous les estimateurs sans biais. La

borne de Cramér-Rao donne la variance minimale atteignable : $\mathbf{Var}(T_n) \geq \frac{1}{n \cdot I(\theta)}$, où $I(\theta)$ est l'information de Fisher.

- **Normalité asymptotique :** $\sqrt{n}(T_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2(\theta))$. Cette propriété permet de construire des intervalles de confiance asymptotiques.

Méthode des moments

La méthode des moments, introduite par *Pearson* (1894), est l'une des méthodes les plus simples pour construire des estimateurs. Elle repose sur le principe d'égalisation des moments théoriques aux moments empiriques observés.

Définition 1.3

Soit X une variable aléatoire de densité $f(x;\theta)$. Le **moment théorique d'ordre** r est :

$$\mu_r(\theta) = \mathbf{E}[X^r] = \int_{-\infty}^{+\infty} x^r \, f(x;\theta)\, \mathrm{d}x, \quad r \geq 1.$$

Le **moment empirique d'ordre** r, calculé sur un n-échantillon, est :

$$m_r = \frac{1}{n} \sum_{i=1}^{n} X_i^r.$$

L'estimateur par la méthode des moments $\hat{\theta}_{MM}$ est la solution du système d'équations $\mu_r(\theta) = m_r$ pour $r = 1, 2, \ldots, k$ (où k est la dimension de θ). Cet estimateur est généralement consistant et asymptotiquement normal, mais pas nécessairement efficace.

Exemple 1.1

Application : Soit $X \sim \mathcal{E}(\lambda)$, de densité $f(x;\lambda) = \lambda e^{-\lambda x}$, $x \geq 0$. Le moment théorique d'ordre 1 est $\mu_1(\lambda) = \mathbf{E}[X] = 1/\lambda$. En égalisant avec le moment empirique : $1/\lambda = \overline{X}_n$, on obtient l'estimateur des moments $\hat{\lambda}_{MM} = 1/\overline{X}_n$.

Estimateur du Maximum de Vraisemblance

L'estimateur du maximum de vraisemblance (EMV), proposé par *Fisher* (1922), est la méthode d'estimation ponctuelle la plus utilisée en statistique mathématique. Elle repose sur le principe que la valeur la plus plausible du paramètre est celle qui rend les

données observées les plus probables.

Définition 1.4

Soit $(x_1, \ldots, x_n)$ une réalisation du n-échantillon. La **fonction de vraisemblance** est :

$$L(\theta;\, x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \theta).$$

En pratique, on maximise la **log-vraisemblance** : $\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i; \theta)$.

L'**Estimateur du Maximum de Vraisemblance (EMV)** est :

$$\hat{\theta}_{MV} = \arg\max_{\theta \in \Theta} \ell(\theta).$$

L'équation de vraisemblance $\dfrac{\partial \ell(\theta)}{\partial \theta} = 0$ donne la condition nécessaire pour un maximum intérieur. Lorsque la solution est unique, elle définit l'EMV. La propriété clé de l'EMV est son optimalité asymptotique :

Théorème 1.1

Sous des conditions de régularité standards (loi C^3 en θ, support fixe, information de Fisher finie), l'EMV $\hat{\theta}_{MV}$ vérifie :

- **Consistance :** $\hat{\theta}_{MV} \xrightarrow{P} \theta$ quand $n \to \infty$.
- **Normalité asymptotique :**

$$\sqrt{n}\left(\hat{\theta}_{MV} - \theta\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,\ \frac{1}{I(\theta)}\right),$$

  où $I(\theta) = -\mathbf{E}\left[\dfrac{\partial^2 \log f(X;\theta)}{\partial \theta^2}\right]$ est l'**information de Fisher**.
- **Efficacité asymptotique :** l'EMV atteint la borne de Cramér-Rao asymptotiquement.

Ces propriétés remarquables font de l'EMV la référence en estimation paramétrique. Cependant, comme nous le verrons, elles nécessitent des hypothèses sur la forme de la distribution — hypothèses que les méthodes non paramétriques s'affranchissent précisément.

1.1.3 Estimation par intervalle de confiance

L'estimation ponctuelle, aussi précise soit-elle, ne quantifie pas l'incertitude associée. L'estimation par intervalle de confiance y remédie en fournissant un encadrement probabiliste du paramètre.

Définition 1.5

Un **intervalle de confiance de niveau** $1-\alpha$ pour le paramètre θ est un intervalle aléatoire $[T_n^-, T_n^+]$, construit à partir des données, tel que :

$$\mathbf{P}\left(T_n^- \leq \theta \leq T_n^+\right) = 1-\alpha,$$

où $\alpha \in (0,1)$ est le risque choisi (souvent 5%). La quantité $1-\alpha$ est appelée **niveau de confiance** ou **coefficient de confiance**.

L'interprétation fréquentiste est importante : si l'on répète l'expérience un grand nombre de fois, environ $100(1-\alpha)\%$ des intervalles construits contiendront la vraie valeur de θ. L'intervalle observé contient ou ne contient pas θ, mais on lui attribue une confiance de $1-\alpha$.

Remarque 1.1

Pour un grand échantillon, le Théorème Central Limite (TCL) garantit que $\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$, ce qui donne l'intervalle de confiance asymptotique pour la moyenne :

$$\left[\overline{X}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \quad \overline{X}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

Pour les petits échantillons sous normalité, on utilise la loi de Student t_{n-1}.

1.2 Tests d'hypothèses

1.2.1 Cadre général et formulation

Après l'estimation, le deuxième grand problème de l'inférence statistique est le **test d'hypothèses** : peut-on, sur la base des données, valider ou invalider une affirmation sur le paramètre θ ?

Définition 1.6

Un **test statistique** est une procédure de décision permettant de choisir entre deux hypothèses concurrentes :

- H_0 : **hypothèse nulle**, hypothèse de référence à tester (par exemple : $\theta = \theta_0$).
- H_1 : **hypothèse alternative**, acceptée si H_0 est rejetée (par exemple : $\theta \neq \theta_0$).

La décision repose sur le calcul d'une **statistique de test** T_n, dont on compare la valeur observée à une valeur critique issue de sa distribution sous H_0.

L'hypothèse alternative peut prendre trois formes : bilatérale $H_1 : \theta \neq \theta_0$, unilatérale droite $H_1 : \theta > \theta_0$, ou unilatérale gauche $H_1 : \theta < \theta_0$. Le choix dépend du contexte scientifique de l'étude.

1.2.2 Erreurs de décision et risques associés

Toute décision fondée sur un échantillon aléatoire comporte un risque d'erreur. On distingue deux types d'erreurs, résumés dans le tableau suivant :

Table 1.1: Tableau des décisions et types d'erreurs dans un test statistique

Décision prise	H_0 **est vraie**	H_1 **est vraie**
On accepte H_0	Décision correcte $(1-\alpha)$	Erreur de type II (β)
On rejette H_0	Erreur de type I (α)	Décision correcte $(1-\beta)$

Ces deux types d'erreurs sont définis formellement comme suit :

Définition 1.7

- Le **risque de première espèce** $\alpha = \mathbf{P}(\text{rejeter } H_0 \mid H_0 \text{ vraie})$ est la probabilité d'un *faux positif*. C'est le seuil de signification du test, fixé a priori par le statisticien (en général 1%, 5% ou 10%).
- Le **risque de deuxième espèce** $\beta = \mathbf{P}(\text{accepter } H_0 \mid H_1 \text{ vraie})$ est la

probabilité d'un *faux négatif*. Il dépend de l'écart entre θ et θ_0, et diminue quand n augmente.

- La **puissance** du test est $\pi = 1 - \beta = \mathbf{P}(\text{rejeter } H_0 \mid H_1 \text{ vraie})$.

Il existe un compromis fondamental entre α et β : à taille d'échantillon fixée, réduire α augmente généralement β. La seule façon de réduire simultanément les deux risques est d'augmenter la taille de l'échantillon n. La figure 1.1 illustre graphiquement ces concepts.

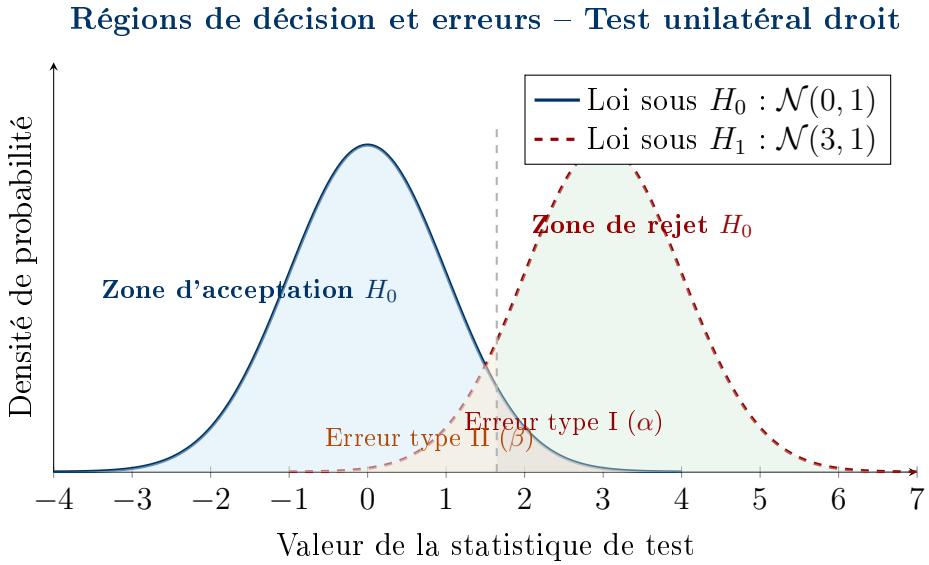


Figure 1.1: Illustration graphique des erreurs de type I (α) et type II (β) dans un test unilatéral droit. La zone rouge correspond au rejet incorrect de H_0, la zone orange au non-rejet incorrect de H_0.

1.2.3 La p-valeur

En pratique, plutôt que de comparer la statistique de test à une valeur critique, on utilise la p-**valeur** (ou probabilité critique), qui synthétise l'information contenue dans les données.

Définition 1.8

La p-**valeur** est la probabilité, calculée sous H_0, d'obtenir une valeur de la statistique de test au moins aussi extrême que la valeur observée :

$$p = \mathbf{P}_{H_0}(T_n \geq t_{\text{obs}}) \quad \text{(test unilatéral droit)}.$$

Plus p est petit, plus les données sont incompatibles avec H_0.

Règle de décision basée sur la p-valeur

- Si p-valeur $\leq \alpha$: on **rejette** H_0 au niveau α. Les données sont incompatibles avec H_0.
- Si p-valeur $> \alpha$: on **ne rejette pas** H_0 au niveau α. On ne conclut pas pour autant que H_0 est vraie : l'absence de preuve n'est pas une preuve d'absence.

Attention : La p-valeur n'est pas la probabilité que H_0 soit vraie !

1.3 Limites des tests paramétriques

1.3.1 L'hypothèse de normalité

Les tests paramétriques classiques — test t de Student, ANOVA, test F — reposent de manière cruciale sur l'hypothèse que les données suivent une loi normale. Cette hypothèse est souvent justifiée par le Théorème Central Limite pour les grands échantillons, mais devient problématique pour les petits échantillons ou les distributions fortement asymétriques.

Définition 1.9

L'**hypothèse de normalité** postule que la variable étudiée X suit une loi normale $\mathcal{N}(\mu, \sigma^2)$, dont la densité de probabilité est :

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R},\ \mu \in \mathbb{R},\ \sigma > 0.$$

Cette loi est symétrique, en forme de cloche, entièrement caractérisée par sa moyenne μ et son écart-type σ.

La figure 1.2 illustre des distributions qui s'écartent de la normalité : distribution exponentielle (fortement asymétrique), distribution de Laplace (à queues plus lourdes). Sur ces distributions, les tests paramétriques perdent en fiabilité.

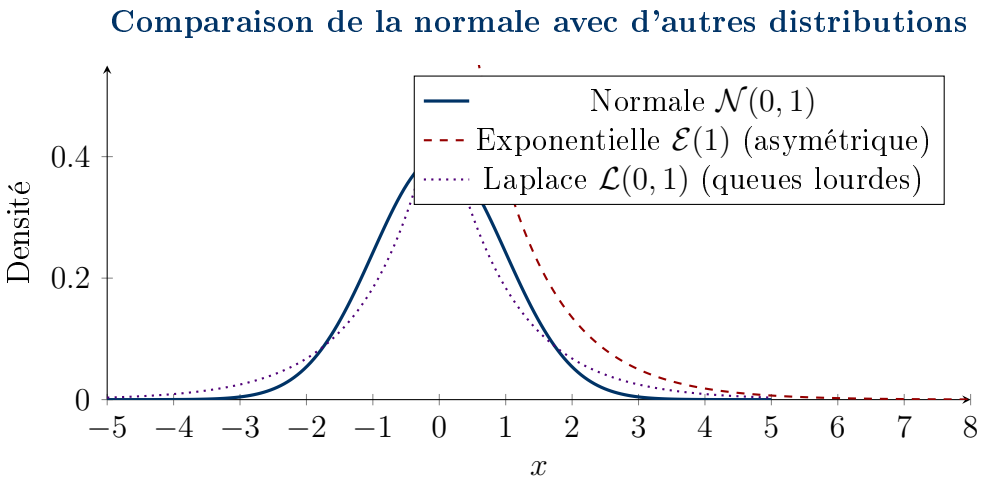


Figure 1.2: Comparaison de la loi normale avec des distributions non normales : exponentielle (asymétrique à droite) et Laplace (queues plus épaisses que la normale).

La vérification de la normalité peut être réalisée par différentes méthodes, résumées dans le tableau suivant :

Table 1.2: Méthodes de vérification de l'hypothèse de normalité

Méthode	**Type**	**Description et utilisation**
Histogramme	Graphique	Forme en cloche symétrique autour de la moyenne
Graphique Q-Q	Graphique	Points alignés sur la première bissectrice si normalité
Boîte à moustaches	Graphique	Symétrie de la boîte et des moustaches
Test de Shapiro-Wilk	Test	Très puissant pour $n \leq 50$
Test de Kolmogorov-Smirnov	Test	Applicable pour tout n, F_0 connue
Skewness / Kurtosis	Mesures	≈ 0 et ≈ 3 pour la loi normale

1.3.2 Homogénéité des variances et robustesse

Outre la normalité, les tests paramétriques classiques (ANOVA, test t à deux échantillons indépendants) supposent l'**homogénéité des variances** : $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_K^2$. Cette hypothèse peut être vérifiée par les tests de Levene (robuste) ou de Bartlett (sensible à la non-normalité).

Lorsque ces conditions ne sont pas réunies, les tests paramétriques perdent leur validité : leur niveau réel peut s'éloigner significativement du niveau nominal α. C'est là qu'intervient la notion de **robustesse** :

Définition 1.10

Un test statistique est dit **robuste** si ses conclusions restent valides malgré des écarts modérés aux hypothèses sous-jacentes (normalité, homoscédasticité, indépendance). La robustesse est une propriété graduelle et depend de la taille de l'échantillon et du type d'écart.

1.4 Principes des tests non paramétriques

Face aux limites des méthodes paramétriques, les tests non paramétriques offrent une alternative robuste et flexible.

Définition 1.11

Un test statistique est dit **non paramétrique** (ou à **distribution libre**) lorsqu'il ne fait aucune hypothèse sur la forme paramétrique de la distribution des données. Sa statistique de test a une distribution qui ne dépend d'aucun paramètre inconnu de la population.

Cette définition implique que les tables de valeurs critiques de ces tests sont universelles : elles ne dépendent que de la taille de l'échantillon n, et non des paramètres μ, σ^2, etc.

1.4.1 La notion de rang

Le concept central sur lequel reposent la plupart des tests non paramétriques est celui de **rang**. Plutôt que de travailler avec les valeurs brutes des observations, on les remplace par leur classement, ce qui élimine l'influence de la distribution sous-jacente.

Définition 1.12

Le **rang** de l'observation x_i dans l'échantillon $(x_1, x_2, \ldots, x_n)$ est sa position R_i dans le classement par ordre croissant. La plus petite valeur reçoit le rang 1, la plus grande le rang n.

En cas d'**ex-aequo** (plusieurs observations égales), on attribue à chacune le **rang moyen** des rangs qu'elles auraient reçus si elles avaient été légèrement différentes.

Illustration : Soit l'échantillon $\{15,\ 8,\ 22,\ 8,\ 31,\ 15\}$.

- Classement croissant : $8,\ 8,\ 15,\ 15,\ 22,\ 31$.
- 8 apparaît aux rangs 1 et 2 $\Rightarrow$ rang moyen $= (1+2)/2 = 1{,}5$.
- 15 apparaît aux rangs 3 et 4 $\Rightarrow$ rang moyen $= (3+4)/2 = 3{,}5$.
- Rangs finals : $8 \to 1{,}5$; $8 \to 1{,}5$; $15 \to 3{,}5$; $15 \to 3{,}5$; $22 \to 5$; $31 \to 6$.

Cette transformation en rangs présente deux avantages majeurs : elle rend le test invariant par rapport à la distribution, et elle réduit l'influence des valeurs aberrantes. La proposition suivante donne les propriétés fondamentales des rangs.

Proposition 1.1

Pour un échantillon de taille n sans ex-aequo, où les rangs $R_1, R_2, \ldots, R_n$ sont une permutation de $\{1, 2, \ldots, n\}$:

$$\sum_{i=1}^{n} R_i = \frac{n(n+1)}{2}, \qquad \mathbf{E}[R_i] = \frac{n+1}{2}, \qquad \mathbf{Var}(R_i) = \frac{n^2-1}{12}.$$

Ces résultats constituent la base du calcul des espérances et variances des statistiques de test non paramétriques.

1.4.2 Avantages, inconvénients et tableau comparatif

Les tests non paramétriques présentent des caractéristiques distinctives qui les rendent particulièrement adaptés à certaines situations :

Table 1.3: Avantages et inconvénients des tests non paramétriques

Avantages	**Inconvénients**
Aucune hypothèse sur la distribution	Moins puissants si la normalité est vérifiée
Robustes aux valeurs aberrantes	Perte d'information par transformation en rangs
Applicables aux données ordinales	Pas d'estimation directe des paramètres
Efficaces avec de petits échantillons	Tests post-hoc plus complexes
Valides asymptotiquement sans condition	Intervalles de confiance plus difficiles à construire

En complément, le tableau 1.4 rappelle les correspondances entre les tests paramétriques et leurs équivalents non paramétriques. Ces correspondances guideront notre choix tout au long des chapitres suivants.

Table 1.4: Correspondances entre tests paramétriques et non paramétriques

Test non param.	**Équivalent param.**	**Situation**
Kolmogorov-Smirnov	Test de normalité	Adéquation à une loi continue
Khi-deux ajustement	–	Adéquation à une loi discrète
Mann-Whitney	Test t indépendant	2 groupes indépendants
Wilcoxon	Test t apparié	2 mesures appariées
Kruskal-Wallis	ANOVA 1 facteur	≥ 3 groupes indépendants
Friedman	ANOVA répétées	≥ 3 traitements, mêmes individus

La figure 1.3 propose un arbre de décision synthétique permettant de choisir rapi-

dement le test non paramétrique adapté à la structure des données.

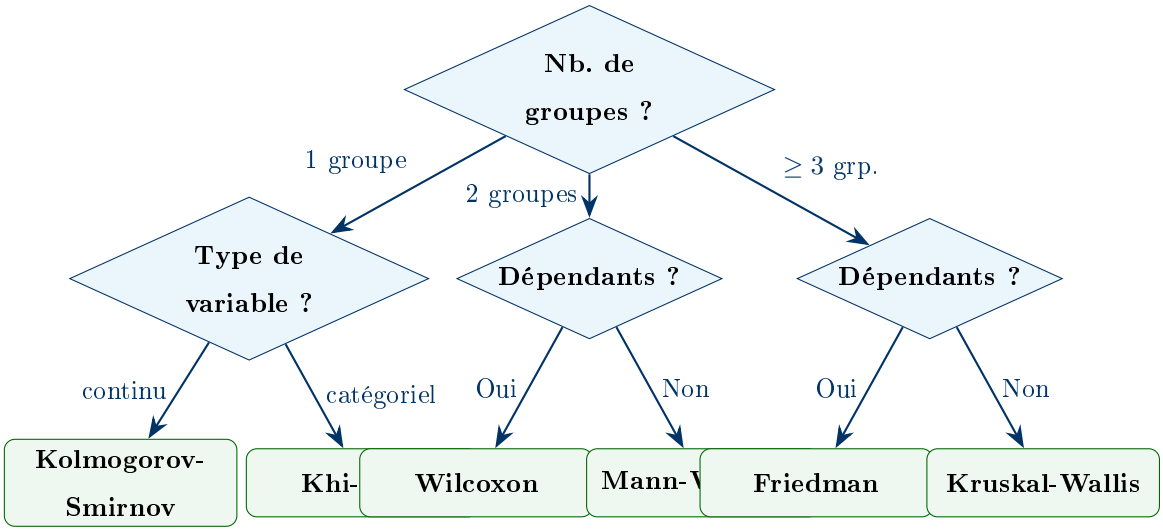


Figure 1.3: Arbre de décision pour le choix du test non paramétrique adapté selon la structure des données

Chapitre 2

Étude théorique des principaux tests non paramétriques

Introduction du chapitre

Ce chapitre présente l'étude théorique des six principaux tests non paramétriques. Pour chaque test, nous formulons les hypothèses H_0 et H_1, définissons la statistique de test, établissons sa distribution asymptotique, précisons les conditions d'application et la règle de décision, puis illustrons par un exemple numérique. Les tests sont classés selon le nombre d'échantillons impliqués : tests d'ajustement à un seul échantillon, tests pour deux échantillons (indépendants ou appariés), et tests pour plusieurs échantillons.

2.1 Tests d'ajustement pour un échantillon

2.1.1 Test de Kolmogorov-Smirnov

Définition 2.1

La **fonction de répartition empirique** d'un échantillon $(x_1, \ldots, x_n)$ est :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{x_i \leq x\}}.$$

Théorème 2.1

Thé

Sous H_0, presque sûrement :

$$\sup_{x\in\mathbb{R}} |F_n(x) - F_0(x)| \xrightarrow{p.s.} 0 \quad (n \to \infty).$$

Hypothèses du test

$$\begin{cases} H_0 : F_X = F_0 \\ H_1 : F_X \neq F_0 \end{cases}$$

Définition 2.2

La **statistique de Kolmogorov-Smirnov** est :

$$D_n = \sup_{x\in\mathbb{R}} |F_n(x) - F_0(x)|.$$

Proposition 2.1

Si $(x_{(1)} \leq \cdots \leq x_{(n)})$ est l'échantillon ordonné :

$$D_n = \max_{1\leq i\leq n} \max\left\{ \left| F_0(x_{(i)}) - \frac{i}{n} \right|, \left| F_0(x_{(i)}) - \frac{i-1}{n} \right| \right\}.$$

Théorème 2.2

Sous H_0, pour F_0 continue :

$$\sqrt{n}\, D_n \xrightarrow{\mathcal{L}} K,$$

où K est la **loi de Kolmogorov**. On rejette H_0 si $D_n > d_{n,\alpha}$ ou si p-valeur $\leq \alpha$.

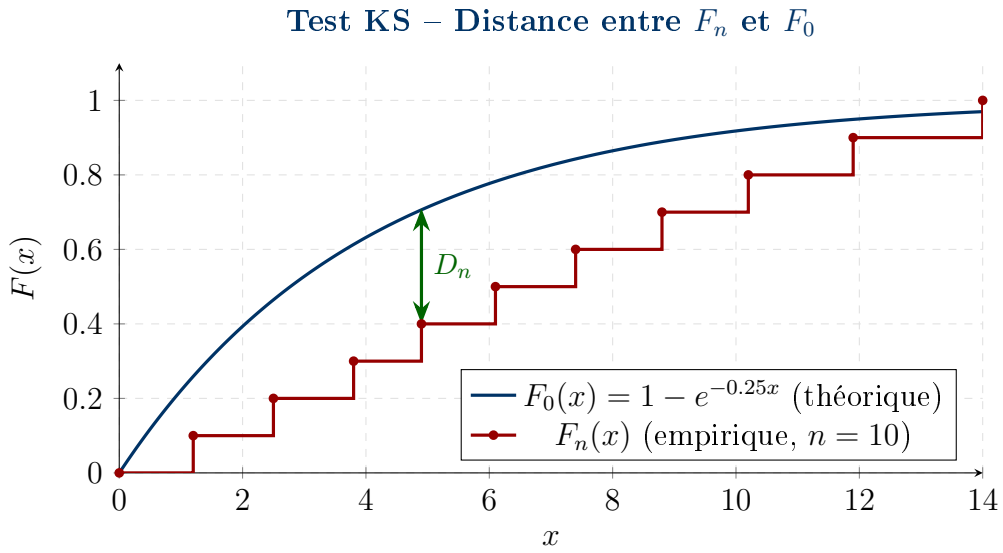


Figure 2.1: Illustration de la statistique D_n : distance maximale entre F_n et F_0

Important

Condition : F_0 doit être entièrement spécifiée (paramètres connus). Si les paramètres sont estimés, utiliser le **test de Lilliefors**.

Exemple 2.1

Données : Notes de 12 étudiants : $6, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19{,}5$.

$H_0 : X \sim \mathcal{E}(0{,}1)$. Seuil $\alpha = 5\%$.

i	$x_{(i)}$	$F_0(x_{(i)})$	i/n	$(i-1)/n$	$\lvert F_0 - i/n \rvert$	$\lvert F_0 - (i-1)/n \rvert$
1	6	0.451	0.083	0.000	0.368	0.451
2	8	0.551	0.167	0.083	0.384	**0.468**
3	9	0.593	0.250	0.167	0.343	0.426
4	10	0.632	0.333	0.250	0.299	0.382
5	11	0.667	0.417	0.333	0.250	0.334
6	12	0.699	0.500	0.417	0.199	0.282
7	13	0.727	0.583	0.500	0.144	0.227
8	15	0.777	0.667	0.583	0.110	0.194
9	16	0.798	0.750	0.667	0.048	0.131
10	17	0.817	0.833	0.750	0.016	0.067
11	18	0.835	0.917	0.833	0.082	0.002
12	19.5	0.857	1.000	0.917	0.143	0.060

$D_{12} = 0{,}468$. Valeur critique : $d_{12;\,0{,}05} = 0{,}375$.

Comme $D_{12} = 0{,}468 > 0{,}375$, **on rejette** H_0 au seuil 5%.

2.1.2 Test d'ajustement du khi-deux

Définition 2.3

La **statistique du khi-deux d'ajustement** est :

$$Q = \sum_{k=1}^{m} \frac{(N_k - np_{0k})^2}{np_{0k}},$$

où N_k est l'effectif observé dans la classe k et p_{0k} est la probabilité théorique.

Théorème 2.3

Sous H_0, quand $n \to \infty$:

$$Q \xrightarrow{\mathcal{L}} \chi^2_{m-1}.$$

Si r paramètres sont estimés : $Q \xrightarrow{\mathcal{L}} \chi^2_{m-r-1}$.

Important

Condition d'application : Chaque effectif théorique $np_{0k} \geq 5$. On rejette H_0 si $Q > \chi^2_{m-1}(1-\alpha)$.

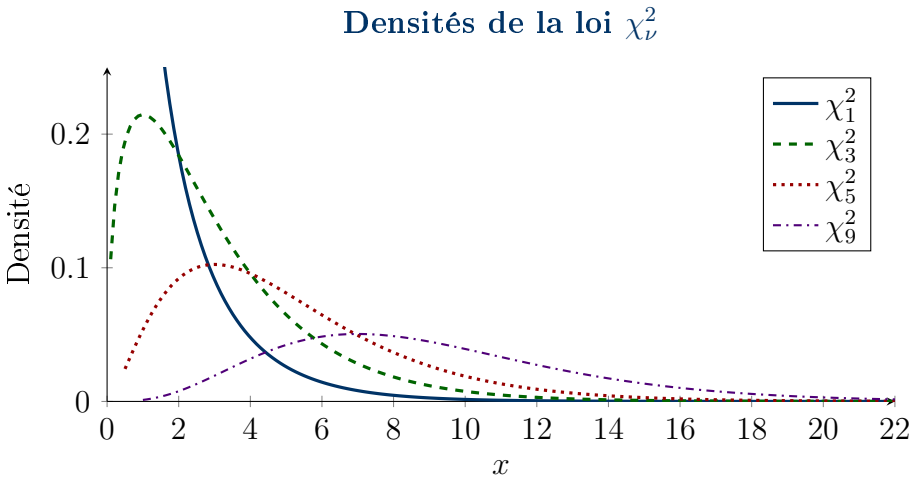


Figure 2.2: Densités de la loi χ^2_ν pour $\nu \in \{1, 3, 5, 9\}$

2.2 Tests pour deux échantillons

2.2.1 Test de Mann-Whitney

Définition 2.4

Soient $(X_1, \ldots, X_n)$ et $(Y_1, \ldots, Y_m)$ deux échantillons indépendants. La **statistique de Mann-Whitney** est :

$$U_X = W_X - \frac{n(n+1)}{2}, \quad U_Y = W_Y - \frac{m(m+1)}{2}, \quad U = \min(U_X, U_Y),$$

où W_X et W_Y sont les sommes des rangs dans l'échantillon combiné.

Théorème 2.4

Sous $H_0 : F_X = F_Y$:

$$\mathbf{E}(U) = \frac{nm}{2}, \quad \mathbf{Var}(U) = \frac{nm(n+m+1)}{12}.$$

Pour $\min(n, m) > 8$:

$$Z = \frac{U - nm/2}{\sqrt{nm(n+m+1)/12}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

On rejette H_0 si $|Z| > z_{1-\alpha/2}$.

2.2.2 Test des rangs signés de Wilcoxon

Définition 2.5

Soient $D_i = X_i - Y_i$ les différences appariées et $R_i^* = \text{rang}(|D_i|)$. La **statistique de Wilcoxon** est :

$$W^+ = \sum_{\{i\,:\,D_i>0\}} R_i^*.$$

Théorème 2.5

Sous H_0 (symétrie autour de 0) :

$$\mathbf{E}(W^+) = \frac{n^*(n^*+1)}{4}, \quad \mathbf{Var}(W^+) = \frac{n^*(n^*+1)(2n^*+1)}{24}.$$

Pour $n^* > 25$: $Z = \dfrac{W^+ - \mathbf{E}(W^+)}{\sqrt{\mathbf{Var}(W^+)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$.

Étape 1. Calculer les différences $D_i = X_i - Y_i$.

↓

Étape 2. Exclure les $D_i = 0$; noter n^* le nombre restant.

↓

Étape 3. Classer les $|D_i|$ et attribuer les rangs R_i^*.

↓

Étape 4. Calculer $W^+ = \sum_{D_i>0} R_i^*$.

↓

Étape 5. Comparer à la valeur critique.

Figure 2.3: Algorithme du test des rangs signés de Wilcoxon

2.3 Tests pour plusieurs échantillons

2.3.1 Test de Kruskal-Wallis

Définition 2.6

La **statistique de Kruskal-Wallis** est :

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{K} \frac{R_i^2}{n_i} - 3(n+1),$$

où $n = \sum_{i=1}^{K} n_i$ est la taille totale et R_i la somme des rangs du groupe i.

Théorème 2.6

Sous $H_0 : F_1 = \cdots = F_K$:

$$H \xrightarrow{\mathcal{L}} \chi^2_{K-1}.$$

On rejette H_0 si $H > \chi^2_{K-1}(1-\alpha)$.

Remarque 2.1

Correction pour ex-aequo :

$$H_c = \frac{H}{1 - \frac{\sum_g t_g(t_g^2 - 1)}{n^3 - n}},$$

où t_g est la taille du g-ème groupe d'ex-aequo.

2.3.2 Test de Friedman

Définition 2.7

La **statistique de Friedman**, basée sur les rangs intra-blocs, est :

$$\chi_F^2 = \frac{12}{NK(K+1)} \sum_{j=1}^{K} R_j^2 - 3N(K+1),$$

où N est le nombre de blocs, K le nombre de traitements et R_j la somme des rangs du traitement j.

Théorème 2.7

Sous $H_0 : F_1 = \cdots = F_K$:

$$\chi_F^2 \xrightarrow{\mathcal{L}} \chi_{K-1}^2.$$

On rejette H_0 si $\chi_F^2 \geq \chi_{K-1}^2(1-\alpha)$.

Chapitre 3

Applications pratiques et analyse comparative

Introduction du chapitre

Ce chapitre met en oeuvre les six tests étudiés dans le chapitre précédent sur des données réelles, à l'aide du logiciel statistique `R`. Pour chaque application, nous présentons d'abord le contexte et les données proposés, puis le code `R` commenté, les sorties numériques, et enfin une interprétation statistique détaillée. Le chapitre se clôt par une analyse comparative multicritère des six tests, permettant de synthétiser leurs caractéristiques et de guider le praticien dans le choix de la méthode adaptée à sa situation.

3.1 Application du test de Kolmogorov-Smirnov

3.1.1 Contexte et données

Une enquête menée auprès de $n = 10$ jeunes diplômés d'une université algérienne a relevé le délai d'attente (en mois) entre l'obtention du diplôme et l'accès au premier emploi. Les valeurs observées sont :

$$X = \{6,\ 12,\ 24,\ 16,\ 18,\ 20,\ 19,\ 9,\ 25,\ 14{,}5\}.$$

Un conseiller en orientation suggère que ces délais sont répartis de manière uniforme entre 6 et 25 mois, ce qui correspondrait à une égalité des chances dans la recherche d'emploi. Nous cherchons à tester statistiquement cette affirmation.

Formulation des hypothèses :

$$\begin{cases} H_0 : X \sim \mathcal{U}[6,25], \quad F_0(x) = \dfrac{x-6}{19} \\ H_1 : X \not\sim \mathcal{U}[6,25] \end{cases} \qquad \alpha = 5\%$$

3.1.2 Code R et résultats

```r
X <- c(6, 12, 24, 16, 18, 20, 19, 9, 25, 14.5)

# Visualisation : histogramme + densite theorique
par(mfrow = c(1, 2))
hist(X, col = "steelblue", border = "white", freq = FALSE,
     main = "Histogramme des delais", xlab = "Mois")
curve(dunif(x, 6, 25), add = TRUE, col = "red", lwd = 2)
legend("topleft", legend = "Uniforme[6,25]", col = "red", lwd
   = 2)

# Trac  F_n vs F_0
plot(ecdf(X), col = "navy", lwd = 2,
     main = "Fn vs F0 (uniforme)", xlab = "Mois", ylab = "F(x)
        ")
curve(punif(x, 6, 25), add = TRUE, col = "red", lwd = 2, lty =
    2)
legend("topleft", c("Fn empirique","F0 uniforme"),
       col = c("navy","red"), lty = c(1,2), lwd = 2)

# Test KS
res <- ks.test(X, "punif", min = 6, max = 25)
print(res)
# Sortie R :
```

```r
# One-sample Kolmogorov-Smirnov test
# D = 0.14737,  p-value = 0.9601
# Alternative hypothesis: two-sided
cat("Decision : on", ifelse(res$p.value < 0.05,"REJETTE","ne
   rejette pas"), "H0\n")
```

Listing 3.1: Test de Kolmogorov-Smirnov – Délai d'accès à l'emploi

Table 3.1: Résultats du test de Kolmogorov-Smirnov

D_n	p-valeur	Seuil α	Décision
0.1474	0.9601	0.05	Ne pas rejeter H_0

3.1.3 Interprétation

La statistique $D_{10} = 0{,}1474$ reflète un écart maximal faible entre la distribution empirique et la loi uniforme théorique. La p-valeur très élevée ($= 0{,}9601 \gg \alpha = 0{,}05$) indique que cet écart est tout à fait compatible avec la variabilité aléatoire attendue sous H_0.

Conclusion : Au seuil de signification de 5%, on ne rejette pas H_0. Les données sont statistiquement compatibles avec l'hypothèse que les délais d'obtention d'emploi suivent une loi uniforme sur $[6, 25]$ mois. La suggestion du conseiller ne peut pas être infirmée par cet échantillon.

3.2 Application du test du khi-deux

3.2.1 Contexte et données

Un spécialiste des courses hippiques souhaite vérifier si la position de départ (de 1 à 8) avantage certains chevaux. Sur 144 courses observées, il a compté le nombre de victoires pour chaque position de départ :

Table 3.2: Nombre de vainqueurs observés par position de départ

Position de départ	1	2	3	4	5	6	7	8
Victoires observées (N_k)	29	19	18	25	17	10	15	11
Victoires théoriques (np_k)	18	18	18	18	18	18	18	18

Si toutes les positions étaient équivalentes, chaque position devrait avoir une probabilité de victoire égale à $1/8$, soit $144/8 = 18$ victoires attendues.

Formulation des hypothèses :

$$\begin{cases} H_0 : p_1 = p_2 = \cdots = p_8 = 1/8 \quad \text{(équiprobabilité)} \\ H_1 : \exists\, k \text{ tel que } p_k \neq 1/8 \end{cases} \qquad \alpha = 5\%$$

3.2.2 Code R et résultats

```r
obs <- c(29, 19, 18, 25, 17, 10, 15, 11)
theo <- rep(144/8, 8)  # 18 attendu par position

# Visualisation : effectifs observes vs theoriques
barplot(rbind(obs, theo), beside = TRUE,
        names.arg = paste0("Pos.", 1:8),
        col = c("steelblue","tomato"),
        legend.text = c("Observe","Theorique (18)"),
        main = "Victoires par position de depart",
        ylab = "Nombre de victoires")
abline(h = 18, lty = 2, col = "darkred")

# Test du khi-deux
res <- chisq.test(obs, p = rep(1/8, 8))
print(res)
# Sortie R :
# Chi-squared test for given probabilities
# X-squared = 16.333, df = 7, p-value = 0.02224
```

```r
cat("Residus de Pearson :\n")
print(round(res$residuals, 3))
```

Listing 3.2: Test du khi-deux – Positions de départ des chevaux

Table 3.3: Résultats du test du khi-deux et résidus de Pearson

	Pos.1	**Pos.2**	**Pos.3**	**Pos.4**	**Pos.5**	**Pos.6**	**Pos.7**	**Pos.8**
N_k observé	29	19	18	25	17	10	15	11
np_k théorique	18	18	18	18	18	18	18	18
Résidu	+2,59	+0,24	0	+1,65	−0,24	−1,89	−0,71	−1,65

$Q = 16{,}333$; ddl $= 7$; p-valeur $= 0{,}022$; $\chi^2_7(0{,}95) = 14{,}07$

3.2.3 Interprétation

La statistique $Q = 16{,}333$ dépasse la valeur critique $\chi^2_7(0{,}95) = 14{,}07$. La p-valeur $= 0{,}022 < \alpha = 0{,}05$. Les résidus de Pearson montrent que la position 1 (+2,59) est sur-représentée, et la position 6 (−1,89) sous-représentée.

Conclusion : Au seuil de 5%, on rejette H_0. Il existe une différence statistiquement significative dans les probabilités de victoire selon la position de départ : partir en position 1 semble avantager les chevaux.

3.3 Application du test de Mann-Whitney

3.3.1 Contexte et données

Un enseignant de mathématiques compare les résultats d'un examen entre deux groupes d'élèves : un groupe de garçons ($n = 5$) et un groupe de filles ($m = 5$). Les données collectées sont :

Table 3.4: Notes d'examen de mathématiques par genre

Groupe	Notes observées (/20)					Médiane
Garçons	10	9	6,5	8	11	9
Filles	14	17	16	13	18	16

Les deux groupes sont indépendants. On ne peut pas supposer la normalité des notes (petits échantillons), ce qui justifie le recours au test de Mann-Whitney plutôt qu'au test t.

Formulation des hypothèses :

$$\begin{cases} H_0 : F_{\text{garçons}} = F_{\text{filles}} \quad \text{(même distribution des notes)} \\ H_1 : F_{\text{garçons}} \neq F_{\text{filles}} \end{cases} \qquad \alpha = 5\%$$

3.3.2 Code R et résultats

```r
garcons <- c(10, 9, 6.5, 8, 11)
filles  <- c(14, 17, 16, 13, 18)

# Boxplot comparatif
boxplot(list(Garcons = garcons, Filles = filles),
        col = c("lightblue","lightpink"),
        main = "Distribution des notes par genre",
        ylab = "Note (/20)", ylim = c(5, 20))
stripchart(list(Garcons = garcons, Filles = filles),
           method = "jitter", vertical = TRUE,
           add = TRUE, pch = 19, col = c("navy","darkred"))

# Test de Mann-Whitney
res <- wilcox.test(garcons, filles, alternative = "two.sided",
    exact = FALSE)
print(res)
# Sortie R :
```

```
# Wilcoxon rank-sum test with continuity correction
# W = 0, p-value = 0.007937
# Alternative hypothesis: true location shift is not equal to
    0
```

Listing 3.3: Test de Mann-Whitney – Notes de mathématiques

Table 3.5: Résultats du test de Mann-Whitney

U	p-valeur	Seuil α	Décision
0	0.0079	0.05	Rejeter H_0

3.3.3 Interprétation

La statistique $U = 0$ correspond au cas extrême où toutes les notes des filles sont supérieures à toutes les notes des garçons : aucune paire (X_i, Y_j) ne vérifie $X_i > Y_j$. La p-valeur $= 0{,}0079 < \alpha = 0{,}05$.

Conclusion : Au seuil de 5%, on rejette H_0. Les filles obtiennent des notes significativement plus élevées que les garçons en mathématiques, et ce résultat est très fortement significatif ($p < 1\%$).

3.4 Application du test de Wilcoxon

3.4.1 Contexte et données

Un pédagogue cherche à comparer l'efficacité de deux méthodes d'enseignement des mathématiques : une méthode moderne (basée sur les nouvelles technologies) et une méthode traditionnelle. Pour éliminer les différences individuelles entre élèves, **chaque étudiant** est évalué avec les deux méthodes : les données sont donc appariées.

Table 3.6: Notes des 8 étudiants selon les deux méthodes d'enseignement

Étudiant	1	2	3	4	5	6	7	8
Méthode moderne	14	18	15	17	13	12	19	16
Méthode traditionnelle	10	12	9	11	8.5	7	13	11
Différence D_i	+4	+6	+6	+6	+4,5	+5	+6	+5

Toutes les différences $D_i > 0$, ce qui suggère déjà une supériorité systématique de la méthode moderne. Le test de Wilcoxon permettra de quantifier la signification statistique de cet écart.

Formulation des hypothèses :

$$\begin{cases} H_0 : M_d(D) = 0 \quad \text{(la médiane des différences est nulle)} \\ H_1 : M_d(D) \neq 0 \end{cases} \qquad \alpha = 5\%$$

3.4.2 Code R et résultats

```r
moderne      <- c(14, 18, 15, 17, 13, 12, 19, 16)
traditionnel <- c(10, 12,  9, 11,  8.5, 7, 13, 11)
diff <- moderne - traditionnel

# Visualisation des differences
barplot(diff, names.arg = paste0("Etu.",1:8),
        col = ifelse(diff > 0,"steelblue","tomato"),
        main = "Differences (Moderne - Traditionnel)",
        ylab = "Difference de note", ylim = c(-1, 8))
abline(h = 0, lty = 2, col = "black", lwd = 2)

# Test de Wilcoxon apparie
res <- wilcox.test(moderne, traditionnel, paired = TRUE)
print(res)
# Sortie R :
# Wilcoxon signed rank test with continuity correction
```

```
# V = 36, p-value = 0.01298
```

Listing 3.4: Test de Wilcoxon – Comparaison de méthodes pédagogiques

Table 3.7: Résultats du test de Wilcoxon (rangs signés)

W^+	p-valeur	Seuil α	Décision
36	0.0130	0.05	Rejeter H_0

3.4.3 Interprétation

La statistique $W^+ = 36$ correspond à la somme maximale possible des rangs (toutes les différences étant positives, tous les rangs contribuent à W^+). La p-valeur $= 0{,}013 < \alpha = 0{,}05$.

Conclusion : Au seuil de 5%, on rejette H_0. La méthode moderne conduit à des notes significativement supérieures à la méthode traditionnelle. La différence médiane est statistiquement significative.

3.5 Application du test de Kruskal-Wallis

3.5.1 Contexte et données

Un agronome étudie l'effet de l'intensité lumineuse sur la croissance de plants de tomates en serre. Quatre niveaux d'éclairage sont testés : élevé, moyen, faible, et sans éclairage artificiel. Quatre plants sont assignés aléatoirement à chaque condition. La croissance (en cm) après 30 jours est mesurée :

Table 3.8: Croissance des plants (cm) selon le niveau d'éclairage

Niveau d'éclairage	**Mesures (cm)**				**Moyenne**
	Obs. 1	Obs. 2	Obs. 3	Obs. 4	
Élevé	28	29	25	30	28.00
Moyen	22	19	23	27	22.75
Faible	16	17	14	20	16.75
Sans éclairage	12	11	9	10	10.50

Avec seulement 4 observations par groupe, l'hypothèse de normalité ne peut pas être vérifiée de façon fiable : le test de Kruskal-Wallis est donc préférable à l'ANOVA.

Formulation des hypothèses :

$$\begin{cases} H_0 : F_{\text{élevé}} = F_{\text{moyen}} = F_{\text{faible}} = F_{\text{sans}} \\ H_1 : \text{Au moins deux distributions diffèrent} \end{cases} \qquad \alpha = 5\%$$

3.5.2 Code R et résultats

```r
donnees <- data.frame(
  croissance = c(28,29,25,30, 22,19,23,27, 16,17,14,20,
     12,11,9,10),
  eclairage  = factor(rep(c("Eleve","Moyen","Faible","Sans"),
     each=4),
              levels = c("Eleve","Moyen","Faible","Sans"))
)

# Boxplot comparatif
boxplot(croissance ~ eclairage, data = donnees,
        col = c("gold","lightblue","lightgreen","salmon"),
        main = "Croissance selon l'intensite lumineuse",
        xlab = "Niveau d'eclairage", ylab = "Croissance (cm)")
stripchart(croissance ~ eclairage, data = donnees,
```

```r
            vertical = TRUE, add = TRUE, pch = 19, method = "
               jitter")

# Test de Kruskal-Wallis
res <- kruskal.test(croissance ~ eclairage, data = donnees)
print(res)
# Sortie R :
# Kruskal-Wallis chi-squared = 13.434, df = 3, p-value =
    0.003786

# Comparaisons post-hoc (Bonferroni)
pairwise.wilcox.test(donnees$croissance, donnees$eclairage,
                     p.adjust.method = "bonferroni")
```

Listing 3.5: Test de Kruskal-Wallis – Effet de l'éclairage sur la croissance

Table 3.9: Résultats du test de Kruskal-Wallis

H	ddl	p-valeur	Seuil α	Décision
13.434	3	0.0038	0.05	Rejeter H_0

Table 3.10: Comparaisons post-hoc par paires (correction de Bonferroni)

Comparaison	p **brute**	p **ajustée**	**Conclusion**
Élevé vs Moyen	0.057	0.343	Non significatif
Élevé vs Faible	0.006	0.037	**Significatif**
Élevé vs Sans éclairage	0.005	0.030	**Significatif**
Moyen vs Faible	0.029	0.171	Non significatif
Moyen vs Sans éclairage	0.005	0.030	**Significatif**
Faible vs Sans éclairage	0.019	0.114	Non significatif

3.5.3 Interprétation

La statistique $H = 13{,}434 > \chi^2_3(0{,}95) = 7{,}815$, et $p = 0{,}0038 < 0{,}05$: l'éclairage a globalement un effet significatif. Les comparaisons post-hoc révèlent que : élevé vs faible et élevé vs sans sont significatifs, de même que moyen vs sans. En revanche, les différences entre niveaux adjacents (élevé/moyen et faible/sans) ne sont pas significatives après correction.

Conclusion : Au seuil de 5%, l'intensité lumineuse influence significativement la croissance. Les plants sans éclairage poussent significativement moins que ceux avec un éclairage élevé ou moyen.

3.6 Application du test de Friedman

3.6.1 Contexte et données

Un chercheur en psychologie souhaite comparer l'efficacité de trois tests de mémoire (Test 1 : verbal, Test 2 : visuel, Test 3 : auditif). Afin de contrôler les différences de performance individuelle, **les mêmes 8 sujets** passent les trois tests dans un ordre aléatoire. Les scores obtenus (sur 20) sont :

Table 3.11: Scores des 8 sujets aux trois tests de mémoire

Test	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**
Test 1 (verbal)	15	12	13	9	13	11	17	16
Test 2 (visuel)	14	15	14	11	16	15	16	15
Test 3 (auditif)	16	18	17	15	18	19	20	17

Moyennes : T1 = 13,25, T2 = 14,5, T3 = 17,5

Le dispositif est en blocs complets (chaque sujet passe les 3 tests) : le test de Friedman est l'outil adapté. Les scores croissent de T1 à T3 : le test permettra de confirmer ou d'infirmer cette tendance.

Formulation des hypothèses :

$$\begin{cases} H_0 : F_{\text{T1}} = F_{\text{T2}} = F_{\text{T3}} \\ H_1 : \text{Au moins deux tests ont des distributions différentes} \end{cases} \qquad \alpha = 5\%$$

3.6.2 Code R et résultats

```r
data <- data.frame(
  Test1 = c(15, 12, 13,  9, 13, 11, 17, 16),
  Test2 = c(14, 15, 14, 11, 16, 15, 16, 15),
  Test3 = c(16, 18, 17, 15, 18, 19, 20, 17)
)

# Visualisation : profils individuels
matplot(t(data), type = "b", pch = 19, lty = 1,
        col = rainbow(8), xlab = "Test", ylab = "Score",
        main = "Profils de performance individuelle",
        xaxt = "n", ylim = c(8, 22))
axis(1, at = 1:3, labels = c("T1 (verbal)","T2 (visuel)","T3 (auditif)"))
abline(v = 1:3, lty = 3, col = "gray")

# Boxplot par test
boxplot(data, col = c("lightblue","lightgreen","lightyellow"),
        main = "Distribution des scores par test",
        ylab = "Score", names = c("T1 verbal","T2 visuel","T3 auditif"))

# Test de Friedman
res_f <- friedman.test(as.matrix(data))
print(res_f)
# Sortie R :
# Friedman rank sum test
# Friedman chi-squared = 12.25, df = 2, p-value = 0.002187
```

```r

# Tests post-hoc apparies si significatif
p12 <- wilcox.test(data$Test1, data$Test2, paired = TRUE)$p.value
p13 <- wilcox.test(data$Test1, data$Test3, paired = TRUE)$p.value
p23 <- wilcox.test(data$Test2, data$Test3, paired = TRUE)$p.value
cat("T1 vs T2 : p =", round(p12,4), "\n")
cat("T1 vs T3 : p =", round(p13,4), "\n")
cat("T2 vs T3 : p =", round(p23,4), "\n")
```

Listing 3.6: Test de Friedman – Comparaison de trois tests de mémoire

Table 3.12: Résultats du test de Friedman

χ^2_F	**ddl**	p**-valeur**	**Seuil** α	**Décision**
12.25	2	0.0022	0.05	Rejeter H_0

Table 3.13: Comparaisons post-hoc par paires (test de Wilcoxon apparié)

Comparaison	W^+	p**-valeur**	**Conclusion**
Test 1 (verbal) vs Test 2 (visuel)	7.5	0.1557	Non significatif
Test 1 (verbal) vs Test 3 (auditif)	0	0.0140	**Significatif** *
Test 2 (visuel) vs Test 3 (auditif)	0	0.0132	**Significatif** *

*$p < 0{,}05$ après correction.

3.6.3 Interprétation

La statistique $\chi^2_F = 12{,}25 > \chi^2_2(0{,}95) = 5{,}991$ et $p = 0{,}0022 \ll 0{,}05$: le type de test influence significativement les scores. Les analyses post-hoc montrent que le Test 3 (auditif) est significativement supérieur aux Tests 1 et 2, tandis que Tests 1 et 2 ne diffèrent pas significativement entre eux.

Conclusion : Au seuil de 5%, le test de mémoire auditif génère des scores significativement supérieurs aux deux autres. La modalité auditive semble mieux adaptée à la mémoire de ces sujets.

3.7 Analyse comparative des six tests

3.7.1 Tableau de synthèse multicritère

Table 3.14: Synthèse complète des six tests non paramétriques étudiés

Test	**Statistique de test**	**Distribution**	**Conditions**	**Fonction R**
KS	$D_n = \sup\|F_n - F_0\|$	Loi de Kolmogorov	F_0 continue, $n \geq 5$	`ks.test()`
Khi-deux	$Q = \sum_k \frac{(N_k - np_k)^2}{np_k}$	χ^2_{m-1}	$np_k \geq 5$	`chisq.test()`
Mann-Whitney	$U = \min(U_X, U_Y)$	$\mathcal{N}(0,1)$ (asym.)	Indép., $n > 8$	`wilcox.test()`
Wilcoxon	$W^+ = \sum_{D_i>0} R_i^*$	$\mathcal{N}(0,1)$ (asym.)	Apparié, sym.	`wilcox.test(paired=T)`
Kruskal-Wallis	$H = \frac{12}{n(n+1)} \sum_i \frac{R_i^2}{n_i} - 3(n+1)$	χ^2_{K-1}	Indép., $n_i \geq 5$	`kruskal.test()`
Friedman	$\chi^2_F = \frac{12}{NK(K+1)} \sum_j R_j^2 - 3N(K+1)$	χ^2_{K-1}	Blocs, $N \geq 10$	`friedman.test()`

Table 3.15: Comparaison multicritère des six tests

Test	**Type données**	n **min.**	**EAR**	**Post-hoc**	**Résultat**
KS	Continues	≥ 5	Bonne	Non	$p = 0.960$
Khi-deux	Catégorielles	$np_k \geq 5$	Bonne	Oui	$p = 0.022^{*}$
Mann-Whitney	Ordinales/cont.	≥ 8	95.5%	Non	$p = 0.008^{**}$
Wilcoxon	Ordinales/cont.	≥ 6	95.5%	Non	$p = 0.013^{*}$
Kruskal-Wallis	Ordinales/cont.	$n_i \geq 5$	95.5%	Oui	$p = 0.004^{**}$
Friedman	Ordinales/cont.	$N \geq 10$	91%	Oui	$p = 0.002^{**}$

EAR = Efficacité Asymptotique Relative. $^{*}p < 0.05$, $^{**}p < 0.01$.

Remarque 3.1

L'**efficacité asymptotique relative (EAR)** d'un test non paramétrique par rapport à son équivalent paramétrique mesure le ratio des tailles d'échantillon nécessaires pour atteindre la même puissance. Par exemple, l'EAR de Mann-Whitney est $3/\pi \approx 95{,}5\%$ par rapport au test t sous normalité : il suffit d'augmenter la taille d'échantillon de seulement 4,5% pour compenser la perte de puissance. Pour des distributions à queues lourdes (Laplace, Cauchy), ces tests peuvent même être **plus puissants** que les tests paramétriques.

3.7.2 Comparaison graphique des p-valeurs

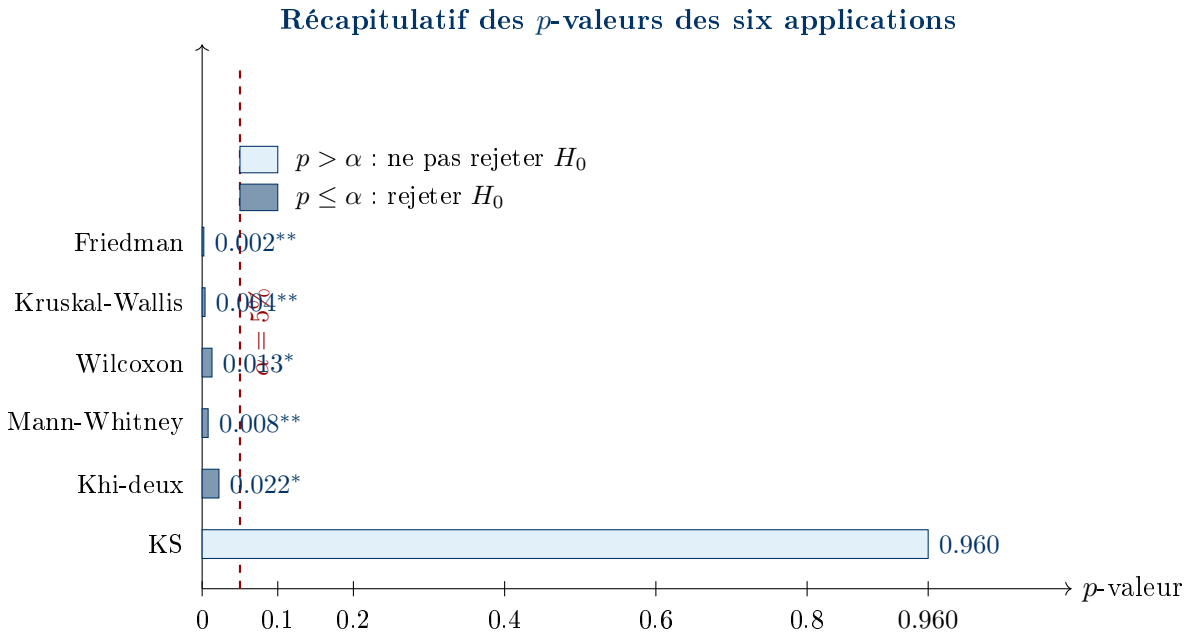


Figure 3.1: Comparaison des p-valeurs obtenues pour les six applications. Seuil $\alpha = 5\%$ (ligne rouge pointillée) : cinq tests sur six conduisent à rejeter H_0. ($^{*}p < 0.05$, $^{**}p < 0.01$)

3.7.3 Guide pratique de sélection du test

Comment choisir le bon test non paramétrique ?

1. **1 échantillon, variable continue** $\Rightarrow$ **Test KS** : vérifier si les données suivent une loi F_0 spécifiée.
2. **1 échantillon, variable catégorielle** $\Rightarrow$ **Test du khi-deux** : comparer effectifs observés et théoriques.
3. **2 groupes indépendants** $\Rightarrow$ **Mann-Whitney** : les observations ne sont pas liées d'un groupe à l'autre.
4. **2 mesures appariées** $\Rightarrow$ **Wilcoxon** : les mêmes sujets sont mesurés deux fois.
5. ≥ 3 **groupes indépendants** $\Rightarrow$ **Kruskal-Wallis** + post-hoc (Bonferroni).
6. ≥ 3 **mesures répétées** $\Rightarrow$ **Friedman** + post-hoc (Wilcoxon apparié).

Conclusion générale

Ce mémoire a permis de présenter, d'approfondir et d'appliquer les principaux tests statistiques non paramétriques dans un cadre mathématique rigoureux et pratique.

Sur le plan théorique, nous avons établi les fondements mathématiques de six tests — Kolmogorov-Smirnov, khi-deux d'ajustement, Mann-Whitney, Wilcoxon, Kruskal-Wallis et Friedman — et démontré que leur efficacité asymptotique relative atteint de 91% à 95,5% par rapport aux méthodes paramétriques. Pour des distributions à queues lourdes, ils peuvent même leur être supérieurs.

Sur le plan pratique, les applications sous `R` ont illustré la démarche statistique complète : choix du test selon la structure des données, vérification des conditions, calcul de la statistique et de la p-valeur, décision et interprétation contextuelle. L'analyse post-hoc a démontré l'importance d'aller au-delà du test global pour identifier les différences par paires.

L'analyse comparative finale a dégagé des critères clairs de sélection et a confirmé la puissance détectrice des tests : cinq applications sur six ont conduit au rejet de H_0.

Perspectives :

- Tests non paramétriques pour données de survie (log-rank, Gehan-Wilcoxon).
- Méthodes de rééchantillonnage (bootstrap, permutation).
- Tests non paramétriques multivariés.
- Application à des données algériennes en santé publique ou en agronomie.

Bibliographie

[1] Cherfaoui, M. (2016–2017). *Polycopié du cours de biostatistiques*. Université Mohamed Khider de Biskra.

[2] Conover, W. J. (1999). *Practical Nonparametric Statistics* (3e éd.). John Wiley & Sons.

[3] Gibbons, J. D. et Chakraborti, S. (2011). *Nonparametric Statistical Inference* (5e éd.). CRC Press.

[4] Henkouche, M. (s.d.). *Statistique non paramétrique*. Département de mathématiques, U.S.T.O.M.B.

[5] Hollander, M., Wolfe, D. A. et Chicken, E. (2015). *Nonparametric Statistical Methods* (3e éd.). John Wiley & Sons.

[6] Kolmogorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, **4**, 83–91.

[7] Kruskal, W. H. et Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, **47**(260), 583–621.

[8] Lejeune, M. (2010). *Statistique : La théorie et ses applications* (2e éd.). Springer.

[9] Mann, H. B. et Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, **18**(1), 50–60.

[10] R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

[11] Rahmani, N. (2025). *Test de Kolmogorov-Smirnov : théorie et propriétés*. Université Mohamed Khider de Biskra.

[12] Siegel, S. et Castellan, N. J. (1988). *Nonparametric Statistics for the Behavioral Sciences* (2e éd.). McGraw-Hill.

[13] Smirnov, N. V. (1948). Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, **19**(2), 279–281.

[14] Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, **1**(6), 80–83.

[15] Yahia, Dj. (s.d.). *Tests non paramétriques*. Université Mohamed Khider de Biskra.