République Algériennne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Mohamed Khider, Biskra

Faculté des Sciences Exactes

Départementde Mathématiques

Mémoire présenté pour obtenir le diplôme de

Master en "**Mathematiques Appliquées**"

Option : **Statistique**

Par

MESBAH Soumia

Titre :

Sur quelques tests statistiques et applications

Devant le Jury :

Pr.	**BENKHELIFA Lazhar**	UMKB	Président
Pr.	**SAYAH Abdallah**	UMKB	Encadrant
Dr.	**BENELMIR Imen**	UMKB	Examinateur

03/06/2025

Dédicace

Tout d'abord, je remercie Dieu, notre créateur, de m'avoir donné la force, la volonté et le courage d'accomplir ce modeste travail.

Je dédie ce travail à ma mère, la source de tendresse et la lumière qui guide mes chemins et me mène vers les voies du succès, pour tous ses sacrifices et ses précieux conseils, pour toute son aide et sa présence dans ma vie.

À mon père, qui m'a beaucoup aidé à atteindre ce niveau de connaissance, j'exprime ma gratitude pour le soutien et l'encouragement qu'il m'a apportés dans l'accomplissement de mon travail.

À mes sœurs.

À mon seul frère.

À mes chères amies.

À tous mes amies.

À tout ce que je sais sans exception.

À tous mes enseignants sans exception, en particulier le Professeur SAYAH Abdallah.

Enfin, je consacre ce mémoire à tous ceux qui ont eu foi en moi et m'ont soutenu, ainsi qu'à toutes les personnes ayant participé à ma trajectoire de réussite. Je vous suis tous profondément reconnaissant

Remerciements

Tout d'abord je tiens à remercier **"Allah"** qui m'a accordé la santé, la force et la patience pour terminer ce travail.

Mes sincères remerciements, mon respect et ma gratitude à mes parents qui ont toujours été là pour moi.

Je remercie mon encadreur : Le **Pr. SAYAH Abdallah**, pour ses efforts et surtout ses conseils précieux et aussi pour sa guidance avec ses encouragements lors des moments de difficulté et sa patience pendant la durée de ce modeste travail.

Mes vifs remerciements vont également aux membres du jury, le **Pr. BENKHELIFA Lazhar**, et le **Dr. BENELMIR Imen** d'avoir accepté d'examiner et d'évaluer ce travail.

Je remercie touts les enseignants du département de Mathématiques.

Je remercie également tous ceux qui étaient à côtés de moi et qui m'ont soutenu dans ma carrière universitaire.

Notations et symbols

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous :

X	:	Variable aléatoire
$X_1, \cdots, X_n$	:	Echantillon de taille n de v.a X
$\overline{X}$	:	Moyenne empirique
σ^2	:	Variance empirique
$\widetilde{S}$	:	L'écart-type
$f(x)$	:	Fonction de densitée
$F(x)$	:	Fonction de répartition
$\mathcal{N}(0,1)$	:	La loi normale centrée réduite.
$\mathcal{N}(\mu, \sigma^2)$	:	Loi normale d'espérance μ et de variance σ^2
χ^2_n	:	Loi de khi-deux à n ddl
$\mathcal{F}(\upsilon_1, \upsilon_2)$	:	Loi de Fisher (υ_1, υ_2) ddl
$\mathcal{T}(\upsilon)$	:	la loi de Student à υ degrés de liberté
$\mathcal{B}(n,p)$	:	Loi binomiale de paramètre et p
$E(X)$	:	Esperance mathématique
$Var(X)$	:	variance mathématique

α	:	Risque de première espèce
β	:	Risque de deuxième espèce
W	:	Région de rejet
$\overline{W}$	:	Région d'acceptation
H_0	:	Hypothèse nulle
H_1	:	Hypothèse alternative
$1-\beta$	:	La puissance du test
π	:	Fonction puissance
θ	:	Paramètre inconnu
Θ_0	:	Ensemble des valeurs de θ
$\mathcal{F}(n_1-1, n_2-1)$	:	Loi de Fisher-Snedecor de (n_1-1, n_2-1) degrés de liberté
Φ	:	La fonction de répartition de la loi normale centrée réduite

Table des matières

Table des figures

Liste des tableaux

Introduction

Le savoir statistique est devenu dans nos sociétés un langage et un outil indispensable dans la plupart des activités professionnelles et même dans la vie courante. Or, la science statistique et ses méthodes suscitent en général maintes difficultés de comprehension, applications inadéquates et biais d'interprétation chez les apprenants et les utilisateurs, comme le decrivent de nombreux auteurs.

William Sealy Gosset, mieux connu sous le nom de Student (1876-1937), a développé un test essentiel en 1908 pour traiter des échantillons réduits lorsque l'écart-type de la population n'est pas connu, et a apporté, grâce à son expérience professionnelle, une contribution critique en termes d'applications.[2]

Abraham Wald (1902-1950) a inventé le concept de tests statistiques séquentiels, ouvrant ainsi un domaine important pour le développement des applications statistiques dans le domaine du contrôle de la qualité industrielle. Eun Pearson (1895-1980) et Jerzy (1894-1981) ont abordé les fondements de la théorie des tests ainsi que Jerzy Neyman.[2]

Les tests statistiques constituent une approche décisionnelle dans les statistiques inférentielles. Il s'agit de décider sur la base d'un échantillon de données si une caractéristique de la population satisfait ou non une certaine spécification que l'on formule sous forme d'hypothèse. L'objet des tests statistiques est de distinguer ce qui est plausible de ce qui est trop peu vraisemblable, il existe deux types de tests : test paramétrique et test non paramétriques.

Notre objectif dans cette recherche est de fournir un aperçu de certains tests statistiques paramétriques en mettant en évidence les bases théoriques de chaque test, en expliquant leurs hypothèses nulles et alternative, ainsi que les statistiques de test et les critères de décision. De plus, nous illustrerons l'application scientifique de ces tests à travers des exemples concrets issus de divers domaines de recherche.

Le travail se compose de trois chapitres comme suit :

Chapitre 1 : Ce chapitre est divisé en deux sections. Dans la première section, nous présenterons quelques concepts et définitions de base de la théorie des probabilités. La deuxième section mentionnera quelques points concernant les tests d'hypothèses (erreur, région de rejet et d'acceptation...).

Chapitre 2 : Ce chapitre se concentre sur l'étude de certains tests statistiques paramétriques, notamment le test de conformité et le test d'homogénéité, avec des exemples illustratifs pour chaque cas, et à la fin de chaque test, les informations sont résumées dans un tableau explicatif.

Chapitre 3 :

Dans ce chapitre, nous nous concentrerons sur l'application de certains tests statistiques paramétriques à des données simulées et réelles, en utilisant le logiciel d'analyse statistique R. Nous conclurons notre travail par conclusion.

Chapitre 1

Généralités sur les tests

Ce chapitre comprend des concepts généraux sur les tests statistiques (formulation d'hypothèses, types de tests,............) comme base utilisés dans divers domaines pour mesurer plusieurs aspects.

1.1 Généralités et principales définitions

1.1.1 Variable statistique

Une variable statistique est un caractère faisant l'objet d'une étude statistique. Elle peut donc être qualitative ou quantitative. Une variable quantitative est appelée :

– Discrète si elle prend un nombre fini ou infini dénombrable de valeurs souvent entières.

– Continue si elle prend toutes les valeurs d'un intervalle fini ou infini.

1.1.2 Individus

Les individus (un client, une région géographique, un animal...) sont les entités de base sur lesquelles on relève un certain nombre de caractéristiques. Les individus peuvent représenter la population entière ou provenir d'un échantillon aléatoire, tiré au hasard dans la population.

1.1.3 Échantillon

Un échantillon est un ensemble d'individus issus d'une population.

1.1.4 Population

Une population est une réunion des individus sur lesquels on étudie une ou plusieurs propriétés. Une population doit être correctement définie afin que l'appartenance d'un individu à cette population soit reconnue sans ambiguïté.

1.1.5 Données

Les données sont les faits et les chiffres qui sont collectés, analysés et résumés pour pouvoir ensuite être interprétés. Toutes les données collectées dans une étude particulière forment l'ensemble de données de l'étude

1.1.6 Moyenne et écart-type corrigé

La moyenne et l'écart-type corrigé des données sont les principales mesures statistiques intervenant en estimation paramétrique. En notant X un caractère numérique, n le nombre d'individus d'un échantillon et $X_1, \cdots, X_n$ les données associées, on définit :

▷ La moyenne de $X_1, \cdots, X_n$:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{1.1}$$

C'est une estimation ponctuelle de la valeur moyenne de X.

▷ L'écart-type corrigé de $X_1, \cdots, X_n$:

$$\widetilde{S} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}. \tag{1.2}$$

C'est une estimation non biaisée de l'écart type

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}. \tag{1.3}$$

C'est une estimation biaisée de l'écart type

1.2 Quelques lois de probabilités continues

1.2.1 Loi normale (ou loi de Gauss)

La loi normale (ou loi gaussienne) de paramètres μ, σ^2 (avec $\sigma > 0$) est la distribution de probabilité définie par la fonction de densité de probabilité :

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

La fonction de répartition F de la loi de gauss n'admet pas d'expression analytique simple. On l'obtient par l'approximation numérique de l'intégrale

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

On parle de loi de gauss centrée réduite si la moyenne est nulle (centrée) et la variance est égale à 1 (réduite).

Pour simplifier l'écriture, on notera parfois $X \sim \mathcal{N}(\mu, \sigma^2)$ pour signifier que la variable aléatoire X suit une loi de gauss de moyenne μ et de variance σ^2.

Remarque 1.2.1 : Si $X \sim \mathcal{N}(\mu, \sigma^2)$ alors sa transformée centrée-réduite $Z = \dfrac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$

Soit $X \sim \mathcal{N}(\mu, \sigma^2)$, alors :

$$P(X \leq x) = P(Z \leq \frac{x-\mu}{\sigma}) \text{ où } Z \sim \mathcal{N}(0,1)$$

soit encore :

$$P(X \leq x) = \Phi(\frac{x-\mu}{\sigma})$$

où Φ est une fonction de répartition de la loi normale centrée-réduite

1.2.2 Loi de χ^2(khi-deux)

La loi de khi-deux (ou loi du khi-carré) à υ degrés de liberté (où υ est un entier positif) est définie par la densité de probabilité :

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{\upsilon}{2}}\Gamma(\frac{\upsilon}{2})} x^{\frac{\upsilon}{2}-1} \exp(-\frac{x}{2}) & \text{si} \quad x \geq 0 \\ 0 & \text{si} \quad x < 0 \end{cases}$$

où $\Gamma(a)$ est la fonction gamma d'euler définie par :

$$\Gamma(a) = \int_{0}^{+\infty} u^{a-1} \exp(-u) du = (a-1)!, \ a \in [0, \ +\infty[$$

Propriétés 1.2.1 :

– Support de définition $x \in [0, +\infty[$

– La loi du khi-deux est asymétrique.

– La moyenne $E(X) = \upsilon$

– La variance $Var(X) = E((X-\upsilon)^2) = 2\upsilon$

Approximation du χ^2 par la loi normale

Lorsque $\upsilon > 30$ on peut admettre que la variable aléatoire $U = \sqrt{2\chi^2_\upsilon} - \sqrt{2\upsilon - 1}$ suit la loi $\mathcal{N}(0,1)$, soit :

$$\chi^2_\upsilon \approx \frac{(U + \sqrt{2\upsilon - 1})^2}{2} \qquad \text{(approximation de Fisher)}$$

Ou (mieux) que :

$$U = \frac{\sqrt{9\upsilon}}{2} \left[\left(\frac{\chi^2_\upsilon}{\upsilon} \right)^{\frac{1}{3}} + \sqrt{\frac{9\upsilon}{2}} - 1 \right]$$

Soit :

$$\chi^2_\upsilon \approx \upsilon \left(U \sqrt{\frac{2}{9\upsilon}} + 1 - \frac{2}{9\upsilon} \right)^3 \qquad \text{(approximation de Wilson-Hilferty)}$$

Cette dernière approximation est très précise et correcte même pour des valeurs faibles de υ. On trouvera en annexe des formules exactes permettant de calculer la fonction de répartition du χ^2_υ.

1.2.3 Loi de Student

Soient Z et Q deux v.a. indépendantes telles que $Z \rightsquigarrow \mathcal{N}(0,1)$ et $Q \rightsquigarrow \chi^2(\upsilon)$. Alors la variable aléatoire :

$$T = \frac{Z}{\sqrt{\frac{Q}{\upsilon}}}$$

Suit une loi appelée loi de Student à υ degrés de liberté, notée $t(\upsilon)$.

La densité de la loi de Student à υ degrés de liberté est :

$$f(x) = \frac{1}{\sqrt{\pi\upsilon}} \frac{\Gamma(\frac{\upsilon+1}{2})}{\Gamma(\frac{\upsilon}{2})} (1 + \frac{x^2}{\upsilon})^{-\frac{\upsilon+1}{2}}, \ x \in \mathbb{R}$$

Propriétés 1.2.2 :

– Moyenne $E(X) = 0$ pour $\upsilon > 1$ et indéfinie pour $\upsilon = 1$

– Variance $Var(X) = E((X-\mu)^2) = \frac{\upsilon}{\upsilon - 2}$ pour $\upsilon > 2$.

1.2.4 Loi de Fisher-Snedecor

Soit U et V deux variables aléatoires indépendantes telles que $U \rightsquigarrow \chi^2(\upsilon_1)$et $V \rightsquigarrow \chi^2(\upsilon_2)$. Alors la variable aléatoire :

$$F = \frac{U / \upsilon_1}{V / \upsilon_2}$$

suit une loi de Fisher-Snedecor à υ_1 degrés de liberté au numérateur et υ_2 degrés de liberté au dénominateur, notée $F(\upsilon_1, \upsilon_2)$, En bref on l'appellera loi de Fisher-Snedecor.

La densité de la loi $F(\upsilon_1, \upsilon_2)$ est :

$$f_{\upsilon_1,\upsilon_1}(x) \ = \frac{\Gamma(\dfrac{\upsilon_1+\upsilon_2}{2})}{\Gamma(\dfrac{\upsilon_1}{2})\Gamma(\dfrac{\upsilon_2}{2})}(\frac{\upsilon_1}{\upsilon_2})^{\dfrac{\upsilon_1}{2}} \frac{x^{\frac{\upsilon_1-2}{2}}}{(1+\frac{\upsilon_1}{\upsilon_2}x)^{\dfrac{\upsilon_1+\upsilon_2}{2}}} \text{ si } x>0$$

Propriétés 1.2.3 :

- Moyenne $E\left(X\right) = \dfrac{\upsilon_2}{\upsilon_2 - 2}$ si $\upsilon_2 > 3$
- Variance $Var(X) = \dfrac{2\upsilon_2^2(\upsilon_1+\upsilon_2-2)}{\upsilon_1(\upsilon_2-2)^2(\upsilon_2-4)}$ si $\upsilon_2 > 5$.

1.2.5 Loi de Fisher

La loi de Fisher permet de modéliser le rapport de deux variables distribuées suivant des lois du khi deux. Soient Z_1 et Z_2 deux variables de loi de khi deux à υ_1 et υ_2 degrés de libertés respectivement et d'écart-types σ_1 et σ_2 alors :

$$\frac{Z_1 / \sigma_1}{Z_2 / \sigma_2}$$

suit une loi de Fisher à (υ_1, υ_2) degrés de liberté.

1.3 Théorème centrale limite

Le théorème central limite est l'un des résultats les plus importants de la théorie des probabilités. De façon informelle, ce théorème donne une estimation très précise de l'erreur que l'on commet en approchant l'espérance mathématique par la moyenne arithmétique. Ce phénomène a d'abord été observé par Gauss qui l'appelait loi des erreurs, mais ce dernier n'en a pas donné de démonstration rigoureuse. La preuve du théorème a été apportée par moivre et laplace, le théorème porte donc parfois leurs noms.

Ce théorème est fondamental car il justifie toutes les approximations par la loi normale.

Théorème 1.3.1 : Si $(X_n)_{n\in\mathbb{N}}$ est une suite de variable aléatoire indépendantes et de même loi $\mathcal{N}(\mu,\sigma^2)$, avec espérance $E(X_i)=\mu$, et variance $Var(X_i)=\sigma^2<\infty$

On définit la moyenne empirique par :

$$\overline{X_n}=\frac{X_1+X_2+\cdots+X_n}{n}.$$

La loi de $\overline{X_n}$, lorsque n est suffisamment grand, est approximativement une $N(\mu,\frac{\sigma^2}{n})$. Plus précisément, si l'on note :

$$Z_n=\sqrt{n}\frac{\overline{X_n}-\mu}{\sigma}\xrightarrow{D}\mathcal{N}(0,1).$$

Est la fonction de répartition de la loi $\mathcal{N}(0,1)$. C'est-à-dire que l'on a :

$$\lim_{n\longrightarrow\infty}P(\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}<x)=\Phi(x).$$

1.4 Test statistique

Définition 1.4.1 : Un test statistique est une procédure de décision permettant de trancher entre deux hypothèses après l'observation d'un échantillon. Les deux hypothèses sont généralement appelées hypothèse nulle et hypothèse alternative

1.5 Hypothèses statistiques

Les hypothèses envisagées a priori s'appellent :

• **L'hypothèse nulle** H_0**.** C'est l'hypothèse selon laquelle on fixe a priori la valeur d'un paramètre.

• **L'hypothèse alternative** H_1. On peut choisir pour cette hypothèse n'importe quelle hypothèse compatible avec le problème étudié, mais différente de H_0. L'hypothèse alternative H_1 peut être :

– **Une Hypothèse bilatéral** si la forme : $H_1 : \theta \neq \theta_0$.

– **Une Hypothèse unilatéral à gauche** si la forme : $H_1 : \theta < \theta_0$.

– **Une Hypothèse unilatéral à droit** si la forme : $H_1 : \theta > \theta_0$.

Pour les tests paramétriques on distingue généralement hypothèses simples et hypothèses composites :

– Une hypothèse simple est du type $H_0 : \theta = \theta_0$, où θ_0 est une valeur isolée du paramètre.

– Une hypothèse composite est du type $H_1 : \theta \in A$, où A est une partie de $\mathbb{R}$ non réduite à un élément.

La plupart dès hypothèses composites se ramènent aux cas : $\theta > \theta_0$, $\theta < \theta_0$, $\theta \neq \theta_0$.

1.5.1 Erreurs et risque

Les hypothèses H_0 et H_1 sont conçues de manière à ce que seulement l'une d'elles soit correcte. Pendant le processus décisionnel qui mènera à la sélection de H_0 ou H_1, quatre scénarios peuvent être envisagés :

– Accepter H_0 et elle est vraie.

– Rejeter H_0 et elle fausse.

– Rejeter H_0 alors qu'elle est vraie.

– Accepter H_0 alors qu'elle est fausse.

Risques de première espèce

Définition 1.5.1 : On appelle risque d'erreur de première espèce la probabilité de rejeter H_0 et d'accepter H_1 alors que H_0 est vraie

$$\alpha = P[\text{rejeter } H_0 \mid H_0 \text{ est vraie}] = P(H_1 \mid H_0)$$

Exemple 1.5.1 : Si l'on cherche à tester l'hypothèse qu'une pièce de monnaie n'est pas « truquée », nous allons adopter la règle de décision suivante :

H_0 : la pièce n'est pas truquée est acceptée si $X \in [40, 60]$, et rejetée si $X \notin [40, 60]$ donc soit $X < 40$ ou $X > 60$ avec X « nombre de faces » obtenus en lançant 100 fois la pièce. Le risque d'erreur de première espèce est :

$$\alpha = 1 - P(\mathcal{B}(100, \frac{1}{2}) \in [40, 60])$$

Risques de deuxième espèce

Définition 1.5.2 : On appelle risque d'erreur de seconde espèce, notée β la probabilité de rejeter H_1 et d'accepter H_0 alors que H_1 est vraie

$$\beta = P[\text{accepter } H_0 \mid H_0 \text{est fausse}] = P(H_0 \mid H_1)$$

Remarque 1.5.1 : Pour évaluer le risque β, il est nécessaire de comprendre la distribution de probabilité de la statistique S selon l'hypothèse H_1.

Exemple 1.5.2 : Si l'on reprend l'exemple précédent de la pièce de monnaie, et que l'on suppose la probabilité d'obtenir face est de 0.6 pour une pièce truquée. En adoptant toujours la même règle de décision suivante :

H_0 : la pièce n'est pas truquée est acceptée si $X \in [40, 60]$, et rejetée si $X \notin [40, 60]$ donc soit $X < 40$ ou $X > 60$ avec X « nombre de faces » obtenues en lançant 100 fois la pièce. Le risque de seconde espèce est :

$$\beta = P(\mathcal{B}(100, \frac{1}{2}) \notin [40, 60]).$$

Soient H_0 et H_1ces deux hypothèses, dont une et une seule est vraie. La décision aboutira à choisir H_0 ou H_1.

Il y'a donc 4 cas possibles schématisés dans le tableau 1.1 avec les probabilités correspondante :

Décision\Réalité	H_0 vrai	H_0 fausse
Accepter H_0	décision correcte $1-\alpha$	erreur de 2ème espèce β
Rejeter H_0	erreur de 1ère espèce α	décision correcte $1-\beta$

TAB. 1.1 – Tableau de décision

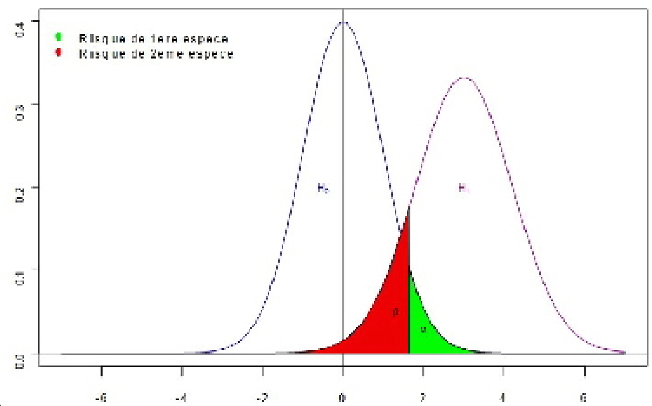
Le graphique ci-dessous représente les risques de première et de deuxième espèce [7] :

Remarque 1.5.1 : Les seuils de signification les plus utilisés sont $\alpha = 0.10,\ 0.05,\ 0.01$.

1.5.2 Variable de décision

Définition 1.5.3 : La variable de décision est une variable qui doit fournir le plus d'informations possible sur la question soulevée et dont la distribution diffèrera en fonction du fait que H_0 ou H_1 soit vrai (sinon, elle serait sans valeur). Il est

01



2.jpg

FIG. 1.1 – Risques de première et deuxième espèce

essentiel que sa loi soit entièrement révélée, au moins si H_0 est valide.

1.5.3 La puissance d'un test

Définition 1.5.4 : La puissance d'un test est la probabilité de prendre une décision correcte si l'hypothèse alternative H_1 est vraie. De la même manière, la puissance d'un test est la probabilité de rejeter l'hypothèse nulle H_0 quand H_1 est vraie.

$$\pi = P[\text{rejet } H_0 \mid H_1 \text{ est vraie}] = 1 - \beta$$

1.5.4 Région de rejet et région critique

Définition 1.5.5 : La région critique notée W est l'ensemble des valeurs de la variable de décision qui conduisent à écarter H_0 au profit de H_1. La forme de la critique est déterminée par la nature de H_1, sa détermination exacte se fait en écrivant que :

$$P(W \mid H_0) = \alpha$$

La région d'acceptation est son complémemaire $\overline{W}$ et l'on a donc :

$$P(\overline{W} \mid H_0) = 1 - \alpha \quad \text{et} \quad P(\overline{W} \mid H_1) = 1 - \beta$$

1.5.5 Règle de décision

Il existe deux stratégies pour prendre une décision en ce qui concerne un test d'hypothèse : la première stratégie fixe à priori la valeur du seuil de signification α et la seconde établit la valeur de la probabilité critique α_{obs} à posteriori.

Règle de décision 1 :

Sous l'hypothèse «H_0 est vraie » et pour un seuil de signification α fixé :

- Si la valeur de la statistique S_{obs} calculée appartient à la région critique alors l'hypothèse H_0 est rejetée au risque d'erreur α et l'hypothèse H_1 est acceptée.

- Si la valeur de la statistique S_{obs} n'appartient pas à la région critique alors l'hypothèse H_0 ne peut être rejetée.

Règle de décision 2 :

La probabilité critique α telle que $P(S \geq S_{obs}) = \alpha_{obs}$ est évaluée

- Si $\alpha_{obs} \geq \alpha$ l'hypothèse H_0 est acceptée car le risque d'erreur de rejeter H_0 alors qu'elle est vrai est trop important.

- Si $\alpha_{obs} \leq \alpha$ l'hypothèse H_0 est rejetée car le risque d'erreur de rejeter H_0 alors qu'elle est vrai est très faible.

1.5.6 La p-valeur

Définition 1.5.6 : La valeur p est la probabilité qui fournit une mesure des preuves fournies par l'échantillon contre l'hypothèse nulle. Plus les valeurs p sont petites, plus les preuves contre H_0 sont fortes.

– Si p-valeur $> \alpha$: on accepte H_0

– Si p-valeur $\leq \alpha$: on rejette H_0

1.5.7 Test uniformément le plus puissant

Définition 1.5.7 : Un test est dit uniformément le plus puissant (UPP) si, quelle que soit la valeur de θ appartenant à l'hypothèse alternative, sa puissance $1-\beta(\theta)$ est supérieure à la puissance de tout autre test.

1.5.8 Démarche d'un test

Démarche de test suit généralement les étapes suivantes :

1. Choix de H_0 et H_1
2. Détermination de la variable de décision.
3. Allure de la région critique en fonction de H_1.
4. Calcul de la région critique en fonction de α.
5. Calcul éventuel de la puissance $1 - \beta$.
6. Calcul de la valeur expérimentale de la variable de décision.
7. Conclusion : rejet ou acceptation de H_0.

1.6 Les tests paramétriques et non paramétriques

1.6.1 Test paramétrique

On parle de tests paramétriques lorsque l'on stipule que les données sont issues d'une distribution paramétrée. Dans ce cas, les caractéristiques des données peuvent être résumées à l'aide de paramètres estimés sur l'échantillon, la procédure de test subséquente ne porte alors que sur ces paramètre sonnées

Il existe plusieurs types de tests paramétriques, dont les suivants :

Tests de conformité :

Ces tests permettent de vérifier si une variable (moyenne, variance, covariance, proportion, etc.) dans une population dont la distribution est connue est égale à une certaine valeur théorique appelée valeur hypothétique du coefficient.

Tests de comparaison (homogénéité) :

Ces tests sont utilisés pour déterminer si les paramètres de deux (ou plusieurs) populations dont les distributions sont connues sont identiques. Par exemple, le test de comparaison des moyennes

1.6.2 Test non paramétrique

Les tests non paramétriques ne font aucune hypothèse sur la distribution sous-jacente des données. On les qualifie souvent de tests « distribution free ». L'étape préalable consistant à estimer les paramètres des distributions avant de procéder au test d'hypothèse proprement dit n'est plus nécessaire.

On identifie plusieurs types de tests non paramétriques, dont les suivants :

Tests d'ajustement :

Ils sont utilisés pour déterminer si un échantillon appartient à une population de distribution spécifiée.

Tests d'indépendance :

Ils sont utilisés pour déterminer si plusieurs populations sont indépendantes les unes des autres.

Chapitre 2

Tests paramétriques

Dans ce chapitre, nous parlerons d'un ensemble important de tests connus sous le nom de tests statistiques paramétriques qui représentent un pilier essentiel pour tirer des conclusions significatives à partir de données quantitatives, mais qui dépendent d'un ensemble d'hypothèses de base sur la nature des données étudiées. Nous nous pencherons sur la compréhension de ces tests, explorerons les hypothèses sur lesquelles ils sont basés, et découvrirons leurs différents types et leurs applications pratiques dans divers domaines.

2.1 Test de conformité

2.1.1 Test de la moyenne d'une population normale

On suppose que l'on a un échantillon $(X_1, \cdots, X_n)$ qui suit une loi normale $\mathcal{N}(\mu, \sigma^2)$, le test statistique de la moyenne empirique

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{2.1}$$

Variance connue :

1. **Test bilatérale :**

$$\begin{cases} H_0 : \mu = \mu_0. \\ H_1 : \mu \neq \mu_0. \end{cases}$$

Sous l'hypothèse H_0 la variable aléatoire suit une loi $\mathcal{N}(\mu, \sigma^2)$ et par conséquent la statistique

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \rightsquigarrow \mathcal{N}(0, 1). \tag{2.2}$$

Pour un risque d'erreur α fixé on a donc:

$$P(|Z| \geq q_{1-\frac{\alpha}{2}}) = \alpha.$$

Puisque la distribution de Z est symétrique alors :

$$P(Z \leq q_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2},$$

où

$$q_{1-\frac{\alpha}{2}} = \Phi^{-1}(1 - \frac{\alpha}{2}).$$

La région critique du test est :

$$W := \left\{ (X_1, \cdots, X_n) \in \mathbb{R}^n : \left| \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \right| \geq q_{1-\frac{\alpha}{2}} \right\}.$$

On rejette H_0 au risque α si $Z \notin] - q_{1-\frac{\alpha}{2}}; q_{1-\frac{\alpha}{2}}[$.

2. **Test unilatérale à droite :**

$$\begin{cases} H_0 : \mu = \mu_0. \\ H_1 : \mu > \mu_0. \end{cases}$$

La statistique Z est définie comme suit :

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Pour un risque d'erreur α fixé on a donc :

$$P(Z \leq q_{1-\alpha}) = 1 - \alpha,$$

où

$$\Phi^1(1-\alpha) = q_{1-\alpha},$$

la région critique du test est :

$$W := \left\{ (X_1, \cdots, X_n) \in \mathbb{R}^n : \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \geq q_{1-\alpha} \right\}$$

On rejette H_0 au risque α si $Z \notin] -\infty; q_{1-\alpha}[$

Exemple 2.1.1. : Un contrôle anti-dopage a été effectué sur 16 sportifs. On a mesuré la variable X de moyenne μ, qui est le taux (dans le sang) d'une certaine substance interdite. Voici les données obtenues :

0.36	0.7	0.61	0.42	0.18	0.54	0.73	0.31
0.24	0.46	0.12	0.72	0.26	0.15	0.75	0.19

La variable X est supposée gaussienne et de variance $\sigma^2 = 0.04$. On veut tester, au $\alpha = 0.05$ l'hypothèse selon laquelle le taux moyen dans le sang de la population des sportifs est égal à 0.4.

Les hypothèses à tester sont : $\begin{cases} H_0 : \mu = 0.4 \\ H_1 : \mu > 0.4 \end{cases}$

$\overline{X} = 0.422,\ n = 16,\ \sigma = \sqrt{0.04}$

Statistique de test 2.2 est : $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{0.422 - 0.4}{\sqrt{\frac{0.04}{16}}} = 0.44$

D'après le tableau de loi $\mathcal{N}(0,\, 1)$ nous trouvons : $q_{1-\alpha} = 1.96$.

$Z = 0.44 < 1.96$, donc on rejette H_0.

3. **Test unilatérale à gauche :**

$$\begin{cases} H_0 : \mu = \mu_0. \\ H_1 : \mu < \mu_0. \end{cases}$$

La statistique Z est définie comme suit :

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Pour un risque d'erreur α fixé on a donc :

$$P(Z \leq q_\alpha) = \alpha,$$

où

$$q_\alpha = \Phi^{-1}(\alpha).$$

La région critique du test est :

$$W := \left\{(X_1, \ \cdots, \ X_n) \in \mathbb{R}^n : \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \leq q_\alpha \right\}.$$

On rejette H_0 au risque α si $Z \notin]q_\alpha; +\infty[$.

Remarque 2.1.1 : Si la variable parente ne suit pas une loi de Gauss, les tests précédents s'appliquent encore dès que n est assez grand ($n > 30$ environ) en raison du théorème centrale-limite.

Variance inconnue :

Comme la variance est inconnue, on l'estime par la variance empirique

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2. \tag{2.3}$$

La statistique T est définie par :

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n-1}} \rightsquigarrow \mathcal{T}_{n-1} \tag{2.4}$$

1. **Test bilatérale** :

On veut tester

$$\begin{cases} H_0 : \mu = \mu_0. \\ H_1 : \mu \neq \mu_0. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(|T| \geq t_{1-\frac{\alpha}{2}}) = \alpha,$$

puisque la distribution de T est symétrique alors :

$$P(T \leq t_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2},$$

où

$$t_{1-\frac{\alpha}{2}} = F^{-1}(1 - \frac{\alpha}{2}).$$

La région critique du test est :

$$W := \left\{ (X_1, \cdots, X_n) \in \mathbb{R}^n : \left| \frac{\overline{X} - \mu_0}{S/\sqrt{n-1}} \right| \geq t_{1-\frac{\alpha}{2}} \right\}.$$

On rejette H_0 au risque α si $T \notin]-t_{1-\frac{\alpha}{2}}; t_{1-\frac{\alpha}{2}}[$.

2. **Test unilatérale à droite :**

$$\begin{cases} H_0 : \mu = \mu_0. \\ H_1 : \mu > \mu_0. \end{cases}$$

Cet par conséquent la statistique

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n-1}},$$

pour un risque d'erreur α fixé on a donc :

$$P(T \geq t_{1-\alpha}) = \alpha,$$

où

$$t_{1-\alpha} = F^{-1}(1 - \alpha).$$

La région critique du test est :

$$W := \left\{ (X_1, \cdots, X_n) \in \mathbb{R}^n : \frac{\overline{X} - \mu_0}{S/\sqrt{n-1}} \geq t_{1-\alpha} \right\}.$$

On rejette H_0 au risque α si $T \notin]-\infty; t_{1-\alpha}[$.

Exemple 2.1.2 : Une usine fabrique un certain type de récipient en plastique. On cherche à montrer, avec un faible risque de se tromper, que le contenu moyen d'un récipient est strictement supérieur à 10 litres. Le contenu de 12 récipients choisis au hasard dans la production est mesuré. Les résultats, en litres, sont :

9.5	8.9	9.2	9.3	9.4	8.8	8.9	9.4	9.2	8.1	9.3	8.6

On suppose que le contenu en litres d'un récipient de cet usine peut être modélisé par une var X suivant une loi normale. ($\alpha = 0.05$). Proposer un test statistique adapté et conclure. Nous effectuons un test d'hypothèse pour la moyenne : $\begin{cases} H_0 : \mu = 9 \\ H_1 : \mu > 9 \end{cases}$

Moyenne empirique : $\overline{X} = 9.05$, $n = 12$

Variance empirique $S^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2 = 0.1525$.

Statistique de test 2.4 est :$T = \dfrac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n-1}}} = \dfrac{9.05 - 9}{\sqrt{\frac{0.1525}{11}}} = 0.4246$.

Dans la table de student à 11 degré de liberté $t_{n-1} = t_{11} = 2.201$.

Décision :

$t_{11} = 2.201 > 0.4246$, donc on accepte H_0.

Sur la base des données de cet échantillon, nous ne pouvons pas conclure que le contenu moyen d'un récipient est strictement supérieur à 10 litres.

3. **Test unilatérale à gauche :**

$$\begin{cases} H_0 : \mu = \mu_0. \\ H_1 : \mu < \mu_0. \end{cases}$$

La statistique T est définie comme suit :

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n-1}}.$$

Pour un risque d'erreur α fixé on a donc :

$$P(T \leq t_\alpha) = \alpha,$$

où

$$t_\alpha = F^{-1}(\alpha).$$

La région critique du test est :

$$W := \left\{ (X_1, \ \cdots \ , \ X_n) \in \mathbb{R}^n : \frac{\overline{X} - \mu_0}{S/\sqrt{n-1}} \leq t_\alpha \right\}.$$

On rejette H_0 au risque α si $T \notin]t_\alpha; +\infty[$.

Exemple 2.1.3 :

Avec le médicament A, la durée moyenne de disparition de la douleur était $30mn$. On a administré le médicament B à 12 malades et relevé les durées de disparition de la douleur suivants ($\alpha = 0.05$) :

23	26	22	32	19	29	36	28	25	30	31	20

Nous effectuons un test d'hypothèse pour la moyenne : $\begin{cases} H_0 : \mu = 30 \\ H_1 : \mu < 30 \end{cases}$

$\overline{X} = 26.75$, $S = 4.95$

La statistique de test est : $T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n-1}} = \dfrac{26.75 - 30}{\frac{4.95}{\sqrt{11}}} = -2.17$.

Dans la table de student à 11 degré de liberté $t_{n-1} = t_{11} = 2.201$.

Décision :

$T = -2.17 < 2.201$, donc on rejette H_0.

Il existe une différence statistiquement significative entre le médicament B et le médicament A.

Ce tableau 2.1 résume les informations ci-dessus :

Paramètres	Hypothèse	Statistique de test	Région critique
Variance connue	$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$	$Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \rightsquigarrow \mathcal{N}(0,1)$	$\lvert Z\rvert \geq q_{1-\frac{\alpha}{2}}$
	$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$		$Z \geq q_\alpha$
	$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$		$Z \leq q_{1-\alpha}$
Variance inconnue	$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$	$T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n-1}} \rightsquigarrow \mathcal{T}_{n-1}$	$\lvert T\rvert \geq t_{1-\frac{\alpha}{2}}$
	$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$		$T \geq t_\alpha$
	$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$		$T \leq t_{1-\alpha}$

TAB. 2.1 – Tableau de test de moyenne

2.1.2 *Test de la variance d'une population* $\mathcal{N}(\mu, \sigma^2)$

Moyenne connue :

Soit une variable aléatoire X de loi normale $\mathcal{N}(\mu, \sigma^2)$, où μ est connu. Soit tout échantillon de n valeurs indépendantes $(n > 1)$, $x_1, \cdots, x_n$ de cette variable X, L'estimateur de la variance sera :

$$V^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

La statistique Z est définie comme suit :

$$Z = \frac{nV^2}{\sigma_0^2} = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma_0}\right)^2 \rightsquigarrow \chi_n^2 \tag{2.5}$$

1. **Test bilatérale :**

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2. \\ H_1 : \sigma^2 \neq \sigma_0^2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc (en choisissant un intervalle symétrique) :

$$P_{\sigma^2=\sigma_0^2}\left(\frac{nV^2}{\sigma_0^2} \leq k_{\frac{\alpha}{2}} \text{ ou } \frac{nV^2}{\sigma_0^2} \geq k_{1-\frac{\alpha}{2}}\right) = \alpha,$$

où

$$P_{\sigma^2=\sigma_0^2}\left(\frac{nV^2}{\sigma_0^2} \geq k_{1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2},$$

$$F^{-1}\left(1 - \frac{\alpha}{2}\right) = k_{1-\frac{\alpha}{2}},$$

où

$$P_{\sigma^2=\sigma_0^2}\left(\frac{nV^2}{\sigma_0^2} \leq k_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}.$$

$$F^{-1}\left(\frac{\alpha}{2}\right) = k_{\frac{\alpha}{2}}.$$

La région critique du test est :

$$W = \left\{ (x_1, \cdots, x_n) \in \mathbb{R}^n \mid \frac{nV^2}{\sigma_0^2} \leq k_{\frac{\alpha}{2}} \text{ ou } \frac{nV^2}{\sigma_0^2} \geq k_{1-\frac{\alpha}{2}} \right\}.$$

On rejette H_0 au risque α si $T \notin]k_{\frac{\alpha}{2}}; k_{1-\frac{\alpha}{2}}[$.

2. **Test unilatérale à droite :**

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2. \\ H_1 : \sigma^2 > \sigma_0^2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P\left(\frac{nV^2}{\sigma_0^2} \geq k_{1-\alpha}\right) = \alpha,$$

où

$$k_{1-\alpha} = F^{-1}(1-\alpha).$$

La région critique du test est :

$$W = \left\{ (x_1, \cdots, x_n) \in \mathbb{R}^n \mid \quad \frac{nV^2}{\sigma_0^2} \ \geq \ k_{1-\alpha} \right\}.$$

On rejette H_0 au risque α si $T \notin]0; k_{1-\alpha}[$.

3. **Test unilatérale à gauche :**

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2. \\ H_1 : \sigma^2 < \sigma_0^2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$\alpha = P(\frac{nV^2}{\sigma_0^2} \leq k_\alpha),$$

où

$$k_\alpha = F^{-1}(\alpha).$$

La région critique du test est :

$$W = \left\{ (x_1, \cdots, x_n) \in \mathbb{R}^n \mid \qquad \frac{nV^2}{\sigma_0^2} \ \leq \ k_\alpha \right\}.$$

On rejette H_0 au risque α si $T \notin]k_\alpha; +\infty[$.

Moyennes inconnues :

Soit une variable aléatoire X de loi normale $\mathcal{N}(\mu,\, \sigma^2)$, où μ est inconnue. Soit tout échantillon de n valeurs indépendantes ($n > 1$), $X_1, \cdots, X_n$ de cette variable X, de variance estimée

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

La statistique K est définie comme suit :

$$K = \frac{n\widetilde{S}^2}{\sigma_0^2} \rightsquigarrow \chi^2_{n-1} \tag{2.6}$$

où χ^2_{n-1} est une loi de khi-deux à $n-1$ degrés de libertés.

1. **Test bilatérale :**

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2. \\ H_1 : \sigma^2 \neq \sigma_0^2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P_{\sigma^2=\sigma_0^2}\left(\frac{n\widetilde{S}^2}{\sigma_0^2} \leq l_{\frac{\alpha}{2}} \text{ ou } \frac{n\widetilde{S}^2}{\sigma_0^2} \geq l_{1-\frac{\alpha}{2}}\right) = \alpha,$$

où

$$P\left(\frac{n\widetilde{S}^2}{\sigma_0^2} \leq l_{1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2},$$

où

$$l_{1-\frac{\alpha}{2}} = F^{-1}(1-\tfrac{\alpha}{2}).$$

$$l_{\frac{\alpha}{2}} = F^{-1}(\tfrac{\alpha}{2}).$$

La région critique du test est :

$$W = \left\{(X_1, \cdots, X_n) \in \mathbb{R}^n \mid \frac{n\widetilde{S}^2}{\sigma_0^2} \leq l_{\frac{\alpha}{2}} \text{ ou } \frac{n\widetilde{S}^2}{\sigma_0^2} \geq l_{1-\frac{\alpha}{2}}\right\}.$$

On rejette H_0 au risque α si $K \notin]l_{\frac{\alpha}{2}}; l_{1-\frac{\alpha}{2}}[$.

2. **Test unilatérale à droite :**

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2. \\ H_1 : \sigma^2 > \sigma_0^2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P\left(\frac{n\widetilde{S}^2}{\sigma_0^2} \geq l_{1-\alpha}\right) = \alpha,$$

où

$$l_{1-\alpha} = F^{-1}(1-\alpha).$$

La région critique du test est :

$$W = \left\{ (X_1, \cdots, X_n) \in \mathbb{R}^n \mid \qquad \frac{n\widetilde{S}^2}{\sigma_0^2} \geq l_{1-\alpha} \right\}.$$

On rejette H_0 au risque α si $K \notin]0; l_{1-\alpha}[$.

3. **Test unilatérale à gauche :**

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2. \\ H_1 : \sigma^2 < \sigma_0^2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$\alpha = P(\frac{n\widetilde{S}^2}{\sigma_0^2}) \leq l_\alpha,$$

où

$$l_\alpha = F^{-1}(\alpha).$$

La région critique du test est :

$$W = \left\{ (X_1, \cdots, X_n) \in \mathbb{R}^n \mid \qquad \frac{n\widetilde{S}^2}{\sigma_0^2} \leq l_\alpha \right\}.$$

On rejette H_0 au risque α si $K \notin]l_\alpha; +\infty[$.

Remarque 2.1.2 : Les tests précédents utilisant la loi du χ_n^2 ne sont valables que dans le cas où X suit une loi de gauss.

Exemple 2.1.5 :

Une technique de dosage de sels nutritifs permet de fabriquer des échantillons calibrés d'eau de mer avec un écart-type de 8. Un nouveau procédé de fabrication sera adopté s'il assure une réduction substantielle de la variabilité. Dix mesures sont réalisées sur des échantillons fabriqués avec la nouvelle méthode : ($\alpha = 0.05$)

722	725	724	719	721	731	718	728	723	729
725	722	727	718	723	731	719	724	726	725

Peut-on adopter la nouvelle technique ?

Les hypothèses à tester sont : $\begin{cases} H_0 : \sigma^2 = 64 \\ H_1 : \sigma^2 < 64 \end{cases}$

$n_1 = 10$, $\overline{X} = 724$, $\widetilde{S}^2 = 18.44$, $\sigma^2 = 64$

Statistique de test est : $\dfrac{n\widetilde{S}^2}{\sigma_0^2} = 2.88$

D'après le tableau de loi khi-deux à 9 degrés de libertés nous trouvons : $\chi^2_{0.05,9}$ $= 3.325$.

Décision :

$\chi^2_{0.05,9} = 3.325 > 2.88$, donc on ne rejette pas H_0,

la nouvelle technique de fabrication réduit considérablement la variabilité des échantillons d'eau de mer.

Ce tableau 2.2 résume les informations ci-dessus :

<table>
<tr><th>Paramètres</th><th>Hypothèse</th><th>Statistique de test</th><th>Région critique</th></tr>
<tr><td rowspan="3">μ connue</td><td>$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases}$</td><td rowspan="3">$Z = \frac{nV^2}{\sigma_0^2} = \sum_{i=1}^{n} (\frac{x_i - \mu}{\sigma_0})^2 \rightsquigarrow \chi^2_n$</td><td>$\frac{nV^2}{\sigma_0^2} \leq k_{\frac{\alpha}{2}}$ ou $\frac{nV^2}{\sigma_0^2} \geq k_{1-\frac{\alpha}{2}}$</td></tr>
<tr><td>$\begin{cases} H_0 : \sigma^2 \geq \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{cases}$</td><td>$\frac{nV^2}{\sigma_0^2} \leq k_{\alpha}$</td></tr>
<tr><td>$\begin{cases} H_0 : \sigma^2 \geq \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases}$</td><td>$\frac{nV^2}{\sigma_0^2} \geq k_{1-\alpha}$</td></tr>
<tr><td rowspan="3">μ inconnue</td><td>$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases}$</td><td rowspan="3">$K = \frac{n\widetilde{S}^2}{\sigma_0^2} \rightsquigarrow \chi^2_{n-1}$</td><td>$\frac{n\widetilde{S}^2}{\sigma_0^2} \leq l_{\frac{\alpha}{2}}$ ou $\frac{n\widetilde{S}^2}{\sigma_0^2} \geq l_{1-\frac{\alpha}{2}}$</td></tr>
<tr><td>$\begin{cases} H_0 : \sigma^2 \geq \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{cases}$</td><td>$\frac{n\widetilde{S}^2}{\sigma_0^2} \leq l_{\alpha}$</td></tr>
<tr><td>$\begin{cases} H_0 : \sigma^2 \geq \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases}$</td><td>$\frac{n\widetilde{S}^2}{\sigma_0^2} \geq l_{1-\alpha}$</td></tr>
</table>

Tab. 2.2 – Tableau de test de variance

2.1.3 Test d'une proportion

On dispose d'une population dans laquelle chaque individu présente ou non un certain caractère, la proportion d'individus présentant le caractère étant notée p, $X_1, \cdots, X_n$ est un échantillon aléatoire de taille n extrait de cette population.

On sait que si $X_1, \cdots, X_n$ sont indépendantes et de même loi de Bernoulli $\mathcal{B}(p)$, la proportion $\sum_{i=1}^{n} X_i$ calculée à partir de l'échantillon est considérée comme une réalisation d'une variable aléatoire de loi binomiale $\mathcal{B}(n;p)$ qu'on peut assimiler, si n est assez grand, à une loi normale $\mathcal{N}(p; \sqrt{\frac{p(1-p)}{n}})$.

La statistique Z est définie comme suit :

$$Z = \frac{\sum_{i=1}^{n} X_i - p}{\sqrt{\frac{p(1-p)}{n}}} \rightsquigarrow \mathcal{N}(0,1) \tag{2.7}$$

1. **Test bilatérale :**

$$\begin{cases} H_0 : p = p_0. \\ H_1 : p \neq p_0. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(|Z| \geq q_{1-\frac{\alpha}{2}}) = \alpha,$$

où

$$q_{1-\frac{\alpha}{2}} = \Phi^{-1}(1 - \frac{\alpha}{2}).$$

La région critique du test est :

$$W = \left\{(X_1, \cdots, X_n) \in \mathbb{R}^n \mid \quad |Z| \leq q_{1-\frac{\alpha}{2}}\right\}.$$

La région d'acceptation est un intervalle de la forme : $]-\infty;\ q_{\frac{\alpha}{2}}[\cup]q_{1-\frac{\alpha}{2}};\ +\infty[$.

2. **Test unilatérale à droite**

$$\begin{cases} H_0 : p = p_0. \\ H_1 : p > p_0. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$\alpha = P(Z \geq q_{1-\alpha}),$$

où

$$q_{1-\alpha} = \Phi^{-1}(1-\alpha).$$

La région critique du test est :

$$W = \left\{(X_1,\ \cdots,\ X_n) \in \mathbb{R}^n \mid \qquad Z > q_{1-\alpha}\right\}.$$

La région d'acceptation est un intervalle de la forme : $]q_{1-\alpha};\ +\infty[$

Exemple 2.1.6 :[5] Du sondage du chapitre 4, on a interrogé $n = 800$ personnes et $\sum_{i=1}^{n} X_i = 420$ d'entre elles ont déclaré vouloir voter pour A. On a donc estimé le pourcentage p de voix qu'obtiendra le candidat A par $\hat{p}_n = \frac{420}{800} = 52.5\%$. Mais on a vu qu'un intervalle de confiance de seuil 5% pour ce pourcentage est $[49\%, 56\%]$, dont une partie est située sous les 50%.

Les hypothèses à tester sont : $\begin{cases} H_0 : p = 0.5 \\ H_1 : p > 0.5 \end{cases}$

Statistique de test 2.7 est :$Z = \dfrac{\sum_{i=1}^{n} X_i - p}{\sqrt{\frac{p(1-p)}{n}}} = \dfrac{420 - 0.5}{\sqrt{\frac{0.5\times0.5}{800}}} = 1.414$.

D'après le tableau de loi $\mathcal{N}(0,\,1)$ nous trouvons : $q_{1-\alpha} = 1.645$.

Décision :

$Z < q_{1-\alpha}$, donc on accepte H_0.

On ne peut pas affirmer que A sera élu avec moins de 5% de chances de se tromper.

3. **Test unilatérale à gauche :**

$$\begin{cases} H_0 : p = p_0. \\ H_1 : p < p_0. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$\alpha = P(Z \leq q_\alpha),$$

où

$$q_\alpha = \Phi^{-1}(\alpha).$$

La région critique du test est :

$$W = \{(X_1,\ \cdots,\ X_n) \in \mathbb{R}^n \mid Z < q_\alpha\}.$$

La région d'acceptation est un intervalle de la forme : $]-\infty; q_\alpha[$.

Exemple 2.1.7 : En 1972, 38% des algériens allaient régulièrement (au moins une fois par semaine) au cinéma. Une enquête réalisée en 1972 indique que sur 900 personnes interrogées, 400 d'entre elles vont régulièrement au cinéma. Y a-t-il une baisse significative de la fréquentation des salles de cinéma algériens entre 1972 et

1982 ? Soit p la proportion de algériens allant régulièrement au cinéma en 1982. ($\alpha = 0.05$)

Les hypothèses à tester sont : $\begin{cases} H_0 : p = 0.38 \\ H_1 : p < 0.38 \end{cases}$

$n = 900$, $\sum_{i=1}^{n} X_i = \frac{400}{900} = 0.44$

Statistique de test 2.7 est : $Z = \frac{\sum_{i=1}^{n} X_i - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.44 - 0.38}{\sqrt{\frac{0.38(1-0.38)}{900}}} = 3.7.$

D'après le tableau de loi $\mathcal{N}(0,1)$ nous trouvons : $q_\alpha = -1.645$.

Décision :

$Z = 3.7 > q_\alpha$, donc on accepte H_0.

Il n'y a pas suffisamment de preuves statistiques pour conclure à une baisse significative de la fréquentation des salles de cinéma Algériens entre 1972 et 1982.

Ce tableau2.3 résume les informations ci-dessus :

Hypothèse	Statistique de test	Région critique
$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}$	$Z = \frac{\sum_{i=1}^{n} X_i - p}{\sqrt{\frac{p(1-p)}{n}}} \rightsquigarrow \mathcal{N}(0,1)$	$\|Z\| \geq q_{1-\frac{\alpha}{2}}$
$\begin{cases} H_0 : p \leq p_0 \\ H_1 : p < p_0 \end{cases}$		$Z < q_\alpha$
$\begin{cases} H_0 : p \leq p_0 \\ H_1 : p > p_0 \end{cases}$	$n \geq 30$	$Z > q_{1-\alpha}$

TAB. 2.3 – Tableau de test d'une proportion :

2.2 Test de comparaison

2.2.1 Test de comparaison de deux moyennes

Soient X_1 et X_2 deux variables aléatoires indépendantes de lois normales $X_1 \rightsquigarrow \mathcal{N}(\mu_1, \sigma_1^2),\ X_2 \rightsquigarrow \mathcal{N}(\mu_2, \sigma_2^2)$. On dispose de deux échantillons indépendants $(X_1^{(1)}, \cdots, X_{n_1}^{(1)}),\ (X_1^{(2)}, \cdots, X_{n_2}^{(2)})$ de tailles n_1 et n_2 respectivement, tels que $X_i^{(1)}$ (respectivement $X_j^{(2)}$) pour $1 \leq i \leq n_1,\ 1 \leq j \leq n_2$ suit la même loi que X_1 (respectivement. X_2).

σ_1^2 et σ_2^2 sont connues :

La variable aléatoire :

$$\overline{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i^{(1)} \rightsquigarrow \mathcal{N}(\mu_1, \frac{\sigma_1^2}{n_1}),$$

et :

$$\overline{X}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} X_j^{(2)} \rightsquigarrow \mathcal{N}(\mu_2, \frac{\sigma_2^2}{n_2}).$$

La statistique Z est définie comme suit :

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightsquigarrow \mathcal{N}(0, 1). \tag{2.8}$$

1. **Test bilatérale :**

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightsquigarrow \mathcal{N}(0, 1).$$

$$\begin{cases} H_0 : \mu_1 = \mu_2. \\ H_1 : \mu_1 \neq \mu_2. \end{cases}$$

pour un risque d'erreur α fixé on a donc :

$$P\left(|Z| \geq q_{1-\frac{\alpha}{2}}\right) = \alpha,$$

où

$$q_{1-\frac{\alpha}{2}} = \Phi^{-1}(1-\frac{\alpha}{2}).$$

La région critique du test est :

$$W = \left\{(x_1^{(1)}, \ \cdots, \ x_n^{(n_1)}; \ x_1^{(1)}, \cdots, \ x_2^{(n_2)}) \in \mathbb{R}^{n_1+n_2} \mid \frac{\left|\overline{X_1} - \overline{X_2}\right|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \geq q_{1-\frac{\alpha}{2}}\right\}.$$

La région d'acceptation est un intervalle de la forme : $]-\infty; -q_{1-\frac{\alpha}{2}}[\cup]q_{1-\frac{\alpha}{2}}; +\infty[$.

2. **Test unilatérale à gauche :**

$$\begin{cases} H_0 : \mu_1 = \mu_2. \\ H_1 : \mu_1 < \mu_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(Z \leq q_\alpha) = \alpha,$$

où

$$q_\alpha = \Phi^{-1}(\alpha).$$

La région critique du test est :

$$W = \left\{ (x_1^{(1)}, \cdots, x_{n_1}^{(1)}; x_1^{(2)}, \cdots, x_{n_2}^{(2)}) \in \mathbb{R}^{n_1+n_2} \mid \frac{\overline{X_1} - \overline{X_2}}{\sqrt{(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})}} \leq q_\alpha \right\}.$$

La région d'acceptation est un intervalle de la forme : $]-\infty; q_\alpha[$.

3. **Test unilatérale à droite :**

$$\begin{cases} H_0 : \mu_1 = \mu_2. \\ H_1 : \mu_1 > \mu_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(Z \geq q_{1-\alpha}) = \alpha,$$

où

$$q_{1-\alpha} = \Phi^{-1}(1-\alpha).$$

La région critique du test est :

$$W = \left\{ (x_1^{(1)}, \cdots, x_{n_1}^{(1)}; x_1^{(2)}, \cdots, x_{n_2}^{(2)}) \in \mathbb{R}^{n_1+n_2} \mid \frac{\overline{X_1} - \overline{X_2}}{\sqrt{(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})}} \geq q_{1-\alpha} \right\}.$$

La région d'acceptation est un intervalle de la forme : $[q_{1-\alpha}; +\infty[$.

σ_1^2 et σ_2^2 sont inconnues :

La variable aléatoire

$$\widetilde{S}_1^2 = \frac{1}{n_1 - 1}\sum(X_i - \overline{X_1})^2,$$

et

$$\widetilde{S}_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n}(X_i - \overline{X_2})^2.$$

a) *n_1 et n_2 supérieurs à* 30

La statistique D est définie comme suit :

$$D = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\tilde{S}_1^2}{n_1} + \frac{\tilde{S}_2^2}{n_2}}} \rightsquigarrow \mathcal{N}(0, 1). \tag{2.9}$$

1. Test bilatérale :

$$\begin{cases} H_0 : \mu_1 = \mu_2. \\ H_1 : \mu_1 \neq \mu_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(|D| \geq q_{1-\frac{\alpha}{2}}) = \alpha,$$

où

$$q_{1-\frac{\alpha}{2}} = \Phi^{-1}(1 - \frac{\alpha}{2}).$$

La région critique du test est :

$$W = \left\{ (X_1^{(1)}, \ \cdots, \ X_{n_1}^{(1)}), \ (X_2^{(2)}, \cdots, \ X_{n_2}^{(2)}) \in \mathbb{R}^{n_1+n_2} \mid \frac{|\overline{X}_1 - \overline{X}_2|}{\sqrt{\frac{\tilde{S}_1^2}{n_1} + \frac{\tilde{S}_2^2}{n_2}}} \geq q_{1-\frac{\alpha}{2}} \right\}.$$

On rejette H_0 au risque α si $D \notin] - q_{1-\frac{\alpha}{2}}; q_{1-\frac{\alpha}{2}}[$.

2. Test unilatérale à gauche :

$$\begin{cases} H_0 : \mu_1 = \mu_2. \\ H_1 : \mu_1 < \mu_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(D \leq q_\alpha) = \alpha,$$

où

$$q_\alpha = \Phi^{-1}(\alpha).$$

La région critique du test est :

$$W = \left\{ (X_1^{(1)}, \ \cdots, \ X_{n_1}^{(1)}), \ (X_2^{(2)}, \ \cdots, \ X_{n_2}^{(2)}) \in \mathbb{R}^{n_1+n_2} \mid \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\widetilde{S}_1^2}{n_1} + \frac{\widetilde{S}_2^2}{n_2}}} \leq q_\alpha \right\}.$$

On rejette H_0 au risque α si $D \notin]q_\alpha; +\infty[$.

Exemple 2.2.2 : On étudie l'efficacité d'un régime hypocholestérolémiant sur des animaux comparativement à des animaux témoins (non traités). On mesure le taux de cholestérol X dans les deux groupes et on obtient les résultats suivants :

Groupe traité : $n_1 = 100, \sum_{i=1}^{n_1} X_i = 300, \sum_{i=1}^{n_1} X_i^2 = 1035$

Groupe témoin : $n_2 = 100, \sum_{j=1}^{n_2} X_j = 450, \sum_{j=1}^{n_2} X_j^2 = 2190$

On souhaite tester l'efficacité du régime hypocholestérolémiant au risque 5%.

Les hypothèses à tester sont : $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$

$\overline{X_1} = \frac{300}{100} = 3, \overline{X_2} = \frac{450}{100} = 4.5$

$$\widetilde{S}_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i - \overline{X_1})^2 = \frac{\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}_1^2}{99} = \frac{1035-100\times 3^2}{99} = 1.3636$$

$$\widetilde{S}_2^2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}(X_j - \overline{X_2})^2 = \frac{\sum_{j=1}^{n_2} X_j^2 - n_2\overline{X}_2^2}{99} = 1.67$$

Statistique de test 2.9 est : $D = \dfrac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{\widetilde{S}_1^2}{n_1} + \frac{\widetilde{S}_2^2}{n_2}}} = \dfrac{3 - 4.5}{\sqrt{\frac{1.3636}{100} + \frac{1.67}{100}}} = -8.617.$

D'après le tableau de loi $\mathcal{N}(0, 1)$ nous trouvons : $q_\alpha = -1.645$,

Décision :

$q_\alpha = -1.645 > -8.617$, donc on rejette H_0.

Au risque de 5%, nous rejetons l'hypothèse nulle. Il existe des preuves statistiques très fortes pour conclure que le régime hypocholestérolémiant est efficace, car il réduit significativement le taux de cholestérol chez les animaux traités par rapport aux animaux témoins.

1. **Test unilatérale à droite :**

$$\begin{cases} H_0 : \mu_1 = \mu_2. \\ H_1 : \mu_1 > \mu_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(D \geq q_{1-\alpha}) = \alpha,$$

où

$$q_{1-\alpha} = \Phi^{-1}(1 - \alpha).$$

La région critique du test est :

$$W = \left\{ (x_1^{(1)}, \cdots, x_{n_1}^{(1)};\ x_1^{(2)}, \ \cdots, \ x_{n_2}^{(2)}) \in \mathbb{R}^{n_1+n_2} \mid \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\widetilde{S}_1^2}{n_1} + \frac{\widetilde{S}_2^2}{n_2}}} \geq q_{1-\alpha} \right\}.$$

On rejette H_0 au risque α si $D \notin]-\infty; q_{1-\alpha}[$.

B). *Si* $n_1 + n_2 - 2$ *est inférieur à* 30 *et* $\sigma_1 = \sigma_2$:

La statistique Z est définie comme suit :

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\widetilde{S}' \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightsquigarrow \mathcal{T}(n_1 + n_2 - 2), \tag{2.10}$$

où $T(n_1+n_2-2)$ est la loi de student à (n_1+n_2-2) degrés de liberté.

On pose :

$$\widetilde{S}' = \sqrt{\frac{n_1 \widetilde{s}_1^2 + n_2 \widetilde{s}_2^2}{n_1 + n_2 - 2}}. \tag{2.11}$$

1. Test bilatérale :

$$\begin{cases} H_0 : \mu_1 = \mu_2. \\ H_1 : \mu_1 \neq \mu_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(|Z| \geq t_{1-\frac{\alpha}{2}}) = \alpha,$$

où

$$t_{1-\frac{\alpha}{2}} = F^{-1}(1 - \frac{\alpha}{2}).$$

La région critique du test est :

$$W = \left\{ (x_1^{(1)}, \ \cdots, \ x_{n_1}^{(1)};\ x_1^{(2)}, \ \cdots, \ x_{n_2}^{(2)}) \in \mathbb{R}^{n_1+n_2} \mid \frac{|\overline{X}_1 - \overline{X}_2|}{\widetilde{S}' \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \geq t_{1-\frac{\alpha}{2}} \right\}.$$

On rejette H_0 au risque α si $Z \notin]-t_{1-\frac{\alpha}{2}}; t_{1-\frac{\alpha}{2}}[$.

Exemple 2.2.3 :[3] On considère deux lots de tasses et on souhaite comparer la solidité de ceux-ci. Pour chacun des deux lots, on dispose d'un échantillon de 10 tasses et on mesure la résistance de chacune d'entre eux. Les résultats sont :

pour le premier échantillon :

31.7	31.98	32.24	32.35	31.18	32.19	32.63	31.19	31.54	31.89

TAB. 2.4 – Table de la solidité d'une tasse du premier lot

pour le deuxième échantillon :

31.61	31.10	31.20	31.11	32.66	31.15	31.71	31.22	31.16	31.21

TAB. 2.5 – Table de la solidité d'une tasse du deuxième lot

La solidité d'une tasse du premier lot peut être modélisée par une var X_1, et celle du tasse du second lot peut être modélisée par une var X_2. On suppose que X_1 et X_2 suivent des lois normales de variances égales.

Peut-on affirmer que ces deux échantillons ne proviennent pas de la même production ?.

L'hypothèse à tester sont : $\begin{cases} H_0 : u_1 = u_2 \\ H_1 : u_1 \neq u_2 \end{cases}$

$n_1 = 10$, $n_2 = 10$, $\overline{X}_1 = 31.989$, $\widetilde{S}_1^2 = 0.1836$, $\overline{X}_2 = 31.413$, $\widetilde{S}_2^2 = 0.2013$,

$\widetilde{S}' = \sqrt{\frac{n_1\widetilde{S}_1^2 + n_2\widetilde{S}_2^2}{n_1+n_2-2}} = \sqrt{0.1925}$,

Statistique de test 2.10 est :$Z = \frac{|\overline{x}_1 - \overline{x}_2|}{\widetilde{S}'\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = 2.937$.

D'après le tableau de loi de student à (18) degrés de liberté nous trouvons : $t_{0.05,18} = 2.101$.

Décision :

$|Z| = 2.937 > 2.101$, donc on rejette H_0.

2. **Test unilatérale à gauche :**

$$\begin{cases} H_0 : \mu_1 = \mu_2. \\ H_1 : \mu_1 < \mu_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(Z \leq t_\alpha) = \alpha,$$

où

$$t_\alpha = F^{-1}(\alpha) = t_\alpha.$$

La région critique du test est :

$$W = \left\{ (x_1^{(1)}, \ \cdots, \ x_{n_1}^{(1)}; \ x_1^{(2)}, \cdots, \ x_{n_2}^{(2)}) \in \mathbb{R}^{n_1+n_2} \mid \frac{\overline{X}_1 - \overline{X}_2}{\widetilde{S}' \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq t_\alpha \right\}.$$

On rejette H_0 au risque α si $Z \notin]t_\alpha; +\infty[$.

3. **Test unilatérale à droite :**

$$\begin{cases} H_0 : \mu_1 = \mu_2. \\ H_1 : \mu_1 > \mu_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(Z \geq t_{1-\alpha}) = \alpha,$$

où

$$t_{1-\alpha} = F^{-1}(1-\alpha).$$

La région critique du test est :

$$W = \left\{ (x_1^{(1)}, \ \cdots, \ x_{n_1}^{(1)}; \ x_1^{(2)}, \ \cdots, \ x_{n_2}^{(2)}) \in \mathbb{R}^{n_1+n_2} \mid \frac{\overline{X}_1 - \overline{X}_2}{\widetilde{S}' \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \geq t_{1-\alpha} \right\}.$$

On rejette H_0 au risque α si $Z \notin] - \infty; t_{1-\alpha}[$.

C). *n_1 ou n_2 inférieur à 30 et $\sigma_1 \neq \sigma_2$*

La statistique T est définie comme suit :

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\widetilde{S}_1^2}{n_1} + \frac{\widetilde{S}_2^2}{n_2}}} \backsimeq t(\nu), \tag{2.12}$$

où

$$v = \frac{[\frac{\widetilde{S}_1^2}{n_1} + \frac{\widetilde{S}_2^2}{n_2}]^2}{\frac{\widetilde{S}_1^4}{n_1^2(n_1-1)} + \frac{\widetilde{S}_2^4}{n_2^2(n_2-1)}}. \tag{2.13}$$

1. **Test bilatérale :**

$$\begin{cases} H_0 : \mu_1 = \mu_2. \\ H_1 : \mu_1 \neq \mu_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(|T| \geq t_{1-\frac{\alpha}{2}}) = \alpha,$$

où

$$t_{1-\frac{\alpha}{2}} = F^{-1}(1 - \frac{\alpha}{2}).$$

La région critique du test est :

$$W = \left\{ (x_1^{(1)}, \ \cdots, \ x_{n_1}^{(1)}; \ x_1^{(2)}, \ \cdots, \ x_{n_2}^{(2)}) \in \mathbb{R}^{n_1+n_2} \mid \frac{|\overline{X}_1 - \overline{X}_2|}{\sqrt{\frac{\widetilde{S}_1^2}{n_1} + \frac{\widetilde{S}_2^2}{n_2}}} \geq t_{1-\frac{\alpha}{2}} \right\}.$$

On rejette H_0 au risque α si $Z \notin]-t_{1-\frac{\alpha}{2}}; t_{1-\frac{\alpha}{2}}[$.

1. **Exemple 2.2.4 :**[3] . On dispose de deux lots de boites de sauce algérienne emballées conditionnées de la même manière mais provenant de producteurs différents. On s'intéresse à la teneur en grammes de viande dans celles-ci.

▶ On extrait 7 boîtes provenant du premier producteur et on mesure leur teneur de viande. Les résultats, en grammes, sont :

12.12	12.03	13.58	13.38	11.81	15.92	13.65

Tab. 2.6 – Teneur en viande dans les pots de sauce algérienne du premier producteur (en grammes)

▶ On extrait 6 boîtes provenant du deuxième producteur et on mesure leur teneur de viande. Les résultats, en grammes, sont :

14.81	13.93	14.91	15.87	15.62	15.39

Tab. 2.7 – Teneur en viande des boîtes de sauce algérienne du deuxième producteur (en grammes)

La teneur en grammes de viande dans une boîte provenant du premier producteur peut être modélisée par un var X_1, et celle dans une boîte provenant du deuxième producteur peut être modélisée par un var X_2. On suppose que X_1 et X_2 suivent des lois normales.

Peut-on affirmer qu'il y a une différence entre les producteurs quant à la teneur moyenne en viande dans les boîtes ?.

On tester l'hypothèse : $\begin{cases} H_0 : \mu_1 = \mu_2. \\ H_1 : \mu_1 \neq \mu_2. \end{cases}$

$n_1 = 7$, $\overline{X}_1 = 13.21$, $\widetilde{S}_1^2 = 1.65$.

$n_2 = 6$, $\overline{X}_2 = 15.09$, $\widetilde{S}_2^2 = 0.45$.

Statistique de test 2.12 est : $|T| = \left| \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\tilde{S}_1^2}{n_1} + \frac{\tilde{S}_2^2}{n_2}}} \right| = |-3.37| = 3.37.$

$v = \frac{[\frac{\tilde{S}_1^2}{n_1} + \frac{\tilde{S}_2^2}{n_2}]^2}{\frac{\tilde{S}_1^4}{n_1^2(n_1-1)} + \frac{\tilde{S}_2^4}{n_2^2(n_2-1)}} = 9.29$

D'après le tableau de loi de student à (9) degrés de liberté nous trouvons : $t_{0.05,9} = 2.262$.

Décision :

$|T| = 3.37 > 2.262$, donc on rejette H_0.

2. **Test unilatérale à gauche :**

$$\begin{cases} H_0 : \mu_1 = \mu_2. \\ H_1 : \mu_1 < \mu_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(T \leq t_\alpha) = \alpha,$$

où

$$t_\alpha = F^{-1}(\alpha).$$

La région critique du test est :

$$W = \left\{ (x_1^{(1)}, \; \cdots, \; x_{n_1}^{(1)}; \; x_1^{(2)}, \cdots, \; x_{n_2}^{(2)}) \in \mathbb{R}^{n_1+n_2} \mid \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\tilde{S}_1^2}{n_1} + \frac{\tilde{S}_2^2}{n_2}}} \leq t_\alpha \right\}.$$

On rejette H_0 au risque α si $Z \notin]t_\alpha; +\infty[$.

3. **Test unilatérale à droite :**

$$\begin{cases} H_0 : \mu_1 = \mu_2. \\ H_1 : \mu_1 > \mu_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(T \geq t_{1-\alpha}) = \alpha,$$

où

$$t_{1-\alpha} = F^{-1}(1-\alpha).$$

La région critique du test est :

$$W = \left\{ (x_1^{(1)}, \cdots, x_{n_1}^{(1)};\ x_1^{(2)}, \cdots, x_{n_2}^{(2)}) \in \mathbb{R}^{n_1+n_2} \mid \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\widetilde{S}_1^2}{n_1} + \frac{\widetilde{S}_2^2}{n_2}}} \geq t_{1-\alpha} \right\}.$$

On rejette H_0 au risque α si $T \notin]-\infty; t_{1-\alpha}[$.

Remarque 2.2.1 : Le test de Student est assez robuste mais si l'on s'éloigne trop des conditions de normalité,

il est préférable d'utiliser un test non paramétrique.

Ce tableau 2.4 résume les informations ci-dessus :

2.2.2 Test de comparaison de deux variances

Soient X_1 et X_2 deux variables aléatoires indépendantes de lois normales $X_1 \rightsquigarrow \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \rightsquigarrow \mathcal{N}(\mu_2, \sigma_2^2)$.

On dispose de deux échantillons indépendants $(X_1^{(1)}, \cdots, X_{n_1}^{(1)})$, $(X_1^{(2)}, \cdots, X_{n_2}^{(2)})$ de tailles n_1 et n_2 respectivement, tels que X_i^1 (respectivement X_j^2) pour $1 \leq i \leq$

Paramètres	Hypothèse	Statistique de test	Région critique
σconnue	$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$	$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})}} \rightsquigarrow \mathcal{N}(0,1)$	$\lvert Z\rvert \geq q_{1-\frac{\alpha}{2}}$
	$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$		$Z < q_\alpha$
	$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases}$		$Z > q_{1-\alpha}$
Variance inconnue $n_1, n_2 \geq 30$	$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$	$D = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\widetilde{S}_1^2}{n_1} + \frac{\widetilde{S}_2^2}{n_2}}} \rightsquigarrow \mathcal{N}(0,1)$	$\lvert D\rvert \geq q_{1-\frac{\alpha}{2}}$
	$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$		$D < q_\alpha$
	$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases}$		$D > q_{1-\alpha}$
Variance inconnues $n_1+n_2-2 \leq 30$ $\sigma_1 = \sigma_2$	$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$	$Z = \frac{\overline{X}_1 - \overline{X}_2}{\widetilde{S}'\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightsquigarrow \mathcal{T}(n_1+n_2-2)$	$\lvert Z\rvert \geq t_{1-\frac{\alpha}{2}}$
	$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$		$Z < t_\alpha$
	$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases}$		$Z > t_{1-\alpha}$
Variance inconnues $n_1, n_2 \leq 30$ $\sigma_1 \neq \sigma_2$	$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$	$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\widetilde{S}_1^2}{n_1} + \frac{\widetilde{S}_2^2}{n_2}}} \rightsquigarrow \mathcal{T}(\nu)$	$\lvert T\rvert \geq t_{1-\frac{\alpha}{2}}$
	$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$		$T < t_\alpha$
	$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases}$		$T > t_{1-\alpha}$

TAB. 2.8 – Tableau de test de comparaison de deux moyennes

$n_1,\ 1 \leq j \leq n_2$ suit la même loi que X_1(respectivement X_2).

Moyennes connues :

On considère

$$\overline{S}_1^2 = \frac{1}{n_2} \sum_{i=1}^{n_1} (X_i^{(1)} - \mu_1)^2 \tag{2.14}$$

et

$$\overline{S}_2^2 = \frac{1}{n_1}\sum_{j=1}^{n_2}(X_j^{(2)} - \mu_2)^2 \tag{2.15}$$

La statistique V est définie comme suit :

$$V = \frac{n_1\overline{S}_1^2}{\sigma_1^2} \times \frac{\sigma_2^2}{n_2\overline{S}_2^2} = \frac{\frac{\overline{S}_1^2}{\sigma_1^2}}{\frac{\overline{S}_2^2}{\sigma_2^2}} \rightsquigarrow \mathcal{F}(n_1,\ n_2) \tag{2.16}$$

La statistique V suit une loi de Fisher $\mathcal{F}(n_1, n_2)$ à n_1 et n_2 degrés de liberté, telles que :

$$\frac{n_1\overline{S}_1^2}{\sigma_1^2} = \sum_{i=1}^{n_1}(\frac{x_i^{(1)} - \mu_1}{\sigma_1})^2 \rightsquigarrow \chi^2_{n_1}$$

$$\frac{n_2\overline{S}_2^2}{\sigma_2^2} = \sum_{i=1}^{n_2}(\frac{x_i^{(2)} - \mu_2}{\sigma_2})^2 \rightsquigarrow \chi^2_{n_2}$$

1. **Test bilatérale** :

$$\begin{cases} H_0 : \sigma_1 = \sigma_2. \\ H_1 : \sigma_1 \neq \sigma_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(|V| \leq f_{1-\frac{\alpha}{2}}) = \alpha,$$

où

$$P(V \geq f_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2},$$
$$f_{1-\frac{\alpha}{2}} = F^{-1}(1 - \frac{\alpha}{2}),$$

où

$$P(V \leq f_{\frac{\alpha}{2}}) = \frac{\alpha}{2}.$$

$$f_{\frac{\alpha}{2}} = F^{-1}(\frac{\alpha}{2}).$$

La région critique du test est :

$$W = \left\{(x_1^{(1)},\ \cdots,\ x_{n_1}^{(1)};\ x_1^{(2)},\ \cdots,\ x_{n_2}^{(2)}) \in \mathbb{R}^{n_1+n_2} \mid \quad V \geq f_{1-\frac{\alpha}{2}} \text{ ou } V \leq f_{\frac{\alpha}{2}}\right\}.$$

On rejette H_0 au risque α si $V \notin]f_{\frac{\alpha}{2}}; f_{1-\frac{\alpha}{2}}[$.

2. **Test unilatérale à droite :**

$$\begin{cases} H_0 : \sigma_1 = \sigma_2. \\ H_1 : \sigma_1 > \sigma_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(V \geq f_{1-\alpha}) = \alpha,$$

où

$$f_{1-\alpha} = F^{-1}(1-\alpha).$$

La région critique du test est :

$$W = \left\{(x_1^{(1)}, \cdots,\ x_{n_1}^{(1)};\ x_1^{(2)}, \cdots,\ x_{n_2}^{(2)}) \in \mathbb{R}^{n_1+n_2} \mid \quad V \geq f_{1-\alpha}\right\}.$$

On rejette H_0 au risque α si $V \notin]0;\ f_{1-\alpha}[$.

3. **Test unilatérale à gauche :**

$$\begin{cases} H_0 : \sigma_1 = \sigma_2. \\ H_1 : \sigma_1 < \sigma_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(V \leq f_\alpha) = \alpha,$$

où

$$f_\alpha = F^{-1}(\alpha).$$

La région critique du test est :

$$W = \left\{(x_1^{(1)}, \cdots,\ x_{n_1}^{(1)};\ x_1^{(2)}, \cdots,\ x_{n_2}^{(2)}) \in \mathbb{R}^{n_1+n_2} \mid V \leq f_\alpha \right\}.$$

On rejette H_0 au risque α si $Z \notin]f_\alpha;\ +\infty[$.

Moyennes inconnues :

On considère :

$$\widetilde{S}_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i^{(1)} - \overline{X}),$$

et

$$\widetilde{S}_2^2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (X_j^{(2)} - \overline{X}).$$

La statistique U est définie comme suit :

$$U = \frac{\frac{(n_1-1)\widetilde{S}_1^2}{\sigma_1^2(n_1-1)}}{\frac{(n_2-1)\widetilde{S}_2^2}{\sigma_2^2(n_2-1)}} = \frac{\tilde{S}_1^2}{\tilde{S}_2^2} \rightsquigarrow \mathcal{F}(n_1-1,\ n_2-1) \tag{2.17}$$

$\mathcal{F}(n_1-1,\, n_2-1)$ loi de ficher-snedecor$(n_1-1,\, n_2-1)$ dégréer de liberté.

1. **Test bilatérale** :

$$\begin{cases} H_0 : \sigma_1 = \sigma_2. \\ H_1 : \sigma_1 \neq \sigma_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(U \leq f_{\frac{\alpha}{2}} \quad \text{ou} \quad U \geq f_{1-\frac{\alpha}{2}}\ ) = \alpha,$$

où

$$P(U \leq f_{\frac{\alpha}{2}}) = \frac{\alpha}{2},$$
$$f_{\frac{\alpha}{2}} = F^{-1}(\frac{\alpha}{2}),$$

où

$$P(U \leq f_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}.$$
$$f_{1-\frac{\alpha}{2}} = F^{-1}(1 - \frac{\alpha}{2}).$$

La région critique du test est :

$$W = \left\{(x_1^{(1)},\ \cdots,\ x_{n_1}^{(1)};\ x_1^{(2)},\ \cdots,\ x_{n_2}^{(2)}) \in \mathbb{R}^{n_1+n_2} \mid \quad U \geq f_{1-\frac{\alpha}{2}} \ \text{ ou } \ U \leq f_{\frac{\alpha}{2}}\right\}.$$

On rejette H_0 au risque α si $U \notin]f_{\frac{\alpha}{2}};\, f_{1-\frac{\alpha}{2}}[$.

Remarque 2.2.2 : Le quantile d'ordre α de la loi $f(n_1-1,\, n_2-1)$ est égal à l'inverse du quantile d'ordre $1-\alpha$ de la loi $f(n_2-1, n_1-1)$, c-à-d **:** $f(n_1-1, n_2-1) = \frac{1}{f(n_2-1,n_1-1)}$

Exemple 2.2.5 : Pour fabriquer une solution contenant une certaine espèce chimique, on dispose de 2 procédés A et B. On se demande si les deux procédés fournissent des solutions dont le titre est bien homogène ou s'il existe une différence entre les deux procédés. On fabrique alors deux séries de solutions, l'une avec le procédé A, l'autre avec le procédé B, dont on vérifie le titre. L'homogénéité est quantifiée par la variance observée sur les résultats obtenus pour un même procédé (plus une méthode est homogène, plus la variance est petite)

On tester les hypothèses : $\begin{cases} H_0 : \sigma_1 = \sigma_2 \\ H_1 : \sigma_1 \neq \sigma_2 \end{cases}$

On réalise 30 aliquotes de solution avec A et 35 avec B. On a donc

$n_A = 19$ et $n_B = 24$. On trouve $\widetilde{S}_A^2 = 36$ et $\widetilde{S}_B^2 = 40$.

$P(U \leq f_{\frac{\alpha}{2}}(n_1-1,\, n_2-1)) = 0.025 \implies P(U \leq f_{0.025}(18,\, 23)) = 0.025$

D'après le tableau de loi de ficher-Snedecor à (18, 23) degrés de liberté nous trouvons :

$f_{0.025}(18,\, 23) = 2.07$

$P(U \geq f_{1-\frac{\alpha}{2}}(n_1-1,\, n_2-1)) = 0.025 \implies P(U \leq f_{0.975}(18,\, 23)) = 0.025$

$f_{1-\frac{\alpha}{2}}(n_2-1,\, n_1-1) = \frac{1}{f(18,23)} = 0.48$

Statistique de test 2.17 est :$U = \frac{36}{40} = 0.9$

Décision :

$U \in]f_{\frac{\alpha}{2}};\, f_{1-\frac{\alpha}{2}}[$, donc on accepte H_0.

Il n'y a pas de différence significative entre les variances des deux procédés. Autrement dit, les deux méthodes sont également homogènes du point de vue de la dispersion.

2. **Test unilatérale à droite :**

$$\begin{cases} H_0 : \sigma_1 = \sigma_2. \\ H_1 : \sigma_1 > \sigma_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(U \geq f_{1-\alpha}) = \alpha,$$

où

$$f_{1-\alpha} = F^{-1}(1-\alpha).$$

La région critique du test est :

$$W = \left\{(x_1^{(1)}, \ \cdots, \ x_{n_1}^{(1)};\ x_1^{(2)}, \ \cdots, \ x_{n_2}^{(2)}) \in \mathbb{R}^{n_1+n_2} \mid U \geq f_{1-\alpha}\right\}.$$

On rejette H_0 au risque α si $U \notin]0; f_{1-\alpha}[$.

3. **Test unilatérale à gauche :**

$$\begin{cases} H_0 : \sigma_1 = \sigma_2. \\ H_1 : \sigma_1 < \sigma_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(U \leq f_{\alpha}) = \alpha,$$

où

$$f_{\alpha} = F^{-1}(\alpha).$$

La région critique du test est :

$$W = \left\{(x_1^{(1)}, \cdots, x_{n_1}^{(1)}; x_1^{(2)}, \cdots, x_{n_2}^{(2)}) \in \mathbb{R}^{n_1+n_2} \mid U \leq f_\alpha\right\}.$$

On rejette H_0 au risque α si $U \notin]f_\alpha; +\infty[$.

Exemple 2.2.6 :

On désire savoir si, chez les individus qui consomment régulièrement de l'huile d'olive, le risque cardio-vasculaire est diminué. On utilise pour cela le logarithme du dosage en d-dimères, modélisé par une loi normale. Sur un échantillon de 10 individus consommant de l'huile d'arachide, on a observé une moyenne de -0.82, avec un écart type de 0.29. Sur un échantillon de 13 individus consommant de l'huile d'olive, on a observé une moyenne de -0.96, avec un écart-type de 0.34. La variance du premier échantillon peut-elle être inférieure à la variance du second échantillon au seuil de 0.05 ?

On tester l'hypothèse : $\begin{cases} H_0 : \sigma_1 = \sigma_2 \\ H_1 : \sigma_1 < \sigma_2 \end{cases}$

$n_1 = 10, n_2 = 13$, $\overline{X_1} = -0.82$, $\overline{X_2} = -0.96$, $\tilde{S}_1 = 0.29$, $\tilde{S}_2 = 0.34$

Statistique de test 2.17 est : $U = \dfrac{\frac{(n_1-1)\tilde{S}_1^2}{\sigma_1^2(n_1-1)}}{\frac{(n_2-1)\tilde{S}_2^2}{\sigma_2^2(n_2-1)}} = \dfrac{\tilde{S}_1^2}{\tilde{S}_2^2} = 0.727$

D'après le tableau de a loi de ficher-Snedecor à $(12, 9)$ dégrées de liberté nous trouvons : $f(12, 9) = 3.07$

$f_{(9,12)} = \frac{1}{(12,9)} = 0.357$

Décision :

$U = 0.727 > 0.357$, donc on accepte H_0.

Ce tableau 2.5 résume les informations ci-dessus :

Paramètres	Hypothèse	Statistique de test	Région critique
μconnues	$\begin{cases} H_0:\sigma_1=\sigma_2 \\ H_1:\sigma_1\neq\sigma_2 \end{cases}$	$V=\frac{n_1\overline{S}_1^2}{\sigma_1^2}\times\frac{\sigma_2^2}{n_2\overline{S}_2^2}\rightsquigarrow\mathcal{F}(n_1,n_2)$	$V\geq f_{1-\frac{\alpha}{2}}$ ou $V\leq f_{\frac{\alpha}{2}}$
	$\begin{cases} H_0:\sigma_1=\sigma_2 \\ H_1:\sigma_1>\sigma_2 \end{cases}$		$V>f_{1-\alpha}$
	$\begin{cases} H_0:\sigma_1=\sigma_2 \\ H_1:\sigma_1<\sigma_2 \end{cases}$		$V<f_{\alpha}$
μinconnues	$\begin{cases} H_0:\sigma_1=\sigma_2 \\ H_1:\sigma_1\neq\sigma_2 \end{cases}$	$U=\frac{\tilde{S}_1^2}{\tilde{S}_2^2}\rightsquigarrow\mathcal{F}(n_1-1,n_2-1)$	$U\geq f_{1-\frac{\alpha}{2}}$ ou $U\leq f_{\frac{\alpha}{2}}$
	$\begin{cases} H_0:\sigma_1=\sigma_2 \\ H_1:\sigma_1>\sigma_2 \end{cases}$		$U>f_{1-\alpha}$
	$\begin{cases} H_0:\sigma_1=\sigma_2 \\ H_1:\sigma_1<\sigma_2 \end{cases}$		$U<f_{\alpha}$

TAB. 2.9 – Tableau de test de comparaison de deux variances

2.2.3 Test de comparaison de deux proportions

Dans cette section, on supposera que les deux échantillons sont indépendants et de loi de Bernoulli et on comparera leurs paramètres.

Soit $X_1 = (X_1^{(1)}, \cdots, X_{n_1}^{(1)})$sont indépendantes et de même loi $\mathcal{B}(p_1)$ et $X_2 = (X_1^{(2)}, \cdots, X_{n_2}^{(2)})$sont indépendantes et de même loi $\mathcal{B}(p_2)$. On a :

$$T_1 = \tfrac{1}{n}\sum_{i=1}^{n_1} X_i^{(1)} \qquad \text{et} \qquad T_2 = \tfrac{1}{n}\sum_{j=1}^{n_2} X_j^{(2)}$$

qui sont indépendantes et respectivement de lois binomiales $\mathcal{B}(n_1, p_1)$, $\mathcal{B}(n_2, p_2)$.

$$\hat{p}_1 = \frac{T_1}{n_1} \qquad \text{et} \qquad \hat{p}_2 = \frac{T_2}{n_2}$$

$\hat{p}_1$ et $\hat{p}_2$ sont indépendantes et (asymptotiquement) de lois normales $\mathcal{N}(p_1; \frac{p_1(1-p_1)}{n_1})$ et $\mathcal{N}(p_2; \frac{p_2(1-p_2)}{n_2})$ donc :

$$\hat{p}_1 - \hat{p}_2 \rightsquigarrow \mathcal{N}(p_1 - p_2, \ \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}),$$

sous H_0, $p_1 = p_2$. On peut donc poser $p = p_1 = p_2$.

Alors :

$$\hat{p}_1 - \hat{p}_2 \rightsquigarrow \mathcal{N}(0,\ p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})),$$

et

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})}} \rightsquigarrow \mathcal{N}(0,1).$$

En général p est inconnu et, sous l'hypothèse (H_0) on réunit les deux échantillons, alors en estime p par : $\hat{p} = \dfrac{T_1 + T_2}{n_1 + n_2}$.

La statistique U est définie comme suit :

$$U = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} \rightsquigarrow \mathcal{N}(0,1), \tag{2.18}$$

conditions d'applications : $n_1 + n_2 > 30$, $n_1 p_1 \geq 5$, $n_2 p_2 \geq 5$, $n_1(1-p_1) \geq 5$, $n_2(1-p_2) \geq 5$.

1. **Test bilatérale :**

$$\begin{cases} H_0 : p_1 = p_2. \\ H_1 : p_1 \neq p_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(|U| \geq q_{1-\frac{\alpha}{2}}) = \alpha,$$

où

$$P(U \leq q_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2},$$

où

$$q_{1-\frac{\alpha}{2}} = \Phi^{-1}(1-\frac{\alpha}{2}).$$

La région critique est :

$$W = \left\{ (x_1^{(1)}, \cdots, x_{n_1}^{(1)}; x_1^{(2)}, \cdots, x_{n_2}^{(2)}) \in \{0,1\}^{n_1+n_2} / \qquad |U| \geq q_{1-\frac{\alpha}{2}} \right\}.$$

On rejette H_0 au risque si $U \notin]-q_{1-\frac{\alpha}{2}}; q_{1-\frac{\alpha}{2}}[$.

2. **Test unilatérale à droite :**

$$\begin{cases} H_0 : p_1 = p_2. \\ H_1 : p_1 > p_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(U \geq q_{1-\alpha}) = \alpha,$$

où

$$q_{1-\alpha} = \Phi^{-1}(1-\alpha).$$

La région critique est :

$$W = \left\{ (x_1^{(1)}, \ \cdots, \ x_{n_1}^{(1)}; \ x_1^{(2)}, \ \cdots, \ x_{n_2}^{(2)}) \in \{0,1\}^{n_1+n_2} \mid U \geq q_{1-\alpha} \right\}.$$

On rejette H_0 au risque si $U \notin]-\infty; q_{1-\alpha}[$.

3. **Test unilatérale à gauche :**

$$\begin{cases} H_0 : p_1 = p_2. \\ H_1 : p_1 < p_2. \end{cases}$$

Pour un risque d'erreur α fixé on a donc :

$$P(U \leq q_\alpha) = \alpha,$$

où

$$q_\alpha = \Phi^{-1}(\alpha).$$

La région critique est :

$$W = \left\{(x_1^{(1)}, \ \cdots, \ x_{n_1}^{(1)}; \ x_1^{(2)}, \ \cdots, \ x_{n_2}^{(2)}) \in \{0,1\}^{n_1+n_2} \mid \qquad U \leq q_\alpha\right\}.$$

On rejette H_0 au risque si $U \notin]q_\alpha; +\infty[$.

Exemple 2.2.7 :

Soit p_1 la probabilité de guérison d'une maladie donnée grâce à un traitement A. Un groupe de 60 malades est soumis à ce traitement et 38 guérissent. Un autre traitement B permet de soigner cette maladie, avec probabilité p_2. Sur 70 malades soumis à ce nouveau traitement, 40 guérissent.

1.Quel test proposez-vous pour décider si le nouveau traitement est meilleur que l'ancien ?

2. Donnez la p-valeur de ce test pour les données de l'énoncé. Quelle est votre conclusion ?

Les hypothèses du test sont les suivantes : $\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 < p_2 \end{cases}$

Pour le traitement A : $n_A = 60$, $X_A = 38$, $\widehat{p}_A = \frac{39}{60} = 0.633$

Pour le traitement B : $n_B = 70$, $X_B = 40$, $\widehat{p}_B = \frac{48}{70} = 0.57$

Statistique de test 2.18 : $U = \dfrac{\widehat{p}_A - \widehat{p}_B}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} = \dfrac{0.633 - 0.57}{\sqrt{0.6(1-0.6)(\frac{1}{(60} + \frac{1}{70})}} \approx$

0.73

$\widehat{p} = \dfrac{38+40}{60+70} = 0.6$

p-valeur $= P(U \leq 0.73) \approx \Phi^{-1}(0.73) \approx 0.7673$

Décision :

p-valeur $= 0.7673 > 0.05$, donc on accepte H_0.

Il n'y a pas de preuve statistique suffisante pour affirmer que le traitement B est meilleur que le traitement A au niveau de signification 5%.

Ce tableau 2.6 résume les informations ci-dessus :

Hypothèse	Statistique de test	Région critique
$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$		$\lvert U \rvert \geq q_{1-\frac{\alpha}{2}}$
$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 > p_2 \end{cases}$	$U = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})}} \rightsquigarrow \mathcal{N}(0,1)$	$U \geq q_{1-\alpha}$
$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 < p_2 \end{cases}$		$U \leq q_{\alpha}$

TAB. 2.10 – Tableau de test de comparaison de deux proportions

Chapitre 3

Application sous R

Dans ce chapitre, nous appliquerons les exemples de tests statistiques paramétriques étudiés, en utilisant le logiciel d'analyse statistique R.

3.1 Test de la moyenne d'une population normale :

▷ **Instruction R** : on utilise la fonction t.test()

Application de l'exemple2.1.1 :(cas unilatéral à gauche) :

Application sous R :

```
medic= c(23,26,22,32,19,29,36,28,25,30,31,20)
t.test(medic,alternative="less",mu=30)
One Sample t-test
data : medic
t = -2.1768, df = 11, p-value = 0.02608
alternative hypothesis : true mean is less than 30
95 percent confidence interval :
-Inf    29.43133
sample estimates :
mean of x
26.75
```

Résultats obtenus :

data :

t	df	p-value
-2.1768	11	0.02608

Interval de confidence à 95% :$]-\infty; 29.43133[$

sample estimates (moyenne observée) :

mean of x =	26.75

Commentaire :

p- value est inférieure à 0.05, donc on rejette l'hypothèse H_0.

Le (t.test) montre que l'efficacité moyenne du traitement B est significativement inférieure à celle du traitement

la p-valeur faible (2.6%) confirme cette différence, on conclura que B est plus efficace que A.

3.2 Test d'une proportion :

▷ **Instruction R** : on utilise la fonction prop.test()

Application de l'exemple2.1.3 : (cas unilatéral à droite)

Application sous R :

```
prop.test(420, 800, p=0.5, alternative="greater")

1-sample proportions test with continuity correction

data : 420 out of 800, null probability 0.5

X-squared = 1.9013, df = 1, p-value = 0.08397

alternative hypothesis : true p is greater than 0.5

95 percent confidence interval :

0.4952991  1.0000000

sample estimates :

p

0.525
```

Résultats obtenus :

data :

X-squared	df	p-value
1.9013	1	0.08397

Interval de confidence à 95% : $]0.4952991; 1.0000000[$

sample estimates :

p	=	0.525

Commentaire :

∗ p- value est superieur à 0.05, donc on accepte H_0.

L'intervalle de confiance indique que les résultats sont trop serrés pour que l'on puisse affirmer en toute confiance que le candidat A obtiendra 50% ou plus des voix de l'ensemble de la population.

3.3 Test de comparaison de deux moyennes :

Instruction R : on utilise la fonction : t.test()

Application de l'exemple 2.2.1 :(cas bilatéral)

Application sous R

```
x1 = c(31.70, 31.98, 32.24, 32.35, 31.18, 32.19, 32.63, 31.19, 31.54, 31.89)
x2 = c(31.61, 31.10, 31.20, 31.11, 32.66, 31.15, 31.71, 31.22, 31.16, 31.21)
t.test(x1, x2, var.equal = T)
Two Sample t-test
data : x1 and x2
t = 2.1875, df = 18, p-value = 0.04214
alternative hypothesis : true difference in means is not equal to 0
95 percent confidence interval :
0.01884893   0.93315107
sample estimates :
mean of x mean of y
31.889   31.413
```

Résultats obtenus :

t	df	p-value
2.1875	18	0.04214

Interval de confidence à 95% :]0.01884893; 0.93315107[

sample estimates :

mean of x	mean of y
31.889	31.413

Commentaire :

p-value = 0.04214 est inferieur à 0.05, donc on rejette H_0.

Le (t.test) révèle une différence significative entre les deux groupes, avec une p-valeur de 4.2%. L'intervalle de confiance ne contenant pas 0 confirme cette conclusion. On peut donc affirmer qu'il y a une réelle différence entre les deux échantillons au seuil de signification de 5%.

3.4 Test de comparaison de deux variances :

Instruction R : on utilise la fonction var.test

Application de l'exemple2.2.2(cas unilatérale à gauche)

Application sous R :

```
>groupe1=rnorm(10, mean = -0.82, sd =0.29)
>groupe2=rnorm(13, mean = -0.96,sd =0.34)
>resultat_test= var.test(groupe1, groupe2, alternative= "less")
>print(resultat_test)
F test to compare two variances
data : groupe1 and groupe2
F = 1.2064, num df = 9, denom df = 12, p-value = 0.6274
alternative hypothesis : true ratio of variances is less than 1
95 percent confidence interval :
0.000000 3.707222
sample estimates :
ratio of variances
1.206406
```

Résultats obtenus :

data :

F	num df	denom df	p-value
1.2064	9	12	0.6274

confidence Interval de confidence à 95% : $]0; 3.707222[$

sample estimates :

ratio of variances =	1.206406

Commentaire :

p-value $=0.6274 > 0.05$, donc on accepte H_0.

Au seuil de signification de 5%, nous ne pouvons pas affirmer que la variance du premier échantillon est inférieure à celle du second échantillon.

3.5 Test de comparaison de deux proportions :

$\triangleright$ **Instruction R** : on utilise la fonction prop.test

Application de l'exemple2.2.3 : (cas unilatérale à gauche)

Application sous R :

```
>rop.test(c(38,40),c(60,70),alternative="less", correct=F)
p2-sample test for equality of proportions without continuity correction
data : c(38, 40) out of c(60, 70)
X-squared = 0.51587, df = 1, p-value = 0.7637
alternative hypothesis : less
95 percent confidence interval :
-1.0000000  0.2031029
sample estimates :
prop 1 prop 2
0.6333333   0.5714286
```

Résultats obtenus :

data :

X-squared	df	p-value
0.51587	1	0.7637

Interval de confidence à 95% : $]-1.0000000;\ 0.2031029[$

sample estimates :

prop 1	prop 2
0.6333333	0.5714286

Commentaire :

p-valeur $= 0.7637 > 0.05$, donc on accepte H_0.

Le (rop.test) nous permet de conclure que le traitement B est ne meilleur pas que le traitement A au niveau designification 5%.

Conclusion

En conclusion, ce mémoire a porté sur l'étude des tests statistiques, qui sont un outil crucial dans la recherche scientifique et l'analyse de données. Une compréhension approfondie des principes des tests et de leurs hypothèses, ainsi que la manière d'interpréter correctement leurs résultats, permet d'éviter les erreurs et de prendre la meilleure décision.

Les tests statistiques ont été mis en avant en tant que branche importante de la statistique inférentielle. Certains d'entre eux ont été étudiés, tels que le test de conformité et le test d'homogénéité, ainsi que la manière de comparer la moyenne, la variance et les proportions, en mentionnant des exemples illustratifs.

Les tests statistiques ont été mis en avant car ils constituent une branche importante de la statistique inférentielle, et j'en ai étudié certains : le test de conformité, le test d'homogénéité, ainsi que la comparaison des moyennes, des variances et des proportions, avec des exemples explicatifs. J'ai utilisé le langage R, les différents packages et les commandes qui conviennent à chaque test statistique pour vérifier la validité des résultats.

Bibliographie

[1] Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., & Cochran, J. J. (2017). Statistiques pour l'économie et la gestion (C. Borsenberger, Trad. ; 5 éd.). De Boeck Supérieur. (Traduction de la 7 édition américaine).

[2] Boulay, J.-P. (n.d.). Statistique mathématique : Applications commentées (Collection Technosup). Ellipses.

[3] Christophe Chesneau. Licence France 2016. Introduction aux tests statistique avec R. Université de Caen.

[4] Dumont, F. (s.d.). Statistiques pour laboratoires. [Matériel de cours]. Fire Testing Laboratory, Université de Liège.

[5] Gaudoin Oliver. Principes et méthodes statistique, note de cours..

[6] Jean-Jacques Ruch. (2012.2013). Statistique : tests d'hypothèses.

[7] Merghni Djamel, Necir Abdelhakim. Cours de tests statistique master1. UMKB.

[8] Michel Lejeune. Statistique, la théorie et ses applications. 2eme version. Université de Grenoble 2.

[9] Monbet.V. 2009. Tests statistique, Notes de cours.

[10] Saporta, G. (2006). Probabilités, analyse des données et statistique. Editions Technip.

[11] Veysseyre, R. (s.d.). Aide-mémoire : Statistique et probabilités pour l'ingénieur (2e éd.). Dunod.

RÉSUMÉ

Les tests statistiques sont un outil précieux dans le domaine de la statistique utilisés pour analyser des données liées aux hypothèses d'un ensemble statistique, dans le but de prendre des décisions scientifiques fiables. Ces tests jouent un rôle important dans plusieurs domaines scientifiques.

L'objectif de ce travail est à connaître les bases des tests statistiques et à étudier certains tests paramétriques (test de conformité, test d'homogénéité, test des proportions) avec des exemples scientifiques et l'utilisation de certaines simulations dans le programme R.

Mots-clés : test statistique, hypothèse, acceptation, rejet, décision, niveau de signification,

valeur P.

الملخص

تعد الاختبارات الإحصائية أداة ثمينة في مجال الإحصاء تُستخدم لتحليل البيانات المتعلقة بفرضيات مجموعة إحصائية، بهدف اتخاذ قرارات علمية موثوقة، تلعب هذه الاختبارات دورًا مهمًا في العديد من المجالات العلمية. الهدف من هذا العمل هو التعرف على أساسيات الاختبارات الإحصائية ودراسة بعض الاختبارات المعلمية (اختبار التطابق، اختبار التجانس، اختبار النسب) مع أمثلة علمية واستخدام بعض المحاكاة في برنامج R

الكلمات المفتاحية: اختبارإحصائي، فرضية ، قبول، رفض، قرار، مستوى الدلالة، قيمة P.

Abstract

Statistical tests are a valuable tool in the field of statistics used to analyze data related to the hypotheses of a statistical set, with the aim of making reliable scientific decisions. These tests Play an important role in several scientific fields.

The objective of this work is to understand the basics of statistical tests and to study certain parametric tests (goodness-of-fit test, homogeneity test, proportion test) with scientific examples and the use of certain simulations in the R program.

Keywords: statistical test, hypothesis, acceptance, rejection, decision, significance level, P-value.