République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

DÉPARTEMENT DE MATHÉMATIQUES

Mémoire présenté en vue de l'obtention du Diplôme **:**

MASTER en Mathématiques

Option : **Statistique**

Par

BOUBECHE Noureddine

Titre :

Test d'hypothèse statistique

Membres du Comité d'Examen :

Dr.	**CHINE AMEL**	UMKB	Président
Dr.	**GHOZLEINE BENBRAIKA**	UMKB	Encadreur
Dr.	**OUANOUGHI YASMINA**	UMKB	Examinateur

September 2020

Dédicace

Tous les mots ne sauraient exprimer la gratitude, l'amour, le respect, la reconnaissance, c'est tous simplement que :

Je dédie ce travail à :

A Ma tendre Mère Djamila .

A Mon très cher Père Messaoud.

A mon cher frères : Aziz,Mohammed,Ali.

A mes sœurs : Naima, Razika ,Donia et Sonia.

A mes très chère amis :Mohammed , Khaled, Houssem ,Imen ,Samira,Soumia.

A tous les membres de ma promotion.

A tous mes enseignants depuis mes premières années d'études.

A tous ceux qui me sens chers et que j'ai omis de citer

Boubeche Noureddine

REMERCIEMENTS

Ce travail de mémoire de master est la première expérience dans l'activité de recherche.

Il n'aurait pas été aussi fructueux sans l'aide de plusieurs personnes.

Je vais donc m'essayer à trouver les mots justes pour exprimer spécifiquement ma reconnaissance à tous ceux qui ont contribue de près ou de loin à ce travail.

Avant tout, Je tient à remercier **ALLAH** le tout puissant qui m'aidée, et donnée la santé, la volonté et la patience pour achever ce travail.

Je tiens à remercier mon encadreur :

Ghozleine Benbraika pour la suivi et l'aide qu'elle m'apporté pour l'élaboration et pour ses précieux conseils.

Je voudrai également remercier les membres de mon jury :

CHINE AMEL

OUANOUGHI YASMINA

pour l'honneur qu'ils m'ont fait en portant leur attention sur ce travail

Merci à toutes et à tous.

Boubeche Noureddine

Table des matières

Table des figures

Liste des tableaux

Introduction

La statistique est la science dont l'objet est de recueillir, de traiter et d'analyser des données issues de l'observation de phénomènes aléatoires, c'est-à-dire dans les quels le hasard intervient.

La statistique peut être divisée en deux composantes majeures : la statistique descriptive et la statistique inférentielle . La statistique descriptive s'occupe de la collection ainsi que du calcul des paramètres descriptifs (moyenne, variance, écart-type, ...) et de la présentation des données. La statistique inférentielle va au-delà des données échantillonnées pour les généraliser&^{2}à une population entière, ce qui suppose par conséquent la formulation de prédictions et comme toute prédiction comporte un certain degré d'incertitude, une décision issue de l'inférence statistique n'est jamais sûre à 100%.

Dans tous les domaines, de l'expérimentation scientifique à la vie quotidienne, on est amené à prendre des décisions sur une activité risquée au vu de résultats d'expériences ou d'observation de phénomènes dans un contexte incertain, par exemple :

Informatique/ au vu des résultats des tests d'un nouveau système informatique, on doit décider si ce système est suffisamment fiable et performant pour être mis en vente.

Essais thérapeutiques : décider si un nouveau traitement médical est meilleur qu'un ancien au vu du résultat de son expérimentation sur des malades.

Finance : au vu du marché, décider si on doit ou pas se lancer dans une opération financière donnée.

Dans chaque cas, le problème de décision consiste à trancher, au vu d'observations, entre une hypothèse appelée hypothèse nulle , et une autre hypothèse dite hypothèse alternative

En général, on suppose qu'une et une seule de ces deux hypothèses est vraie.

Ce mémoire est divisé en trois chapitres :

Dans le premier chapitre, nous nous limitons à un bref rappel sur des notions et définitions de base sur les paramètres statistique et sur quelques éléments fondamentaux(variable aléatoire et fonction de repartions et le théorème central-limite) que nous jugeons utiles pour la suite de notre travail.

Dans le deuxième chapitre nous nous intéressons essentiellement aux tests d'hypothèse , nous exposons l'idée des notions générales des tests et nous étudions deux types de tests : tests de conformité et tests de d'homogénéité.

Dans le troisième chapitre, est, quant à lui, consacré à l'application sous logiciel R de quelques procédures présentées dans les chapitres antérieurs.

Chapitre 1

Généralités sur les paramètres

La base de toute étude statistiques, il y a une population formée d'individus sur lequel on observe un ou des caractères. Par exemple pour fixer les idées, on peut penser en terme de population humaine, les individus sont des personnes et les caractères observés peuvent être : morphologiques (taille, poids, couleur des yeux,..), physiologiques (groupe sanguin, taux de cholestérol,..) et psychologiques (réaction à un test, sondage,...).

1.1 Les caractères statistique

1.1.1 Population

En statistique, on travaille sur des populations. Ce terme vient du fait que la démographie, étude des populations humaines, a occupé une place centrale aux débuts de la statistique, notamment au travers des recensements de population. Mais, en statistique, le terme de population s'applique à tout objet statistique étudié, qu'il s'agisse d'étudiants (d'une université ou d'un pays), de ménages ou de n'importe quel autre ensemble sur lequel on fait des observations statistiques.

Définition 1.1.1 *On appelle population l'ensemble sur lequel porte notre étude statistique. Cette ensemble est noté Ω, par exemple : Ensemble des habitants d'une ville, ensemble des étudiants d'une école.*

1.1.2 Echantillon

De nombreux problèmes faisant intervenir le calcul des probabilités se ramènent aux problèmes de tirer des échantillons de taille r dans un ensemble de taille n, appelé population, quelle que soit la nature de ses éléments. Suivant la règle du tirage, cet échantillon est :

- ordonné ou non.
- avec ou sans répétitions (on dit aussi avec ou sans remise).

Deux autres espaces interviennent souvent dans des problèmes élémentaires, l'espace des sous-populations de taille r avec répétitions et l'espace des permutations de n objets.

Remarque 1.1.1 *Choisir un élément au hasard, signifie que les divers choix possibles sont équiprobables donc que l'ensemble Ω est muni de la loi de probabilité uniforme. Dans ce cas, tous les calculs sont simples et se ramènent souvent à des calculs d'analyse combinatoire.*

1.1.3 Individu (unité statistique)

Une population est composée d'individus. Les individus qui composent une population statistique sont appelés unités statistiques.

Définition 1.1.2 *On appelle individu tout élément de la population Ω, il est noté ω (ω dans Ω).*

1.1.4 Caractère (variable statistique)

La statistique « descriptive », comme son nom l'indique cherche à décrire une population donnée. Nous nous intéressons au caractéristique des unités qui peuvent prendre différentes valeurs.

Définition 1.1.3 *On appelle caractère (ou variable statistique, dé notée v.s) toute application*

$$X : \Omega \to C.$$

L'ensemble C est dit : ensemble des valeurs du caractère X (c'est ce qui est mesuré ou observé sur les individus).

Exemple 1.1.1 *Taille, température, nationalité, couleur des yeux, ...*

1.1.5 Variable aléatoire réelle

On appelle variable aléatoire réelle (abréviation : v.a.r) toute application de Ω dans $\mathbb{R}$.

$$X : \Omega \to \mathbb{R}.$$

$$\omega \mapsto X(\omega).$$

• I un sous-ensemble de $\mathbb{R}$.L'ensemble $X^{-1}(I) = \{\omega \in \Omega \ / X(\omega) \in I\}$,qui est un sous-ensemble de Ω,est un évènement aléatoire : on le note$(X \in I)$.

On dit que la variable aléatoire X est :

• discrète finie si l'ensemble $X(\Omega)$ est fini, discrète infinie si l'ensemble $X(\Omega)$ est infini dénombrable.

• continue si l'ensemble $X(\Omega)$ est un intervalle de $\mathbb{R}$ non réduit à un point(ou une réunion d'intervalles de $\mathbb{R}$).

1.1.6 Fonction de répartion

Soit X une variable aléatoire réelle, on appelle fonction de répartion de X, que l'on note F(ou F_x), la fonction définie par :

$$F : \mathbb{R} \to [0,1] .$$
$$x \mapsto F(x) = \mathbb{P}(X \leqslant x) = \mathbb{P}_x(] - \infty, x]).$$

qui est croissante monotone et telle que $\lim\limits_{x\to-\infty} F(x) = 0$ et $\lim\limits_{x\to+\infty} F(x) = 1$.

- Si la variable aléatoire X absolument continue la dérivée de F notée f est appelée densité de probabilité de la variable aléatoire X.La fonction f ,définie sur tout $\mathbb{R}$,vérifie : $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$.Et ona la relation fondamentale suivante :

$$\forall x \in \mathbb{R},\ F(x) = \int\limits_{-\infty}^{x} f(t)dt.$$

1.1.7 Le théorème central-limite

L'étude de sommes de variables indépendantes et de même loi joue un rôle capital en statistique.Le théorème suivant connu sous le nom de théorème central-limite (il vaudrait mieux dire théorème de la limite centrée) établit la convergence vers la loi de Gauss sous des hypothèses peu contraignantes.

Théorème 1.1.1 *Soit (X_n) une suite de variables aléatoires indépendantes de même loi d'espérance μ et d'écart-type δ.Alors :*

$$\sum_{i=1}^{n} \frac{X_i - \mu}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \left(\frac{X_1 + X_2 + ... + X_n - n\mu}{\sigma} \right) \rightsquigarrow N(0;1).$$

1.1.8 Les différents types de caractères

Un caractère peut être qualitatif ou quantitatif. Les méthodes d'analyse d'une population différente suivant la nature du caractère étudié.

a) caractère qualitatif

Définition 1.1.4 *Un caractère qualitatif est un caractère dont les modalités échappent à la mesure.*

b) caractère quantitatif

Définition 1.1.5 *Un caractère est qualifié de quantitatif lorsqu'il est mesurable ou repérable.*

Remarque 1.1.2 *Quand il est quantitatif, un caractère peut être :*

- *discret, quand il prend un nombre fini de valeurs.*
- *continu, quand il prend toute valeur comprise entre deux nombres donnés.*

1.1.9 Modalités

Les modalités d'une variable statistique sont les différentes valeurs que peut prendre celle-ci par exemple le variable est " situation familiale" et modalités sont " célibataire, marié, divorcé "

1.1.10 Les effectifs et les fréquences

Les effectifs

L'effectif total est le nombre d'individus appartenant à la population.

Définition 1.1.6 *L'effectif d'une modalité x_i d'un caractère X est le nombre d'individus présentant cette modalité On compte ainsi le nombre de fois que cette modalité du caractère apparait dans la population étudiée. L'effectif total est noté N. L'effectif correspondant à la modalité x_i du caractère X est noté n_i Ona donc :*

$$\sum_{i=1}^{k} n_i = N.$$

Les fréquences

Définition 1.1.7 *La fréquence d'une modalité x_i d'une caractère X est la proportion d'undividus de la population totale qui présentent cette modalité. On la note généralement f_i.Elle est égale à :*

$$f_i = \frac{n_i}{N}$$

1.2 Paramètres de position(les valeurs centrales)

Il existe trois valeurs centrales : le mode, la moyenne arithmétique et la médiane.

1.2.1 Le mode

Valeur la plus fréquente pour une distribution discrète,classe correspondant au pic de l'histogramme pour une variable continue sa détermination est malisée et dépend du découpage en classes.

1.2.2 La moyenne arithmétique

Pour calculer la moyenne arithmétique, deux cas sont à distinguer selon la façon dont les donnée sont été recueillies.

• Cas 1 : n données non réparties en classes :

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Exemple 1.2.1 *Les nombres d'enfants de 6 familles sont les suivants : 0, 1, 2, 2, 3,4.*
La moyenne est : $\overline{X} = \dfrac{0+1+2+2+3+4}{6} = 2.$

• Cas 2 : n données réparties en k classes, x_i la valeur de la ième modalité et n_i l'effectif correspondant à cette modalité et en utilisant la fréquence f_i :

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} f_i x_i.$$

Exemple 1.2.2 *On considère le tableau :*

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

TAB. 1.1 – Les valeurs distinctes et les effectifs

$$\overline{X} = \frac{0\times 1 + 1\times 1 + 2\times 2 + 3\times 1 + 4\times 1}{6} = 2.$$

1.2.3 La médiane

La médiane est la valeur que partage la lise des observations, parélabelement classées en ordre croissant en deux sous-listes qui contiennent le meme nombre d'observations. Soit N est le nombre d'observations classés alors, la médiane notée $\widetilde{X}$ vaut :

• Si N est impair, la médiane est :

$$\widetilde{X} = x_{\frac{N+1}{2}}$$

• Si N est pair, la médiane est :

$$\widetilde{X} = \frac{x_{\frac{N}{2}+1} + x_{\frac{N}{2}}}{2}$$

1.3 Paramètres dispersion

Les indicateurs statistiques de dispersion sont : l'étendue, la variance, l'écart type, les quartiles et l'écart interquartile.

1.3.1 L'étendue

La différence entre la plus grande valeur et la plus petite valeur du caractère, donnée par la quantité :

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

1.3.2 La variance

Soit une série de valeurs d'une variable X : $\{x_1, x_2, ..., x_k\}$ .Soit les effectifs associés $\{n_1, n_2, ..., n_k\}$,la variance de cette série s'écrit :

1- $\sigma^2 = \frac{1}{n}\sum\limits_{i=1}^{k} n_i(x_i - \overline{x})^2$.Si l'effectif considéré est celui d'une population.

2- $\sigma^2 = \frac{1}{n-1}\sum\limits_{i=1}^{k} n_i(x_i - \overline{x})^2$.Si l'effectif considéré est celui d'un échantillon.

1.3.3 L'écart-type

L'écart type est égal à la racine carrée de la varince :

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \overline{x})^2}$$

1.3.4 Les quartiles

Pour une comparaison différente, on peut déterminer les quartiles.

Définition 1.3.1 *Les quartiles d'une série statistique sont les trois valeurs* Q_1,Q_2 *et* Q_3.*Du caractère qui partagent la population en quatre parties de même effectif :*

• *Premier quartile* Q_1 *:est la plus petite valeur du caractère telle qu'au moins* 25% *des données lui sont inférieures ou égales.*

• *Deuxième quartile* Q_2 : *C'est la médiane.*

• *Troisième quartile* Q_3 : *est la plus petite valeur du caractère telle qu'au moins* 75% *des données lui sont inférieures ou égales.*

1.3.5 L'écart interquartile

C'est la différence entre le troisième et le premier quartile

$$\text{Ecart interquartile} = Q_3 - Q_1.$$

Chapitre 2

Test statistique

2.1 Introduction

Les tests d'hypothèses font appel à un certain nombre d'hypothèses concernant la nature de la population dont provient l'échantillon étudié (normalité de la variable, égalité des variances, etc). En fonction de l'hypothèse testée, plusieurs types de tests peuvent être réalisés :

• Les tests destinés à vérifier si un échantillon peut être considéré comme extrait d'une population donnée, vis-à-vis d'un paramètre comme la moyenne ou la fréquence observée (tests de conformité) ou par rapport à sa distribution observée (tests d'ajustement). Dans ce cas la loi théorique du paramètre est connue au niveau de la population.

• Les tests destinés à comparer plusieurs populations à l'aide d'un nombre équivalent d'échantillons (tests d'égalité ou d'homogénéité) sont les plus couramment utilisés. Dans ce cas la loi théorique du paramètre est inconnue au niveau des populations.

2.2 Notions générales

2.2.1 Hypothèse

Un test statistique permet de trancher entre deux hypothèses :

1- Hypothèse nulle notée $\mathbf{H}_0$: c'est l'hypothèse à tester.

2- Hypothèse alternative notée $\mathbf{H}_1$: hypothèse contraire.

2.2.2 Différentes formes d'hypothèses

Dans le cas d'un test paramétrique, Il existe deux sorte d'hypothèses : hypothèse simple et hypothèse multiple (composite).

- Une hypothèse simple est du type H : " $\theta = \theta_0$ ", où $\theta_0$est une valeur isolée du paramètre.$\theta_0 \in \Theta$ (Θ est un ensemble de touts les valeure de θ).
- Une hypothèse composite est du type $H : \theta \in I$, où I est une partie de $\mathbb{R}$ non réduite à un élément.

Remarque 2.2.1 *On dit que le test est :*

- *Bilatéral si elle est la forme : " $\theta \neq \theta_0$ ".*
- *Unilatéral à gauche si elle est la forme : " $\theta < \theta_0$ ".*
- *Unilatéral à droite si elle est la forme : " $\theta > \theta_0$ ".*

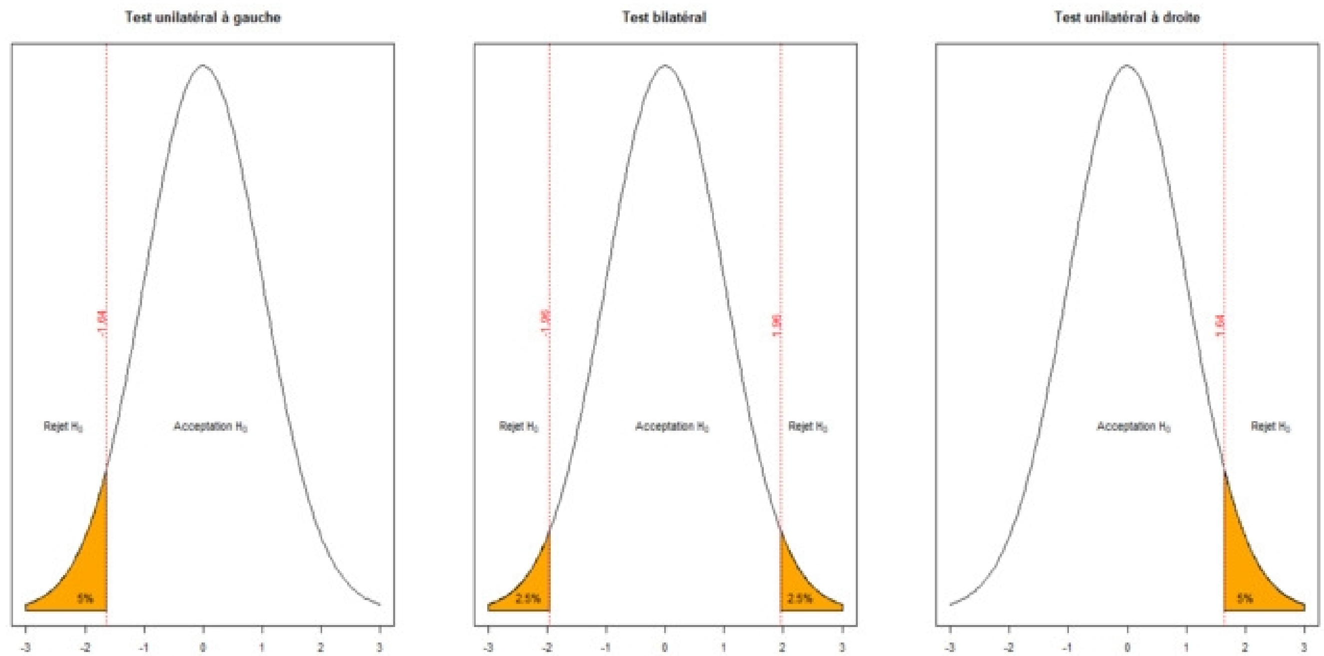

FIG. 2.1 – Les trois tests :" Unilatéral à gauche,Bélatéral,Unilatéral à droit "

2.2.3 Types d'erreurs

Lorsqu'on effectue un test statistique quatre situations sont possibles :

1) Accepter H_0 alors qu'elle est vraie.

2) Rejeter H_0 alors qu'elle est fausse.

3) Accepter H_0 alors qu'elle est fausse.

4) Rejeter H_0 alors qu'elle est vraie.

On distingue deux types de erreurs :

• Erreur de première espèce : on rejeter H_0 alors qu'elle est vraie.

• Erreur de deuxième espèce : on accepter H_0 alors qu'elle est fausse.

2.2.4 Risques de première et deuxième espèce

Définition 2.2.1 *On appelle risque d'erreur de première espèce, notée "α" la probabilité de rejeter H_0 et d'accepter H_1 alors que H_0 est vraie,telle que :*

$$\alpha = \mathbb{P}(rejeterH_0|H_0vraie) = \mathbb{P}(H_1|H_0).$$

Définition 2.2.2 *On appelle risque d'erreur de deuxième espèce ,notée "β" la probabilité de rejeter H_1et d'accepter H_0 ,alors que H_1 est vraie, telle que :*

$$\beta = \mathbb{P}(accepterH_0|H_0fausse) = \mathbb{P}(H_0|H_1).$$

En résumé, nous avons le tableau suivant :

vérité **décision**	H_0	H_1
H_0	$1-\alpha$	β
H_1	α	$1-\beta$

Tab. 2.1 – Les Risques associés à un test

Définition 2.2.3 $1-\alpha$ *et* $1-\beta$ *sont les bornes decision telle que :*

- $1-\alpha = \mathbb{P}(\overline{H_1}|H_0) = \mathbb{P}(H_0|H_0)$.
- $1-\beta = \mathbb{P}(\overline{H_0}|H_1) = \mathbb{P}(H_1|H_1)$.

Remarque 2.2.2 *Dans la pratique des tests statistiques on fixe α :*

$$\alpha \in \{0.05, 0.01, 0.1\}.$$

2.2.5 Région critique et région d'acceptation

Définition 2.2.4 *La région critique d'un test,est un ensemble des valeurs* $(x_1, x_2, ..., x_n)$ *de la variable de décision, qui conduisent à écarter* H_0 *au profit de* H_1*, et la note :*

$$W = \{(x_1, x_2, ..., x_n) \in \mathbb{R}^n / H_0 \text{ est rejetée}\}.$$

Elle est défini par :

$$\alpha = \mathbb{P}(W|H_0).$$

La région complémentaire est appelé région d'acceptation,est on la note $\overline{W}$*.*

Elle est défini par :

$$1 - \alpha = \mathbb{P}(\overline{W}|H_0).$$

Définition 2.2.5 *Pour un test d'hypothèse nulle multiple ,on appelle niveau du test (ou seuil du test)* α *si :*

$$\alpha = \sup_{\theta \in \Theta_0} \alpha(\theta).$$

2.2.6 La valeur p ou p-value

Pour conclure quant au rejet ou non de l'hypothèse nulle, il convient de comparer la réalisation de la statistique de test à la région critique. Une autre façon de conclure consiste à utiliser la valeur p. Nous avons vu, que pour une réalisation donnée de la statistique de test $T_n(x)$, la conclusion du test peut changer lorsque l'on modifie le niveau de risque. L'idée de la p-value consiste à déterminer le plus petit niveau pour lequel on peut rejeter l'hypothèse nulle.

Il existe une façon directe de déterminer la p-value .La règle est alors suivante :

Définition 2.2.6 *Suivant la nature du test (unilatéral ou bilatéral), la p-value associée à une réalisation $T_n(x)$ est égale à :*

Test unilatéral droit : p-value$= 1 - F_{T_n}(T_n(x))$.

Test unilatéral gauche : p-value $= F_{T_n}(T_n(x))$.

Test bilatéral : p-value $= 2 \times F_{T_n}(-|T_n(x)|)$.

où $F_{T_n}(.)$ désigne la fonction de répartition de la statistique de test T_n sous l'hypothèse nulle H_0.

règle de décision :

$$p - value < \ \alpha \Rightarrow \ rejet \ \ de \ H_0.$$

$$p - value > \ \alpha \Rightarrow \ ne \ rejet \ \ de \ H_0.$$

2.2.7 Puissance d'un test

Définition 2.2.7 *On appelle la puissance d 'un test la quantité " $1 - \beta$ ",qui correspond à la probabilité de rejeter H_0 quand H_1est vraie :*

$$Puissance = 1 - \beta = \mathbb{P}(W|H_1).$$

On peut aussi voir cela sur le graphique suivant :

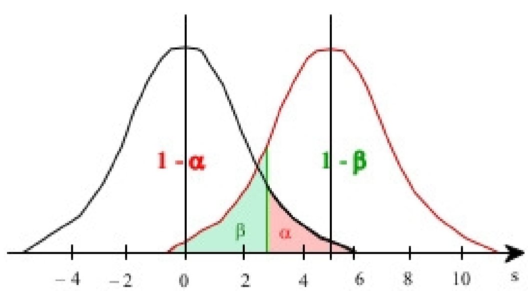

Fig. 2.2 – La relation entre les deux risques d'erreur

2.2.8 Fonction de puissance

Définition 2.2.8 *On appelle fonction puissance la fonction qui à une valeur θ du paramètre associe :*

$$\pi : \Theta \to \mathbb{R}.$$

$$\theta \mapsto \pi(\theta) = \mathbb{P}(rejet\ H_0).$$

la valeur de la fonction puissance pour : $\begin{cases} \theta \in \Theta_0 \text{ est le seuil du test} : \pi(\theta) = \alpha(\theta). \\ \theta \in \Theta_1 \text{ est la puissance } : \pi(\theta) = 1 - \beta(\theta). \end{cases}$

Exemple 2.2.1 *Soit $(X_1, ...X_n)$ un échantillon,de taille $n = 9$, d'une population normale d'espérance μ et de variance 1. Testons* $\begin{cases} H_0 : \mu \leqslant 5. \\ H_1 : \mu > 5. \end{cases}$ *. on considère la région critique $W = \{(x_1, ..., x_n) \in \mathbb{R}^n / \overline{X} > 5 + \frac{1}{\sqrt{9}}\}$. Pour un niveau $\alpha = 0.05$, la fonction puissance est définie par : $\pi(\mu) = \mathbb{P}(\overline{X} > 5 + 1.645(\frac{1}{\sqrt{9}}))$ pour $\overline{X} \rightsquigarrow N(\mu; \frac{1}{9})$.En posant $Z = \sqrt{9}(\overline{X} - \mu) \rightsquigarrow N(0; 1)$ on obtient :*

$$\pi(\mu) = \mathbb{P}(Z > 1.645 - \sqrt{9}(\mu - 5) = \mathbb{P}(Z > \sqrt{9}(5.55 - \mu) = 1 - \Phi(\sqrt{9}(5.55 - \mu)).$$

Cette valeur croît avec μ comme le montre la figure

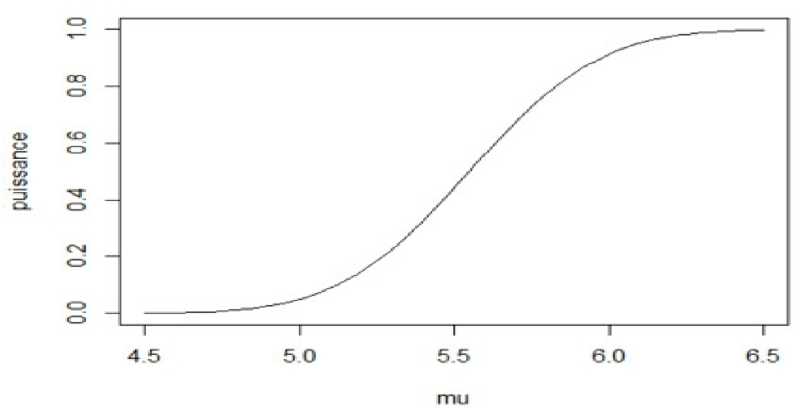

FIG. 2.3 – Fonction puissance pour $H_0 : \mu \leqslant 5$.sur une loi $N(0; 1)$

2.3 Tests de conformité

Le test de conformité consiste à confronter un paramètre calculé sur l'échantillon à une valeur pré-établie. Les plus connus sont certainement les tests portant sur la moyenne, la variance ou sur les proportions.

2.3.1 Tests sur la moyenne d'une population Normale

Cas où σ^2 est connu

Soit $X = (X_1, X_2, ..., X_n)$ un n-échantillon d'une loi gaussienne $N(\mu; \sigma^2)$. Alors la loi exacte de $\overline{X}$ est $N(\mu; \frac{\sigma^2}{n})$.on définit la statistique Z avec :

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

D'aprés la théorèm central limite la statistique Z suit une loi normale centrée réduite.$N(0; 1)$.

a) Pour une hypothèses bilatéral $\begin{cases} H_0 : \mu = \mu_0. \\ H_1 : \mu \neq \mu_0. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$\left| \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right| \geqslant z_{\frac{\alpha}{2}}.$$

où $z_{\frac{\alpha}{2}}$le quantile d'ordre $1 - \frac{\alpha}{2}$ de loi $N(0; 1)$ telle que :

$$\mathbb{P}(|Z| = \left| \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right| \geqslant z_{\frac{\alpha}{2}}) = \frac{\alpha}{2} \Rightarrow \mathbb{P}(Z \leqslant z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$$
$$\Rightarrow \Phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} \Rightarrow z_{\frac{\alpha}{2}} = \Phi^{-1}(1 - \frac{\alpha}{2}).$$

où Φ est la fonction de répartition de la loi normale centré réduite.

Donc la zone de rejet est :

$$I_{rejet} = \left]-\infty, \mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right[\cup \left]\mu_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, +\infty\right[.$$

b) Pour une hypothèses unilatéral à droite $\begin{cases} H_0 : \mu = \mu_0. \\ H_1 : \mu > \mu_0. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geqslant z_\alpha.$$

Donc la zone de rejet est :

$$I_{rejet} = \left]\mu_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, +\infty\right[.$$

Exemple 2.3.1 *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 :*

0.35	0.4	0.65	0.27	0.44	0.59	0.73	0.23
0.24	0.48	0.42	0.70	0.21	0.13	0.74	0.18

TAB. 2.2 – Les donnée de taux (dans le sang) d'une certaine substance interdite.

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

On testons : $\begin{cases} H_0 : \mu = \mu_0 = 0.4. \\ H_1 : \mu > \mu_0 = 0.4. \end{cases}$ *avec :* $\alpha = 0.05$

Sous l'hypothèse H_0, on obtient :

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{0.42 - 0.4}{\frac{0.2}{\sqrt{16}}} = 4.$$

Pour un niveau $\alpha = 0.05$,on trove : $z_\alpha = 1.96$ alors on rejette H_0 car : $Z = 4 \geqslant 1.96 = z_\alpha$.

Conclusion : le taux moyen dans le sang de la population des sportifs est non égal 0.4 *au risque* 5%.

c) Pour une hypothèses unilatéral à gauche $\begin{cases} H_0 : \mu = \mu_0. \\ H_1 : \mu < \mu_0. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leqslant -z_\alpha.$$

Donc la zone de rejet est :

$$I_{rejet} = \left]-\infty, \mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right[.$$

Cas où σ^2 est inconnu

Dans ce cas la variance σ^2 est estimée par son estimateur empirique $\widetilde{S^2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

On définit la statistique T avec :

$$T = \frac{\overline{X} - \mu}{\frac{\widetilde{S}}{\sqrt{n-1}}} \rightsquigarrow T_{n-1}.$$

Donc T suit une loi de Student à $(n-1)$ degrés de liberté.

a) Pour une hypothèses bilatéral $\begin{cases} H_0 : \mu = \mu_0. \\ H_1 : \mu \neq \mu_0. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$\left|\frac{\overline{X} - \mu}{\frac{\widetilde{S}}{\sqrt{n}}}\right| \geqslant t_{\frac{\alpha}{2}}.$$

où $t_{\frac{\alpha}{2}}$le quantile d'ordre $1 - \frac{\alpha}{2}$ de loi T_{n-1} telle que :

$$\mathbb{P}(|T_{n-1}| = \left|\frac{\overline{X} - \mu}{\frac{\widetilde{S}}{\sqrt{n}}}\right| \geqslant t_{\frac{\alpha}{2}}) = \frac{\alpha}{2} \Rightarrow \mathbb{P}(T_{n-1} \leqslant t_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$$
$$\Rightarrow F(t_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} \Rightarrow t_{\frac{\alpha}{2}} = F^{-1}(1 - \frac{\alpha}{2}).$$

où F est la fonction de répartition de la loi Student.

Donc la zone de rejet est :

$$I_{rejet} = \left]-\infty, \mu_0 - t_{\frac{\alpha}{2}} \frac{\widetilde{S}}{\sqrt{n}}\right[\cup \left]\mu_0 + t_{\frac{\alpha}{2}} \frac{\widetilde{S}}{\sqrt{n}}, +\infty\right[.$$

b) Pour une hypothèses unilatéral à droite $\begin{cases} H_0 : \mu = \mu_0. \\ H_1 : \mu > \mu_0. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$\frac{\overline{X} - \mu}{\frac{\widetilde{S}}{\sqrt{n}}} \geqslant t_\alpha.$$

Donc la zone de rejet est :

$$I_{rejet} = \left]\mu_0 + t_{\frac{\alpha}{2}} \frac{\widetilde{S}}{\sqrt{n}}, +\infty\right[.$$

c) Pour une hypothèses unilatéral à gauche $\begin{cases} H_0 : \mu = \mu_0. \\ H_1 : \mu < \mu_0. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$\frac{\overline{X} - \mu}{\frac{\widetilde{S}}{\sqrt{n}}} \leqslant -t_\alpha.$$

Donc la zone de rejet est :

$$I_{rejet} = \left]-\infty, \mu_0 - t_{\frac{\alpha}{2}} \frac{\widetilde{S}}{\sqrt{n}}\right[.$$

Remarque 2.3.1 *Si la variable X ne suit pas une loi normale les tests précédents s'appliquent encore dés que n est assez grand($n > 30$),en raison du théorème centrale limite.*

2.3.2 Tests sur la variance d'une population Normale

Cas où μ est connu

Lorsque la moyenne est connue,on prend comme variable de décision $V^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ telle que :

$$\frac{nV^2}{\sigma^2} \rightsquigarrow \chi_n^2.$$

où χ_n^2 est une loi du khi-deux à (n) degrés de libertés.

a) Pour une hypothèses bilatéral $\begin{cases} H_0 : \sigma^2 = \sigma_0^2. \\ H_1 : \sigma^2 \neq \sigma_0^2. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$\frac{nV^2}{\sigma^2} \leqslant l_1 \text{ ou } \frac{nV^2}{\sigma^2} \geqslant l_2.$$

où l_1 le quantile d'ordre $\frac{\alpha}{2}$, et l_2 le quantile d'ordre $1-\frac{\alpha}{2}$ de loi χ_n^2 telle que :

$$\begin{cases} \mathbb{P}(\chi_n^2 \leqslant l_1) = \frac{\alpha}{2} \\ \text{et} \\ \mathbb{P}(\chi_n^2 \geqslant l_2) = \frac{\alpha}{2} \end{cases} \Rightarrow \begin{cases} \mathbb{P}(\chi_n^2 \leqslant l_1) = \frac{\alpha}{2} \\ \text{et} \\ \mathbb{P}(\chi_n^2 \leqslant l_2) = 1-\frac{\alpha}{2} \end{cases} \Rightarrow \begin{cases} l_1 = F^{-1}(\frac{\alpha}{2}) \\ \text{et} \\ l_2 = F^{-1}(1-\frac{\alpha}{2}) \end{cases}.$$

où F est la fonction de répartition de la loi χ_n^2.

Donc la zone de rejet est :

$$I_{rejet} =]0, l_1[\cup]l_2, +\infty[\,.$$

b) Pour une hypothèses unilatéral à droite $\begin{cases} H_0 : \sigma^2 = \sigma_0^2. \\ H_1 : \sigma^2 > \sigma_0^2. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$\frac{nV^2}{\sigma^2} \geqslant l.$$

où l le quantile d'ordre $1-\alpha$ de loi χ^2_n.

Donc la zone de rejet est :

$$I_{rejet} =]l, +\infty[\,.$$

c) Pour une hypothèses unilatéral à gauche $\begin{cases} H_0 : \sigma^2 = \sigma_0^2. \\ H_1 : \sigma^2 < \sigma_0^2. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$\frac{nV^2}{\sigma^2} \leqslant l.$$

où l le quantile d'ordre α de loi χ^2_n.

Donc la zone de rejet est :

$$I_{rejet} = [0, l[\,.$$

Cas où μ est inconnu

Dans ce cas μ n'est pas connue,ona la statistique $S^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$.Alors :

$$\frac{nS^2}{\sigma^2} \rightsquigarrow \chi^2_{n-1}.$$

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

a) Pour une hypothèses bilatéral $\begin{cases} H_0 : \sigma^2 = \sigma_0^2. \\ H_1 : \sigma^2 \neq \sigma_0^2. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$\frac{nS^2}{\sigma^2} \leqslant k_1 \text{ ou } \frac{nS^2}{\sigma^2} \geqslant k_2.$$

où k_1 le quantile d'ordre $\frac{\alpha}{2}$,et $k_2$le quantile d'ordre $1-\frac{\alpha}{2}$ de loi χ^2_{n-1} telle que :

$$\left\{\begin{array}{c} \mathbb{P}(\chi^2_{n-1} \leqslant k_1) = \frac{\alpha}{2} \\ \text{et} \\ \mathbb{P}(\chi^2_{n-1} \geqslant k_2) = \frac{\alpha}{2} \end{array}\right. \Rightarrow \left\{\begin{array}{c} \mathbb{P}(\chi^2_{n-1} \leqslant k_1) = \frac{\alpha}{2} \\ \text{et} \\ \mathbb{P}(\chi^2_{n-1} \leqslant k_2) = 1-\frac{\alpha}{2} \end{array}\right. \Rightarrow \left\{\begin{array}{c} k_1 = F^{-1}(\frac{\alpha}{2}) \\ \text{et} \\ k_2 = F^{-1}(1-\frac{\alpha}{2}) \end{array}\right. .$$

où F est la fonction de répartition de la loi χ^2_{n-1}.

Donc la zone de rejet est :

$$I_{rejet} =]0, k_1[\, \cup \,]k_2, +\infty[\,.$$

b) Pour une hypothèses unilatéral à droite $\left\{\begin{array}{l} H_0 : \sigma^2 = \sigma_0^2. \\ H_1 : \sigma^2 > \sigma_0^2. \end{array}\right.$

La règle de Décision : on rejette H_0 au risque α si :

$$\frac{nS^2}{\sigma^2} \geqslant k.$$

où k le quantile d'ordre $1-\alpha$ de loi χ^2_{n-1}.

Donc la zone de rejet est :

$$I_{rejet} =]k, +\infty[\,.$$

c) Pour une hypothèses unilatéral à gauche $\left\{\begin{array}{l} H_0 : \sigma^2 = \sigma_0^2. \\ H_1 : \sigma^2 < \sigma_0^2. \end{array}\right.$

La règle de Décision : on rejette H_0 au risque α si :

$$\frac{nS^2}{\sigma^2} \leqslant k.$$

où k le quantile d'ordre α de loi χ^2_{n-1}.

Donc la zone de rejet est :

$$I_{rejet} = [0, k[\,.$$

Remarque 2.3.2 *Les résultats précédents ne sont valables que avec une population normale.*

2.3.3 Tests sur une proportion pour un grand échantillon

Soit p la proportion de la population possédant le caractère considéré.On considère un échantillon-d'individus de taille n de cette population,et soit la variable aléatoire $K_n$le nombre d'individus qui vérifier un certain caractère. La fréquence empirique $F = \frac{K_n}{n}$ est l'estimateur de p,et la loi de F peut être approchée par $N(p; \sqrt{\frac{p(1-p)}{n}})$.Ainsi d'après le théorème de centrale limite :

$$\frac{F-p}{\sqrt{\frac{p(1-p)}{n}}} \rightsquigarrow N(0;1).$$

a) Pour une hypothèses bilatéral $\begin{cases} H_0 : p = p_0. \\ H_1 : p \neq p_0. \end{cases}$

On la note par f la fréquence empirique de p_0.

La règle de Décision : on rejette H_0 au risque α si :

$$\left| \frac{f-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \right| \geqslant z_{\frac{\alpha}{2}}.$$

où $z_{\frac{\alpha}{2}}$ le quantile d'ordre $1-\frac{\alpha}{2}$ de loi $N(0;1)$.

donc la zone de rejet est :

$$I_{rejet} = \left] -\infty, p_0 - z_{\frac{\alpha}{2}}\sqrt{\frac{p_0(1-p_0)}{n}} \right[\cup \left] p_0 + z_{\frac{\alpha}{2}}\sqrt{\frac{p_0(1-p_0)}{n}}, +\infty \right[.$$

b) Pour une hypothèses unilatéral à droite $\begin{cases} H_0 : p = p_0. \\ H_1 : p > p_0. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$\frac{f-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \geqslant z_{\alpha}.$$

Donc la zone de rejet est :

$$I_{rejet} = \left] p_0 + \sqrt{\frac{p_0(1-p_0)}{n}}, +\infty \right[.$$

c) Pour une hypothèses unilatéral à gauche $\begin{cases} H_0 : p = p_0. \\ H_1 : p < p_0. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$\frac{f - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \leqslant -z_\alpha.$$

Donc la zone de rejet est :

$$I_{rejet} = \left] -\infty, p_0 - \sqrt{\frac{p_0(1-p_0)}{n}} \right[.$$

Exemple 2.3.2 *Sachant que* 20% *des personnes vaccinées par la formule* F *d'un vaccin présentent des allergies ou troubles secondaires, un laboratoire pharmaceutique propose une formule améliorée de ce vaccin* Fa *et espère diminuer le taux d'allergies et de troubles secondaires. Sur un échantillon de* 400 *personnes prises au hasard et ayant opté pour la formule*Fa *on observe* 60 *cas d'allergies ou troubles secondaires. On veut tester si la formule améliorée apporte réellement un bénéfice.*

On testons :$\begin{cases} H_0 : p = p_0. \\ H_1 : p < p_0. \end{cases}$ *avec :* $\alpha = 0.05$

Sous l'hypothèse H_0, *on obtient :*

$$Z = \frac{f - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{60}{400} - 0.2}{\sqrt{\frac{0.2(1-0.2)}{400}}} = -2.5.$$

Pour un niveau $\alpha = 0.05$,*on trove :* $-z_\alpha = -1.65$ *alors on rejette* H_0 *car :*$Z = -2.5 \leqslant -1.65 = z_\alpha$.

Conclusion : le taux d'allergies ou troubles secondaires est diminué par la formule Fa *du vaccin au risque* 5%.

2.4 Tests de d'homogénéité

Le test d'homogénéité il vise à comparer deux ou plus de deux séries d'observations afin de savoir si leurs distributions sont ou ne sont pas équivalentes.

2.4.1 Tests de comparaisons de deux moyennes de deux populations Normale

Soient X_1 et X_2 deux variables définies sur deux population,de moyenne noté μ_1 et μ_2 et de variancenoté σ_1^2et σ_2^2 respectivement.On prélève dans chacune des populations un échantillon et noton respectivement n_1 et n_2 les tailles de ces échantillons indépendants.

Cas où σ est connu

On a pour $i = 1, 2$:

$$\overline{X_i} \rightsquigarrow N(\mu_i; \frac{\sigma_i^2}{n_i}).$$

On définit la statistique Z avec :

$$Z = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightsquigarrow N(0; 1).$$

a) Pour une hypothèses bilatéral $\begin{cases} H_0 : \mu_1 = \mu_2. \\ H_1 : \mu_1 \neq \mu_2. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$|Z| = \left| \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right| \geqslant z_{\frac{\alpha}{2}}.$$

où $z_{\frac{\alpha}{2}}$ le quantile d'ordre $1 - \frac{\alpha}{2}$ de loi $N(0; 1)$.

Donc la zone de rejet est :

$$I_{rejet} = \left]-\infty, -z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right[\cup \left] z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}, +\infty\right[.$$

b) Pour une hypothèses unilatéral à droite $\begin{cases} H_0 : \mu_1 = \mu_2. \\ H_1 : \mu_1 > \mu_2. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$Z = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \geqslant z_{_\alpha}.$$

où $z_{_\alpha}$ le quantile d'ordre $1-\alpha$ de loi $N(0;1)$.

Donc la zone de rejet est :

$$I_{rejet} = \left] z_\alpha\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}, +\infty\right[.$$

c) Pour une hypothèses unilatéral à gauche $\begin{cases} H_0 : \mu_1 = \mu_2. \\ H_1 : \mu_1 < \mu_2. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$Z = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \leqslant -z_{_\alpha}.$$

où $z_{_\alpha}$ le quantile d'ordre $1-\alpha$ de loi $N(0;1)$.

Donc la zone de rejet est :

$$I_{rejet} = \left]-\infty, -z_\alpha\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right[.$$

Cas où σ est inconnu et $\sigma_1^2 = \sigma_2^2$

i) Si $n_1$ou $n_2 < 30$:

On a pour $i = 1, 2$:

$$\frac{n_i S_i^2}{\delta^2} \rightsquigarrow \chi^2_{n_i - 1} \ et \ \overline{X_i} \rightsquigarrow N(\mu_i; \frac{\sigma^2}{n_i}).$$

on pose $S^2 = \dfrac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$,et ona la statistique T :

$$T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \rightsquigarrow T_{n_1 + n_2 - 2}.$$

où S_1^2 et S_2^2 sont les estimateurs de σ_1^2 et σ_2^2.

Donc T suit une loi de Student à $n_1 + n_2 - 2$ degrés de liberté.

a) Pour une hypothèses bilatéral $\begin{cases} H_0 : \mu_1 = \mu_2. \\ H_1 : \mu_1 \neq \mu_2. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$\left| T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right| \geqslant t_{\frac{\alpha}{2}}.$$

où $t_{\frac{\alpha}{2}}$le quantile d'ordre $1 - \frac{\alpha}{2}$ de loi $T_{n_1 + n_2 - 2}$.

Donc la zone de rejet est :

$$I_{rejet} = \left] -\infty, -t_{\frac{\alpha}{2}} \times S\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right[\cup \left] t_{\frac{\alpha}{2}} \times S\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, +\infty \right[.$$

b) Pour une hypothèses unilatéral à droite $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \geqslant t_\alpha.$$

où t_α le quantile d'ordre $1-\alpha$ de loi $T_{n_1+n_2-2}$.

Donc la zone de rejet est :

$$I_{rejet} = \left]t_\alpha \times S\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, +\infty\right[.$$

c) Pour une hypothèses unilatéral à gauche $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \leqslant -t_\alpha.$$

où t_α le quantile d'ordre $1-\alpha$ de loi $T_{n_1+n_2-2}$.

Donc la zone de rejet est :

$$I_{rejet} = \left]-\infty, -t_{\frac{\alpha}{2}} \times S\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right[.$$

Exemple 2.4.1 *Pour tester un nouveau médicament qui réduit le cholestérol total. Un test a été effectué, un total de* 30 *échantillons qui ont été sélectionnés au hasard,*15 *personnes ont testé le nouveau médicament tandis que les* 15 *autres ont testé un placebo pendant* 6 *semaines pour chaque groupe.*

Les résultats du taux de cholestérol total de chaque patient sont indiqués dans le tableau ci-dessous :

Traitement	Taille de l'échantillon	Moyenne	Écart-type
Nouveau médicament	15	195.9	28.7
Placebo	15	227.4	30.3

TAB. 2.3 – Statistiques pour mesurer le taux de cholestérol

Nous vérifierons statistiquement si le cholestérol total moyen a diminué chez les patients prenant le nouveau médicament par rapport aux participants prenant le placebo.

On testons : $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$ *;avec :* $\alpha = 0.05$. *Comme les deux échantillons sont petits* (< 30)*, nous utilisons la statistique du test* T *dans lequel :*

$$T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$On\ a : S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} = \frac{15(28.7^2 + 30.3^2)}{15 + 15 - 2} = 933.1.$$

$$donc : T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{195.9 - 227.4}{\sqrt{933.1\left(\frac{1}{15} + \frac{1}{15}\right)}} = \frac{-31.5}{11.14} = -2.73.$$

Nous avons besoin de degrés de liberté(ddl), définis par : ddl $= n_1 + n_2 - 2 = 15 + 15 - 2 = 28$.

Pour un niveau $\alpha = 0.05$ *et ddl* $= 28$ *ona :* $t_\alpha = -1.701$ *alors on rejette* H_0 *car* $T = -2.73 \leqslant t_\alpha = -1.701$.

Enfin nous avons des preuves statistiquement pour un niveau $\alpha = 0.05$,*pour montrer que le taux moyen de cholestérol total est plus faible chez les patients prenant le nouveau médicament pendant* 6 *semaines que chez les patients prenant un placebo.*

ii) Si $n_1$et $n_2 > 30$: on a la statistique T :

$$T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \rightsquigarrow N(0;1).$$

2.4.2 Tests de comparaisons de deux variance de deux populations Normale

La statistique de test est définie par :

$$F = \frac{n_1 S_1^2}{(n_1-1)\,\sigma_1^2} \times \frac{(n_2-1)\,\sigma_2^2}{n_2 S_2^2} \rightsquigarrow \mathcal{F}\,(n_1-1; n_2-1)\,.$$

$$\text{on a pour } i=1,2 : \frac{n_i S_i^2}{(n_i-1)\,\sigma_i^2} \rightsquigarrow \chi^2_{n_i-1}.$$

Sous l'hypothèse H_0, on obtient :

$$F = \frac{n_1 S_1^2}{(n_1-1)} \times \frac{(n_2-1)}{n_2 S_2^2} \rightsquigarrow \mathcal{F}\,(n_1-1; n_2-1)\,.$$

où $\mathcal{F}\,(n_1-1; n_2-1)$ est une loi de Fisher à (n_1-1, n_2-1) degrés de libertés.

a) Pour une hypothèses bilatéral $\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2. \\ H_1 : \sigma_1^2 \neq \sigma_2^2. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$F \leqslant C_1 ou F \geqslant C_2. \text{ où } \mathbb{P}(\mathcal{F}\,(n_1-1; n_2-1) \leqslant C_1) = \mathbb{P}(\mathcal{F}\,(n_1-1; n_2-1) \geqslant C_2) = \frac{\alpha}{2}.$$

Exemple 2.4.2 *On veut comparer les variances de deux populations* $X \rightsquigarrow N(\mu_1; \sigma_1^2)$*et* $Y \rightsquigarrow N(\mu_2; \sigma_2^2)$ *. On dispose de deux échantillons de taille* 25 *et* 30 *respectivement, avec des variances empirique* $S_1^2 = 121.2$ *et* $S_2^2 = 53.8$.*On testons :*$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2. \\ H_1 : \sigma_1^2 \neq \sigma_2^2. \end{cases}$ *;avec :* $\alpha = 0.05$

Sous l'hypothèse H_0, *on obtient :*

$$F = \frac{n_1 S_1^2}{(n_1-1)} \times \frac{(n_2-1)}{n_2 S_2^2} = \frac{25 \times 121.2}{(25-1)} \times \frac{(30-1)}{30 \times 53.8} = 2.27.$$

Pour un niveau $\alpha = 0.05$,*on trove :*$C_1 = 2.15; C_2 = 0.45$ *alors on rejette* H_0 *car :*$F = 2.27 \geqslant$

$0.45 = C_2$ *on conclut Il y a une différence sur les deux variance .*

b) Pour une hypothèses unilatéral à droite $\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2. \\ H_1 : \sigma_1^2 > \sigma_2^2. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$F \geqslant C.\text{où C est solution de : } \mathbb{P}(\mathcal{F}\,(n_1 - 1; n_2 - 1) \geqslant C) = \alpha.$$

c) Pour une hypothèses unilatéral à gauche $\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2. \\ H_1 : \sigma_1^2 < \sigma_2^2. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$F \leqslant d.\text{où d est solution de : } \mathbb{P}(\mathcal{F}\,(n_1 - 1; n_2 - 1) \leqslant d) = \alpha.$$

2.4.3 Tests de comparaison de deux proportions, pour de grands échantillons

On veut comparer deux proportions p_1 et p_2 à partir de deux échantillons de tailles n_1 et n_2 (respectivement), dans la pratique les tailles sont supposées suffisamment grandes : $n_1 > 30$ *et* $n_2 > 30$.et $n_1.p_1(1 - p_1) > 5$ et $n_2.p_2(1 - p_2) > 5$.

Soient F_1 et F_2 deux fréquences empiriques associée à l'échantillons de $p_1$et p_2(respectivement). Ainsi les fréquences empiriques F_1 et F_2 sont indépendantes et de lois normales, et on a pour $i = 1, 2 : F_i \rightsquigarrow N(p_i; \frac{p_i(1-p_i)}{n_i})$.

$$donc : F_1 - F_2 \rightsquigarrow N(p_1 - p_2; \frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}).$$

Sous l'hypothèse $H_0, (p_1 = p_2 = p)$ on obtient :

$$U = \frac{F_1 - F_2}{\sqrt{p(1 - p)(\frac{1}{n_1} + \frac{1}{n_2})}} \rightsquigarrow N(0; 1).$$

Si p est inconnue on la remplace par son estimation :

$$\widehat{p} = f = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}.où\ f_1\ et\ f_2 \text{sont les estimations de } p_1\ et\ p_2.$$

a) Pour une hypothèses bilatéral $\begin{cases} H_0 : p_1 = p_2. \\ H_1 : p_1 \neq p_2. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$\left| U = \frac{f_1 - f_2}{\sqrt{f(1-f)(\frac{1}{n_1} + \frac{1}{n_2})}} \right| \geqslant z_{\frac{\alpha}{2}}.$$

Exemple 2.4.3 *Les données du tableau suivant représentent les mesures de glycémie pour* 120 *patients : Deux médicaments A et B en concurrence ont été administrés à deux groupes distincts*

Glycémie mg/dL	**Centre de la classe**	**Effectif**
[50-100[	75	19
[100-150[	125	25
[150-200[	175	41
[200-250[	225	18
[250-300[	275	10
[300-350[	325	7

TAB. 2.4 – Le taux de sucre dans le sang est pour 120 patients

de 60 *personnes atteintes glycémie. Les patients sont suivis de façon absolument identique. Après* 15 *jours, on remarque que* 55 *personnes parmi les* 60 *traitées par A ont guéri alors qu'il y en a* 51 *guéries parmi les* 60 *personnes traitées par B. On veut tester au risque* 5% *si les deux médicaments sont significativement différents.*

On testons : $\begin{cases} H_0 : p_1 = p_2. \\ H_1 : p_1 \neq p_2. \end{cases}$ *;avec :* $\alpha = 0.05$ *On calcule* f_1 *et* f_2 *:*

$$f_1 = \frac{55}{60} = 0.92. \quad ; \quad f_2 = \frac{51}{60} = 0.85.$$

Les deux échantillons n_1 et n_2 est grandes ($n_1 = n_2 = 60 > 30$) donc :

$$f = \frac{60 \times 0.92 + 60 \times 0.85}{60 + 60} = \frac{106.2}{120} = 0.88.$$

On a la statistique du test U dans lequel :

$$U = \frac{f_1 - f_2}{\sqrt{f(1-f)(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0.92 - 0.85}{\sqrt{0.88(1-0.88)(\frac{1}{60} + \frac{1}{60})}} = \frac{0.07}{0.059} = 1.19.$$

Pour un niveau $\alpha = 0.05$ le quantile $z_{\frac{\alpha}{2}} = 1.96$ alors on ne rejette pas H_0 car : $|U| = 1.19 \leqslant z_{\frac{\alpha}{2}} = 1.96$. Enfin les deux médicaments ne sont pas significativement différents en termes d'efficacité au risque 5%.

b) Pour une hypothèses unilatéral à droite $\begin{cases} H_0 : p_1 = p_2. \\ H_1 : p_1 > p_2. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$U = \frac{f_1 - f_2}{\sqrt{f(1-f)(\frac{1}{n_1} + \frac{1}{n_2})}} \geqslant z_\alpha.$$

c) Pour une hypothèses unilatéral à gauche $\begin{cases} H_0 : p_1 = p_2. \\ H_1 : p_1 < p_2. \end{cases}$

La règle de Décision : on rejette H_0 au risque α si :

$$U = \frac{f_1 - f_2}{\sqrt{f(1-f)(\frac{1}{n_1} + \frac{1}{n_2})}} \leqslant -z_\alpha.$$

Chapitre 3

Partie D'application

Nous allons examiner dans cette section quelques unes des commandes de R, numériques ou graphiques.Ces commandes sont très utiles pour mettre en évidence de manière descriptive certaines particularités des données. Ils permettent d'avoir une première impression d'ensemble sur les données. Les résultats numériques et graphiques obtenus aident beaucoup au choix de modèles et aux hypothèses que l'on sera amené à formuler pour poursuivre une étude statistique.R contient une multitude de fonctions pour calculer des fonctions statistiques de base sur des échantillons de données, qu'elles soient numériques ou catégorielles.

Nous résolvons l'exemple précédent 2.4.3 avec langage R :

3.1 Application sur le 1er chapiter :

En langage R :

```
> X=c(75,125,175,225,275,325)

> N=c(19,25,41,18,10,7)

> Y=rep(X,N)

> mean(Y)

[1] 173.3333
```

```
> n=sum(N)
> n
[1] 120
> median(Y)# le mediane
[1] 175
> var(Y) # la variance
[1] 4661.064
> sqrt(var(Y))# l'écart type
[1] 68.27199
> n*var(Y)/(n-1)# estimateur sans biais de la variance
[1] 4700.233
> sqrt(n*var(Y)/(n-1))#S2
[1] 68.55825
```

3.2 Application sur le 2em chapiter :

En langage R :

```
> tabac <- matrix(c(05, 55, 09, 51), 2)
> colnames(tabac) = c("A", "B")
> rownames(tabac) = c("non_guéries", "guéries")
> tabac
             A  B
non_guéries  5  9
guéries      55 51
```

Si nous appliquons la valeur de la statistique du test a ces données :

$U = \frac{f_1 - f_2}{\sqrt{f(1-f)(\frac{1}{n_1}+\frac{1}{n_2})}}$

Sous R :

```
> eg.prop<-function (m1, n1, m2, n2)
+ {
+ risque=0.05
+ f1=round(m1/n1,2)
+ f2=round(m2/n2,2)
+ f <- round((n1*f1 + n2*f2)/(n1 + n2),2)
+ u <- abs(round((f1 -f2)/sqrt(f * (1 - f) * (1/n1 + 1/n2)),2))
+ p_value <- 2 * (1 - pnorm(u))
+ cat("\n")
+ cat("Test de l'égalité de 2 proportions", "\n")
+ cat("\n")
+ cat("Valeur de U : ", round(u,2), "\n")
+ cat("\n")
+ cat("p-value observé, test bilatéral : ", round(p_value * 100, 2), "%", "\n")
+ cat("\n")
+ if (risque > p_value)
+ {
+ cat("On rejette H0 alors les deux médicaments significativement différents en termes
d'ffcacité au risque
+ 5% ", "\n")
+ }
+ else
+ {
```

```
+ cat("On ne rejette pas H0 alors les deux médicaments ne sont pas significativement
différents en termes d'ffcacité au risque
+ 5%", "\n")
+ }
+ }
> eg.prop(55,60,51,60)
Test de l'égalité de 2 proportions
Valeur de U : 1.18
p-value observé : 23.8 %
```

On ne rejette pas H0 alors les deux médicaments ne sont pas significativement différents en termes d'ffcacité au risque

5%

Nous appliquons la fonction prop.test pour comparer deux proportions observées est :

```
> prop.test(c(55, 51), c(60, 60), correct = F)
 2-sample test for equality of proportions without continuity correction
data : c(55, 51) out of c(60, 60)
X-squared = 1.2938, df = 1, p-value = 0.2553
alternative hypothesis : two.sided
95 percent confidence interval :
 -0.04758676 0.18092010
sample estimates :
 prop 1 prop 2
0.9166667 0.8500000
```

Conclusion

Dans ce travail Nous avons appris à effectuer un certain nombre de test paramétrique et il existe d'autres. Tous fonctionnent sur le même principe. On a basé sur trois chapiter :

Premièrement, on a définit les différents paramètres statistique et sur quelques éléments fondamentau qu'on peut utilse sur le tests.

Deuxièmement, on a parlé sur les principe de tests et on a étudie les tests de conformité et comment comparé une moyenne et une variance aussi une proportion.Plus de ça, les tests d'homogénéité et comment comparer les moyennes et les va-riances aussi les proportions.

Troisièmement, la partie pratique en utilisant le langage R pour faire les tests et vérifier les résultats.

En fin,tout ceci pour faire un test statistique. Il nous permet d'indiquer si oui ou non.

.

Bibliographie

[1] Goldfarb, B., & Pardoux, C. (2011). Introduction à la méthode statistique : manuel et exercices corrigés. Dunod. .

[2] Chekroun, A. Statistiques descriptives et exercices.

[3] Leboucher, L., & Voisin, M. J. (2013). Introduction à la statistique descriptive : cours et exercices avec tableur. Cépaduès éd..

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

[5] Veysseyre, R. (2014). Aide-mémoire-Statistique et probabilités pour les ingénieurs. Dunod. .

[6] Hubler, J. (2007). Statistique descriptive appliquée à la gestion et à l'économie. Editions Bréal.

[7] Velenik, Y. (2011). Probabilités et statistique. Université de Geneve.

[8] Cottet-Emard, F. (2014). Probabilités et tests d'hypothèses. De Boeck Superieur.

[9] Lejeune, M. (2004). Statistique : La théorie et ses applications. Springer Science & Business Media.

[10] JACQUES, J. Statistiques inférentielles.

[11] Monbet, V. (2009). Tests statistiques Notes de cours. Cité en, 133.

[12] AKAKPO, N. (2017). Tests statistiques. Notes de cours issues du module 4M018 Statistique Appliquée. Polycopié disponible sur la page web de l'auteur.

[13] CHAMPELY, S. (2006). Tests statistiques paramétriques : Puissance, taille d'e et et taille d'échantillon (sous R).

[14] JOLION, J. M. (2006). Probabilités et Statistique

[15] H,Schyns. (2010) .Paramètres de position et de dispersion.http ://notesde-cours.drivehq.com/courspdf/StatParam.pdf

[16] Grammont, L.(2003).Statistiques inférentielles Licence d'économie et de gestion.https ://dossier.univ-st-etienne.fr/fac-sciences-maths/www/stat03.pdf.

[17] ELMARHOUM,A. Echantillonnage et estimation. http ://fsjes-agdal.um5.ac.ma/sites/fsjes-agdal.um5.ac.ma/files/Cours%20Echantillonnage%20et%20Estimation.pdf

[18] Dusart, P. (2018) .Statistiques inférentielles. https ://www.unilim.fr/pages_perso/pierre.dusart/Probas/c

[19] LENOIR,J.Les tests d'hypothèse.https ://www.imo.universite-paris-saclay.fr/~pansu/web_ifips/Tests.pdf.

[20] Grammont, L..(2013). Cours de statistique inferentielles Licence d'économie et de gestion. https ://dossier.univ-st-etienne.fr/fac-sciences-maths/www/stat03.pdf.

[21] Ruch,J.J.(2013). Statistique : tests d'hypothèses. Préparation à l'Agrégation Bordeaux, 11.

[22] Brabant, P, Dillmann, C , Legrand ,J.,Manicacci, D. Marchadier,E. ,Ollier S. ,Sicard D. ,De Vienne, D..Biostatistiques http ://moulon.inra.fr/modelstat/images/poly_biostat_sept18.pdf

[23] Bouraine, M., Berdjoudj, L.,(2013), Statistique Inférentielle. https ://cours-examens.org/ images/An_2017_1/ Etudes_superieures/ statistiques/Alg%C3%A9rie/ mi_lessons06_stat_inferentielle.pdf

[24] Ycart, B. (2002). Estimation paramétrique, tests statistiques. Centre de Publication Universitaire.

Annexe A : Logiciel *R*

3.3 Qu'est-ce-que le langage R ?

• Le langage **R** est un langage de programmation et un environnement mathématique utilisés pour le traitement de données. Il permet de faire des analyses statistiques aussi bien simples que complexes comme des modèles linéaires ou non-linéaires, des tests d'hypothèse, de la modélisation de séries chronologiques, de la classification, etc. Il dispose également de nombreuses fonctions graphiques très utiles et de qualité professionnelle.

• **R** a été créé par Ross Ihaka et Robert Gentleman en 1993 à l'Université d'Auckland, Nouvelle Zélande, et est maintenant développé par la R Developement Core Team.

L'origine du nom du langage provient, d'une part, des initiales des prénoms des deux auteurs (Ross Ihaka et Robert Gentleman) et, d'autre part, d'un jeu de mots sur le nom du langage S auquel il est apparenté.

Annexe B : Abréviations et Notations

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

Abreviations ou Notations		Explication
Ω	:	L'ensemble sur lequel porte notre étude statistique
n	:	La taille d'un échantillon.
N	:	La taill d'une population
ω	:	Individu.
$v.s$	:	Variable statistique
X	:	Caractère.
$v.a.r$	:	Variable aléatoire réelle.
$\mathbb{R}$	:	Ensemble des valeurs réelles
I	:	Un sous-ensemble de $\mathbb{R}$
$X^{-1}(I)$	:	Un sous-ensemble de Ω
F	:	Fonction de répartition
f	:	Fonction de densitée.
$\mathbb{P}$	:	Probabilité
X_n	:	Suite de variables aléatoires
μ	:	Espérance ou moyenne d'une va

σ	:	Ecart-type.
σ^2	:	La variance.
$\rightsquigarrow$	:	Converge en distribution.
$N(0;1)$	:	Loi normale standard.
f_i	:	La fréquence
$\overline{X}$	:	La moyenne arithmétique
$\widetilde{X}$	:	La médiane
$Q(.)$	:	Fonction quantile
H_0	:	Hypothèse nulle
H_1	:	Hypothèse alternative
Θ	:	Ensemble de touts les valeure de θ
α	:	Risque d'erreur de première espèce.
β	:	Risque d'erreur de deuxième espèce.
W	:	La région critique.
$\overline{W}$	:	La région d'acceptation.
$1-\beta$	:	La puissance d 'un test.
π	:	Fonction de puissance.
Φ	:	La fonction de répartition de la loi normale centré réduite.
$\widetilde{S^2}$	:	Estimateur empirique de la variance σ^2
T	:	Statistique suit une loi de Student
ddl	:	degrés de libertés
χ^2_n	:	Loi du khi-deux à (n) degrés de libertés
p	:	La proportion de la population
$\mathcal{F}$	:	Est une loi de Fisher.
$\widehat{p}$	:	Estimation de p

اختبار فرضية إحصائية

ملخص :

يهدف هذا البحث إلى دراسة موضوع فحص الفرضيات ودقة الاختبارات الوسيطية وذلك بشرح نوعين رئيسيين هما : فحص المطابقة وفحص التجانس.

قدمنا أيضًا أمثلة لشرح الطرق التي تم النظر فيها لإكمال الدراسة النظرية.

الكلمات المفتاحية: *الاختبارات الوسيطية،* فحص المطابقة ، فحص التجانس.

Test d'hypothèse statistique

Résumé:

Cette recherche vise à étudier les tests d'hypothèse, précisément, les tests paramétriques et à expliquer les deux types; test d'homogénéité et test de conformité.

Et des exemples d'illustration des méthodes étudiées complètent l’étude théorique.

Mots clés : tests paramétriques , test d'homogénéité, test de conformité.

Statistical hypothesis testing

Summary :

This research aims to study the hypothesis tests, precisely, the parametric tests and explain the two types ; homogeneity test and compliance testing.

We also provided examples to explain the methods considered to complete the theoretical study.

Key words : parametric tests, homogeneity test, compliance testing.