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

ZAHNIT Samiha

Titre :

Tests de normalité des données

Membres du Comité d'Examen :

Dr.	DHIABI Samra	UMKB	Encadreur
Dr.	TOUBA Sonia	UMKB	Président
Dr.	DJABER Ibtisem	UMKB	Examinateur

Septembre 2020

Dédicace

À vant tout, je tiens à remercier "ALLAH", est l'Unique qui m'offre le courage et la volonté nécessaire pour affronter les difficultés de la vie.

À ma chère mère.

À mon chère père que Dieu ait pitié de lui.

À tout ma famille.

Remerciements

Tout d'abord merci au **mon Dieu** le tout puissant, de m'avoir donné la force, la patience et la volonté pour réaliser ce travail dans des meilleures circonstances et en bon état.

Je teint à remercier mon encadreur Mademoiselle **DHIABI Samra**.

J'adresse mes remerciements aux présidents et membres du Jury qui ont accepté d'examiner ce mémoire en lui apportant de l'intérêt.

Je remercie l'ensemble des enseignants du département Mathématiques.

Enfin, Je voudrais associer à mes remerciements toutes les personnes qui ont contribué de prés ou de loin à l'aboutissement de ce travail.

Table des matières

Table des figures

Liste des tableaux

Introduction

Un test d'adéquation permet de statuer sur la compatibilité d'une distribution observée avec une distribution théorique associée à une loi de probabilité. Il s'agit de modélisation. Nous résumons une information brute, une série d'observation, à l'aide d'une fonction paramétrée.

Les tests de normalité sont des cas particuliers des tests prennent une place importante en statistique. En effet, de nombreux tests supposent la normalité des distributions pour être applicables. En toute rigueur, il est indispensable de vérifier la normalité avant d'utiliser les tests. Cependant de nombreux tests sont suffisamentrobustes pour être utilisables même si les distributions s'écartent de la loi normale.

L'objectif de ce mémoire est l'étude de ce type de tests. Il est composé de trois chapitres :dans le premier chapitre, on va présenter des généralités sur les tests d'hypothèses, nous donnons quelque concepts de base tests statistiques, les types des tests, le niveau de singnification du test et d'autres concepts de base.

Ensuite, dans le deuxième chapitre, nous présenterons les téchniques descriptives, notamment : méthode d'histogramme, boite à moustache et Q-Q plot,...etc, qui peuvent permettre de visualiser si la distribution empirique suit une loi normale.

Enfin, dans le troisième chapitre, on va présenter quelque tests pour affirmer la normalité d'une distribution comme (test de shapiro-wik, Cramer-von mises et de Jarque bera...etc). Tous ces test ont en comun d'avoir comme hypothèse nulle : la distribution empirique suit une loi gaussienne.

Pour la simulation des données trouvées dans ce mémoire, nous utilisons le logiciel R.

Chapitre 1

Tests d'hypothèses, Généralités

Dans ce chapitre nous énonçons (ou rappelons) un certain nombre de généralités autour des tests d'hypothèse, l'objectif étant d'être capable de bien formuler un test.

1.1 Les tests statistiques

*Les tests statistiques sont des méthodes de la statistique inférentielle, qui permettent d'analyser des obttenus par tirages ou hasard. Ils consistent à généraliser les propriétés constatées sur des observations à la population d'où ces dernières sont extraites. On a deux types des tests statistiques sont : les tests paramétriques et les tests non paramétriques.

1.1.1 Tests Paramétriques

*Un test est dit paramétrique l'orsqu'on fait l'hypothèse que les obsevations qui décrivent les individus sont tirées de distribution dépendent d'un nombre fini de paramétre (dans ce cas la distribution connue), de plus on utilise ses tests lorsque les données sont quantitatives.

1.1.2 Test non parametrique

*L'orsqu'on n'impose pas de distribution sur ces variables on sera dans le cadre de la statistique non paramertrique, de plus on utilise ses tests lorsque les données sont qualitatives.

Parmi les tests non paramétrique, on a les tests d'ajustememt ou d'adéquation. Par exemple les tests de poisson et tests de normalité.

1.2 Principe des tests

1.2.1 Hypothèses de test

*En premier lieu, nous devons formuler les hypothèses. L'hypothèse que nous voulons vérifier sera appelée **hypothèse nulle** et on la notera H_0. Il est l'hypothèse que l'on désire contrôler : elle consiste à dire qu'il n'existe pas de différence entre les paramètres comparés ou que la différence observée n'est pas significative et est due aux fluctuations d'échantillonnage. Cette hypothèse est formulée dans le but d'être rejetée.

D'autre part **L'hypothèse alternative** notée H_1 est la négation de H_0, elle est équivalente à dire « H_0» est fausse. La décision de rejeter H_0 signifie que H_1 est réalisée ou H_1 est vraie.

Il existe une **dissymétrie importante** dans les conclusions des tests. En effet, la décision d'accepter H_0 n'est pas équivalente à « H_0 est vraie et H_1 est fausse ». Cela traduit seulement l'opinion selon laquelle, il n'y a pas d'évidence nette pour que H_0 soit fausse.

1.2.2 Test unilatéral ou bilatéral

La nature de H_0 détermine la façon de formuler H_1 et par conséquence la nature **unilatérale** ou **bilatérale** du test.

Test bilatéral

*Un test bilatéral s'applique quand vous cherchez une différence entre deux paramètres, ou entre un paramètre et une valeur donnée sans se préoccuper du signe ou du sens de la différence. Dans ce cas, la zone de rejet de l'hypothèse principale se fait de part et d'autre de la distribution de référence.

Test unilatéral

*Un test unilatéral s'applique quand vous cherchez à savoir si un paramètre est supérieur (ou inférieur) à un autre ou à une valeur donnée. La zone de rejet de l'hypothèse principale est située d'un seul côté de la distribution de probabilité de référence.

Exemple 1.1 *Vous pouvez avoir comme hypothèse nulle :*

$$H_0 : \text{La moyenne de la population est égale à } \mu_0$$

et, dans ce cas, une hypothèse alternative pourrait être :

$$H_1 : \text{La moyenne de la population est différente de } \mu_0$$

Ou encore :

$$H_1 : \text{La moyenne de la population est strictement plus grande que } \mu_0$$

On dit que le test est **bilatéral** c'est la manière d'écrit ces hypothèses est :

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

C'est le test est **unilatéral**, la manière d'écrit ces hypothèses est :

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

Ou bien :

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

1.2.3 Les deux espèces d'erreur

*Lorsque l'on fait un test d'hypothèse, deux sortes d'erreur sont possibles. On peut rejeter l'hypothèse nulle alors qu'elle est vraie. Ceci se produit si la valeur de la statistique de test tombe dans la région de rejet alors que l'hypothèse H_0 est vraie.
La probabilité de cet évènement est le niveau de signification. On dira aussi que le niveau de signifcation est la probabilité de rejeter l'hypothèse nulle à tort.
Le risque α de première espèce est celui de **rejeter** H_0 alors qu'elle est vraie.

$$\alpha = P(rejeter H_0 / H_0 vrai)$$

Ou **accepter** H_1 alors qu'elle est fausse :

$$\alpha = P(accepter H_1 / H_1 fausse)$$

*Si nous ne rejetons pas l'hypothèse nulle alors qu'elle est fausse nous commettons une erreur de deuxième espèce. C'est le cas si la valeur de la statistique de test tombe dans la région de non rejet (ou d'acceptation) alors que H_0 est fausse(c'est à dire si H_1 est vraie).
Le risque β de deuxième espèce est celui d'**accepter** H_0 alors qu'elle est fausse

$$\beta = P(accepter H_0 / H_0 fausse)$$

ou

$$\beta = P(accepter H_0 / H_1 vrai)$$

Ou **rejeter** H_1 alors qu'elle est vraie :

$$\beta = P(rejeter H_1 / H_1 vrai)$$

Nous pouvons résumer la situation avec le tableau suivant :

		*D*écision	
		H_0 est vrais	H_1 est vrais
*R*éalité	H_0 est vrais	Bonne décision	Erreur de deuxième espèce β
	H_1 est vrais	*E*rreur de première espèce α	Bonne décision

TAB. 1.1 – Résumé les cas de deux espaces

1.2.4 Niveau de signification du test

***Le niveau de signification** (ou niveau α) est un seuil qui détermine si le résultat d'une étude peut être considéré comme statistiquement significatif après que les tests statistiques prévus ont été réalisés. Il s'apelle aussi le risque de première espèce maximal :

$$\alpha = \sup \theta \qquad \text{tellque} \quad \theta \in \theta_1$$

***Le niveau de signification** est le plus souvent défini sur $\mathbf{5\%(ou\ 0,05)}$. Cependant, d'autres niveaux peuvent être utilisés en fonction de l'étude. Cela représente la probabilité de rejeter l'hypothèse nulle lorsqu'elle est vraie.

1.2.5 Variable de décision

*Si la valeur expérimentale de notre variable est dans la région critique, on l'hypothèse H_0 n'est plus tenable et on accepte H_1 avec un risque d'erreur α. Sinon, on conserve H_0 et on rejette H_1.

Le rejet ou non de H_1 dépend du niveau de signification du test (du choix de α) **et** de la taille de l'échantillon.

1.2.6 Région critique

*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 région critique est déterminée par la nature de H_1.

La région critique dépend aussi de la nature de test à réaliser. La nature de H_0 détermine la façonde formuler H_1 et par conséquence, et selon la nature unilatérale ou bilatérale du test,

la définition de la région critique varie.

Dans la plupart des situations que vous rencontrerez dans la suite, la région critique W peut être reliée au risque d'erreur de première espèce α par :

$$P_{H_0}(W) = \alpha$$

1.2.7 Région d'acceptation

*La région d'acceptation notée $\bar{W}$, ou encore appelée **zone d'acceptation** est la région complémentaire de la région critique W. Elle correspond à l'intervalle dans lequel les différences observées entre les réalisations et la théorie sont attribuables aux fluctuations d'échantillonnage.

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

Remarque 1.1 *Dans la plupart des situations que vous rencontrerez dans la suite, la région d'acceptation W peut être reliée au risque d'erreur de première espèce α par :*

$$P_{H_0}(\bar{W}) = 1 - \alpha$$

1.2.8 P-Valeur

*La p-valeur est la plus petite réel $\alpha \in]0, 1[$ calculé à partir des données tel que l'on puisse se permettre de rejeter H_0 au risque $(\alpha\%)$. Autrement écrit, la p-valeur est une estimation ponctuelle de la probabilité critique de se tromper en rejtant H_0 alors que H_0 est vraie.

1.2.9 Démarche d'un test statistique

Les étapes à suivre pour tester une hypothèse sont les suivantes :

1. Définir l'hypothèse nulle (notée H_0) à contrôler.

2. Choisir un test statistique ou une statistique pour contrôler H_0.
3. Définir la distribution de la statistique sous l'hypothèse « H_0 est réalisée ».
4. Définir le niveau de signification du test ou région critique notée α.
5. Calculer à partir des données fournies par l'échantillon, la valeur de la statistique.
6. Prendre une décision concernant l'hypothèse posée et conclure.

Chapitre 2

Techniques empiriques et méthodes graphiques

*L'appréhension d'un jeu de données passe systématiquement par les statistiques descriptives. Elles donnent une image globale. Bien souvent, elles permettent de se faire une idée sur les techniques que l'on pourrait utiliser et les dangers ou artefacts dont il faudra se méfier.
Bien avant les techniques complexes et les ratios savants, quelques indicateurs usuels et des graphiques judicieusement choisis sont le bienvenu. Ces outils sont disponibles dans tous les outils de traitement exploratoire des données.

2.1 Histogramme

*L'histogramme est un outil graphique à barres verticales accolée, obtenu aprés découpage en classes des observations d'une variable continue. IL est analogues à la densité d'une variable aléatoire continue. En l'utilise pour comparer la distribution des données analysées en les représentant sous forme d'histogramme avec une courbe représentant la loi normale.

2.1.1 programmation en code R

hist(x) %pour présenter l'histogramme des fréquence de x.

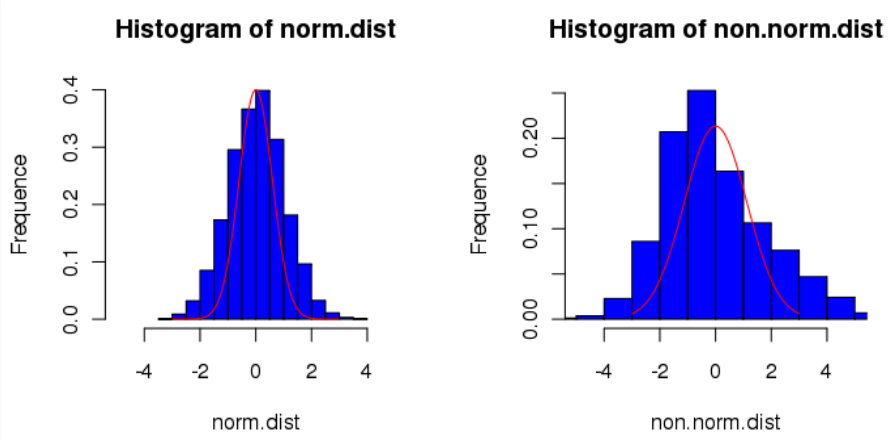

FIG. 2.1 – Histogramme de distribution normale et distribution non normale

Exemple 2.1 *Dans cet exemple, nous présentons l'histogramme pour faire la comparaison entre la loi normale et la loi uniforme avec* $n = 30$.

programme :

op=par(mfrow=c(1,2))

poids=rnorm(30,10,2) %création d'un échantillon de 30 poids provenant d'une distribution normale ayant pour moyenne 10 et pour écart type 2.

poids

[1] 12.245167 10.490593 8.205654 13.059189 10.459846 9.813346 9.859615

[8] 8.124903 11.648303 9.335432 11.307787 12.720500 9.494156 11.172032

[15] 9.099665 9.020339 11.290081 12.228375 7.175643 8.461143 10.449054

[22] 8.151913 9.180147 9.355788 9.925532 11.516052 10.454256 8.027909

[29] 11.290209 9.168457

hist(poids,prob=T)

curve(dnorm(x,mean(poids),sd(poids)),add=T,col="red")

poids1=runif((10,8,13),runif(20,13,15)) %création d'un échantillon de 30 poids.

poids1

[1] 12.942454 11.073317 8.169298 12.145237 8.526026 10.658278 10.807352

[8] 11.845429 9.919003 11.465058 13.897973 14.929379 14.550819 13.725673

[15] 14.008494 14.204234 13.532013 14.105695 13.296534 13.048093 14.891749

[22] 13.736529 14.428254 13.749061 13.382411 13.362361 14.741129 13.550638

[29] 13.036630 13.494419

hist(poids1,prob=T)

curve(dnorm(x,mean(poids1),sd(poids1)),add=T,col="red")

Résultat de la commande :

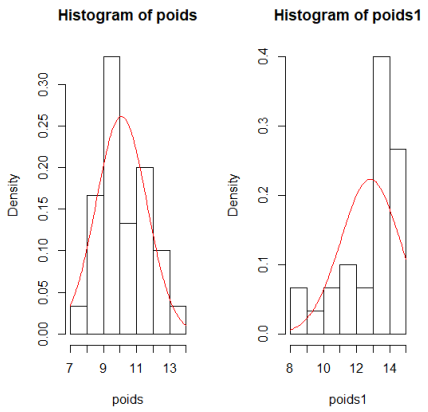

FIG. 2.2 – Histogramme de la la loi normale et loi uniforme avec n=30

commentaire : On observe que dans le premier cas, les données sont centrées et semblent s'ajuster à la courbe de la loi normale alors que dans le deuxième cas, les données sont plus dispersées et s'éloigne plus fortement de la loi normale.

Exemple 2.2 *La comparaison entre la loi normale centrée et la loi expontielle de paramètre* 4 *avec* $n = 400$.

programme :

op=par(mfrow=c(1,2))

poids2=rnorm(400,0,1)

poids2

hist(poids2,prob=T)

curve(dnorm(x,mean(poids2),sd(poids2)),add=T,col="red")

poids3=rexp(400,4)

poids3

hist(poids3,prob=T)

curve(dnorm(x,mean(poids2),sd(poids2)),add=T,col="red")

Résultat de la commande :

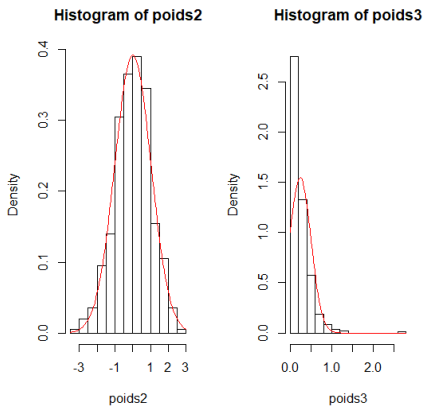

FIG. 2.3 – *H*istograme de loi normale centrée et loi expontielle avec n=400

Commentaire : On a les mêmes remarques que l'exemple précédent.

2.2 Boite à moustache

*La boite à moustache (en anglais $box-plot$) est un outil graphique trés pratique représentant une distribution empirique à l'aide de quelques paramètres de localisation : la médiane (M), le $1^{ère}$ quartile($Q1$) et $3^{ème}$ quartile($Q3$), minimum et maximum, telle que :

1. **La médiane** : Cest la valeur "centrale" de la série. On dit quelle partage la série en deux moitiés.
2. **Les quartiles** : partagent la série en 4, en a donc :

– Le $1^{ère}$ quartile($Q1$) : est la plus petite valeur, telle que 25% des données lui soit inférieures ou égales.

– Le $2^{ème}$ quartile($Q2$) : est la médiane.

– Le $3^{ème}$ quartile($Q3$) : est la plus petite valeur, telle que 75% des données lui soit inférieures ou égales.

*La boîte à moustache permet d'observer les valeurs extrêmes (outliers) mais également d'avoir une idée sur la symétrie de la distribution. La symétrie d'une distribution n'affirme pas la normalité, mais une distribution normale est forcément symétrique. Une boîte à moustache est dite symétrique lorsque la position de la médiane se situe au milieu de la boîte à moustache et qu'il y a symétrie des moustaches.

Remarque 2.1 *On a aussi :*

- **L'intervalle** $[Q1; Q3]$ s'appelle L'intervalle interquartile.
- **Le nombre** $(Q3 - Q1)$ s'appelle l'écart interquartile.
- **La moustache inférieure** : valeur de la série immédiatement supérieure la frontière basse, avec la frontière basse qui égale $(Q1 - 1.5 * (Q3 - Q1))$.
- **La moustache supérieure** : valeur de la série immédiatement inférieure la frontière haute, avec la frontière haute égale $(Q3 + 1.5 * (Q3 - Q1))$.

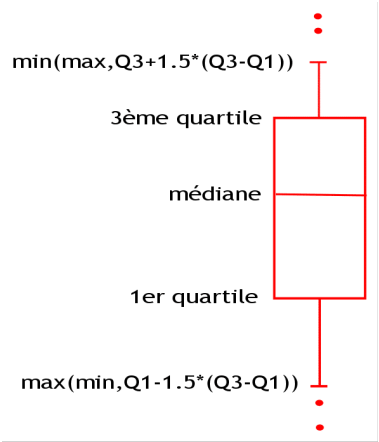

FIG. 2.4 – Diagramme de boite à moustache

2.2.1 programmation en code R

boxplot(x) %pour présenter la Graphe de boite à moustache de x.

Exemple 2.3 *Dans cet exemple, nous présentons le graphe de boite à moustache pour faire la comparaison entre la loi normale de moyenne 10 et l'écart type 2 et la loi uniforme avec* $n = 30$.

programme :

```
op=par(mfrow=c(1,2))
poids=rnorm(30,10,2)                %création d'un échantillon de 30 poids provenant
d'une distribution normale ayant pour moyenne 10 et pour écart type 2.
poids
boxplot(poids,main="poids")
curve(dnorm(x,mean(poids),sd(poids)),add=T,col="red")
poids1<-runif((10,8,13),runif(20,13,15))        %création d'un échantillon de 30
poids.
poids1
boxplot(poids1,main="poids1")
curve(dnorm(x,mean(poids1),sd(poids1)),add=T,col="red")
```

Résultat de la commande :

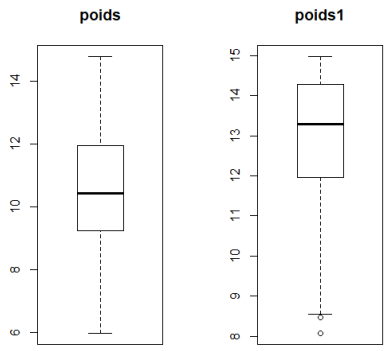

FIG. 2.5 – Graphe de boite à moustache de loi normale et loi uniforme avec n=30

commentaire : On observe, dans l'exemple 1 (poids), que la médiane se situe légèrement dans la partie supérieure de la boîte à moustache et que le minimum et le maximum sont légèrement asymétriques. Il y a également une mesure qui est fortement séparée des autres et qui pourrait être un outlier. Dans l'exemple 2 (poids1), la médiane se situe à l'extrémité haute de la boîte à moustache et les moustaches sont fortement asymétriques.

Exemple 2.4 *La comparaison entre la loi normale de paramètres* $\mu = 0$ *et* $\sigma = 1$ *et loi*

expontielle de paramètre 4 *avec* $n = 400$.

programmation :

```
op=par(mfrow=c(1,2))
poids2=rnorm(400,0,1)
boxplot(poids2,main="poids2")
poids3=rexp(400,4)
boxplot(poids3,main="poids3")
```

Résultat de la commande :

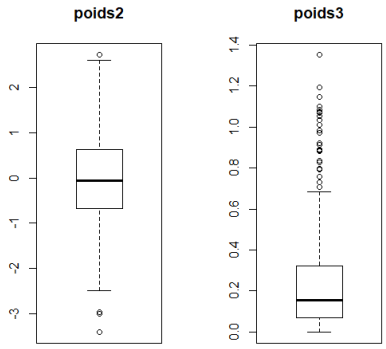

FIG. 2.6 – Graphe de boite à moustache de loi normale centrée et loi expontille avec n=400

commentaire : D'aprés le figure on observe que la première boite à moustache (poids2) est symétrique qui présente la loi normale et la deuxième boite (poids3) n'est pas symétrique pour la distribution expontielle.

2.3 $QQ - Plot$ (**quantille-quantille plot**)

*Le "**diagramme Quantile-Quantile**" ou "**diagramme Q-Q**" ou "**Q-Q plot**" est un outil graphique permettant d'évaluer la pertinence de l'ajustement d'une distribution donnée à un modèle théorique. à partir de la série statistique observée, on calcule alors un certain nombre de quantiles. Si la série statistique suit bien la distribution théorique choisie, on devrait avoir les quantiles observés égaux aux quantiles associés au modèle théorique.

*Le $QQ-plot$ gaussien est un outil graphique qui permet d'apprécier visuellement la normalité d'une variable quantitative. La fonction à utiliser est la fonction $qnorm()$.

Où : $qqnorm$ permet de comparer graphiquement la distribution d'un échantillon avec une distribution normale (si points alignés alors la distribution est normale).

*Les échantillons ne sont pas forcément de même taille. Il se peut également, et c'est ce qui nous intéresse dans le cas présent, qu'un des ensembles de données soient générées à partir d'une loi de probabilité qui sert de référentiel.

Concrètement, il sagit :

1. De trier les données de manière croissante pour former la série $x_{(i)}$.
2. Chaque valeur $x_{(i)}$, nous associons la fonction de répartition empirique $F_i = \frac{i-0,375}{n+0,25}$.
3. Nous calculons les quantiles successifs $Z_{(i)}$ d'ordre F_i en utilisant l'inverse de la loi normale centrée et réduite.
4. En n, les données initiales n'étant pas centrées et réduites, nous dé-normalisons les données en appliquant la transformation.

$$x_{(i)} = Z_{(i)} * s * \bar{x}$$

Telle que :

$$\bar{x} = \frac{1}{n} \sum_i x_i$$

Où : $\bar{x}$ est la moyenne empirique, calculé partir les observation de l'échantillon $(x_1, ..., x_n)$.

$$Z_{(i)} = \frac{x_i - \bar{x}}{s}$$

Et : S est l'estimateur de l'écart type $\sigma(X)$; défini par :

$$S = \sqrt{\frac{1}{n-1} \sum_i (x_i - \bar{x})^2}$$

2.3.1 programmation en code R

qqplot(x) %pour présenter graphe des quantiles.

Exemple 2.5 *On peut utilisé même exemple précédent (exemple de boite à moustache)*

programme

op=par(mfrow=c(1,2))

poids=rnorm(30,10,2) %création d'un échantillon de 30 poids provenant d'une distribution normale ayant pour moyenne 10 et pour écart type 2.

qqnorm(poids,datax=TRUE,main="poids")

qqline(poids,datax=TRUE) %pour dessiné droit de henry.

poids1<-runif((10,8,13),runif(20,13,15)) %création d'un échantillon de 30 poids.

qqnorm(poids1,datax=TRUE,main="poids1")

qqline(poids1,datax=TRUE) %pour dessiné droit de henry.

Résultat de la commande :

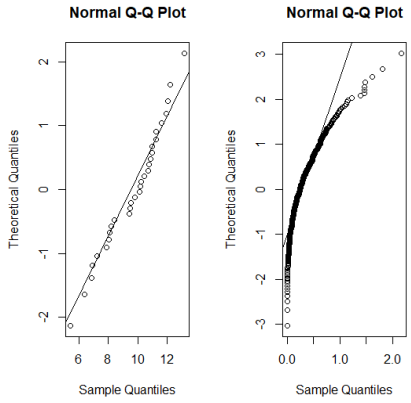

FIG. 2.7 – Graphe de quantile de loi normale et quantile de loi uniforme avec n=30.

commentaire : Plus les données (points) se rapprochent de la droite, plus la distribution empirique est dite normale. Les données de l'exemple 1 (poids) sont proches de la droite tandis que les données de l'exemple 2 (poids1) sont plus éloignées.

Programme :

x=rnorm(500,0,1)

qqnorm(x) %pour dessiner les quantile de loi normale N(0 ;1) avec les quantiles de x.

qqline(x,col=2) % ajouter une ligne au graphe qqnorm.

Résultat de la commande :

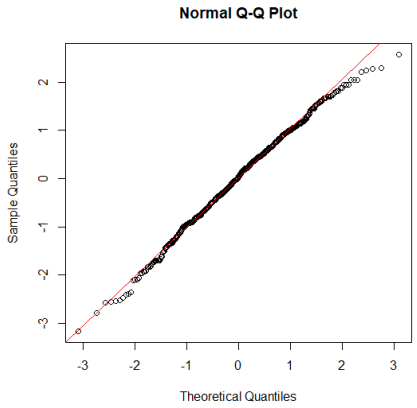

FIG. 2.8 – Graphe des quantiles de loi normale centrée et les quantiles de x

Commentaire :

On observe que les points de quantiles alignées. Donc x suit la loi normale.

Chapitre 3

Tests statistiques

Très commodes, les approches empiriques n'ont pas la rigueur des techniques statistiques. Dans ce chapitre, nous présentons les tests de compatibilité à la loi normale. Encore une fois, il s'agit bien de vérifier l'adéquation (la compatibilité) à la loi normale et non pas déterminer la loi de distribution.

3.1 Tests de normalité

*Le principe de ces tests (tests de normalité) est de comparer la fonction de répartition théorique $F(x)$ spécifiée sous H_0, et la fonction de répartition empirique $F_n(x)$ vue précédemment et définie par :

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{(X_i \leq t)} = \begin{cases} 0 & X_{(1)} > t \\ \frac{i}{n} & X_{(i)} \leq t \leq X_{(i+1)} \\ 1 & X_{(n)} \leq t \end{cases} \tag{3.1}$$

Telle que :

$$\mathbb{I}_{(X_i \leq t)} = \begin{cases} 1 \quad si & X_i \leq t \\ 0 & \text{sin}\, ous \end{cases} \tag{3.2}$$

X_i :représentent des variables de la loi normale.

Et l'hypothèse s'écrit sous forme de :

$$\begin{cases} H_0 : F_n = F \\ H_1 : F_n \neq F \end{cases}$$

Tell que la distribution d'une variable aléatoire X est décrite par sa fonction de répartition c'est a dire par la fonction.

$$F(x) = P(X \leq x)$$

Exemple 3.1 *On appelle loi normale de paramètres $\mu \in \mathbb{R}$ et $\sigma \geq 0$ la loi d'une variable aléatoire continue X prenant toutes les valeurs réelles de densité de probabilité, la fonction définie pour $x \in \mathbb{R}$ par :*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\right] \tag{3.3}$$

Et sa fonction de répartition est donnée par :

$$F_X(x) = \int_{-\infty}^{x} f(t)dt = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\right] dt \tag{3.4}$$

3.1.1 Test de kolmogorov-Smirnov

*Le principe du test de **Kolmogorov-Smirnov** est de comparer deux distributions au moyen de la fonction de répartition. Dans ce cas, on va comparer la fonction de répartition empirique $F_n(x)$ avec la fonction de répartition théorique $F(x)$ d'une loi normale dont les paramètres sont estimés à partir de l'échantillon.

1. La statistique D_n de **Kolmogorov-Smirnov** est définie par :

$$Dn = sup|F_n(x) - F(x)| \tag{3.5}$$

$F(x)$: la fonction de répartition de la loi normale.

X_i : représentent des variables de la loi normale et continue.

2. La region critique : Si $\begin{cases} Dn > d_{n,\alpha} & \text{On rejette } H_0 \\ Dn < d_{n,\alpha} & \text{On rejette } H_1 \end{cases}$

Où : $d_{n,\alpha}$ est le quantile théorique lu à partir la table de **Kolmogorov-Smirnov** et $(\alpha = 5\%)$.

Remarque 3.1 (***Tests de Kolmogorov-Smirnov de comparaison de deux échantillons***)

On considère deux échantillons indépendants : $X_1, ..., X_n$ $(i.i.d)$, de fonction de répartition F_0 et $Y_1, ..., Ym$ $(i.i.d)$. de fonction de répartition F_1. On veut tester $\{H_0 : F_0 = F_1 \; contre \; H_1 : F_0 \neq F_1\}$. Soit F_n la fonction de répartition empirique de l'échantillon $(X_1, ..., X_n)$ et G_m celle de l'échantillon $(Y_1, ..., Ym)$.

Alors, Le test **d'homogénieté de Kolmogorov-Smirnov** est défini par la statistique :

$$D_{n,m} = \sqrt{\frac{nm}{n+m}} sup|F_n(X) - G_m(X)| \tag{3.6}$$

Il consiste à rejeter l'hypothèse H_0 si $D_{n,m} \geq d_{n,m,1-\alpha}$.

Programmation en code R

On compare les p-value de tout les tests par 0.01, c-à-d :

$\begin{cases} si \quad \text{p-value} > 0.01 \quad \text{on accepte } H_0 \\ si \quad \text{p-value} \leq 0.01 \quad \text{on rejette } H_0 \end{cases}$

Exemple 3.2 *Nous utilisons le même exemple pour appliquer tous les tests.*

x=rexp(10,5)

ks.test(x,x)

Résultat de la commande :

Two-sample Kolmogorov-Smirnov test

data : x and x

Ks = 0, p-value = 1

alternative hypothesis : two-sided

Message d'avis :

In ks.test(x, x) : les valeurs "p" seront approximées en présence d'ex-aequos.

3.1.2 Test de Lilliefors

*Le test de Lilliefors est une variante du test de Kolmogorov-Smirnov où les paramètres de la loi (μ et σ) sont estimées à partir des données. La statistique du test est calculée de la même manière. Mais sa loi est tabulée différemment, les valeurs critiques sont modifiées pour un même risque α. Elles ont été obtenues par simulation.

1. La statistique D de Lilliefors est définie par :

$$D = \sqrt{n}D_n = \max_{i=1,\dots,n}(F_i - \frac{i}{n}, F_i - \frac{i-1}{n}) \tag{3.7}$$

où F_i est la fréquence théorique de la fonction de répartition de la loi normale centrée réduite associée à la valeur standardisée $Z(i) = \frac{X_{(i)}-\bar{X}}{\delta_x}$ tel que : $\begin{cases} \bar{X} : moyenne \ \ empirique \\ \delta_x = écart - type \ empirique \end{cases}$

2. La région critique : La table des valeurs critiques D_{crit} pour les petites valeurs de n et diférentes valeurs de α doivent être utilisées. Lorsque les effectifs sont élevés, typiquement $n \geq 30$; il est possible d'approcher la valeur critique à l'aide de formules simples suivent :

α	0.1	0.05	0.01
D_{crit}	$\frac{0.805}{\sqrt{n}}$	$\frac{0.886}{\sqrt{n}}$	$\frac{1.031}{\sqrt{n}}$

TAB. 3.1 – Tableau des valeurs critiques de lillifors

Et la région critique définie par :

$$D > D_{crit}$$

programmation en code R

x=rexp(50,5)

lillie.test(x,x)

Résultat de la commande :

Two-sample Kolmogorov-Smirnov test

data : x and x

D = 0.292, p-value =0.02

Commentaire :

L'exemple ci-dessus renvoie une p-value non significative (p-value=0.02>0.01). Alors l'échantillon accepte la normalité pour la loi expotielle.

3.1.3 Test de Shapiro-Wilk

*Le test de Shapiro et Wilk (1965) a été initialement restreint pour des échantillons de moins de 50 ($n \leq 50$). Ce test était le premier qui pouvait détecter des écarts par rapport à la normalité en raison de l'asymétrie ou de la kurtose, ou les deux (Althouse et al., 1998). Il est devenu le test préféré en raison de ses bonnes propriétés énergétiques (Mendes et Pala, 2003). Compte tenu d'un échantillon aléatoire ordonné.

Le test de **Shapiro-Wilk** est basé sur la statistique W. Il est trés popoulaire en comparaison des autres tests.

1. La statistique W de **Shapiro-Wilk** est définie par :

$$W = \frac{\left[\sum_{i=1}^{\left[\frac{n}{2}\right]} a_i (x_{(n-i+1)} - x_{(i)})\right]^2}{\sum_i (x_i - \bar{x})} \tag{3.8}$$

Où

x_i : correspond à la série des données triées.

$[\frac{n}{2}]$: est la partie entière du rapport $\frac{n}{2}$.

a_i : sont des constantes générées à partir de la moyenne et de la matrice de variance (covariance) des quantiles d'un échantillon de taille "n" suivant la loi normale. Ces constantes sont fournies dans des tables spécifques.

2. La region critique : Si $\begin{cases} W \geq W_{crit} & \text{On rejette } H_0 \\ W < W_{crit} & \text{On rejette } H_1 \end{cases}$

les valeurs seuils W_{crit} pour différents risques α et effectifs n sont lues dans table de **Shapiro-Wilk.** La table de Shapiro-wilk donne W_{crit} pour des tailles inférieur ou égale à 50 seulement.

programmation en code R

x=rexp(10,5)

shapiro.test(x)

Résultat de la commande :

Shapiro-Wilk normality test

data : x

W = 0.9851, p-value = 0.3213

Commentaire :

L'exemple ci-dessus renvoie une p-value non significative (p-value=0.3213>0.01). Alors l'échantillon accepte la normalité pour la loi expontielle.

Exemple 3.3 *On trouve un autre exemple :*

shapiro.test(runif(100, min = 2, max = 4))

Résultat de la commande :

Shapiro-Wilk normality test

data : runif(100, min = 2, max = 4)

W = 0.9451, p-value = 0.0004003

Commentaire :

Dans lexemple ci-dessus, le p-value est significatif (p-value=0.0004003<0.01). Alors l'échantillon ne suit donc pas une loi normale, c-a-d rejette la normalité pour la loi uniforme.

3.1.4 Test de Cramer-Von Mises

*Le test de Cramer-Von Mises est une variante au test de Kolmogoro-Smirnov. Il permet égalment de tester toute forme de différenciation entre les distributions. Sa particularité est qu'il exploite différemment les fonctions de répertition empirique au lieu de se focaliser sur l'écart maximal, il compile tous les écarts sous la forme d'une somme des carrés des différences.

1. La statistique CM : elle s'écrit

$$\mathbf{CM}= \mathbf{n} \int_{-\infty}^{+\infty} \left[F_n(x) - F_0(x)\right]^2 dF_n(x) \tag{3.9}$$
$$= \frac{1}{12n} + \sum_{i=1}^{n} \left[\frac{2i-1}{2n} - F(x_i)\right]^2$$

2. La region critique : cette statistique admet une loi limite permettant de déterminer la valeur de **"c"** définissant la région critique, et on rejette H_0 si

$$CM \geq c$$

pour un niveau α donnée. La valeur de "c" calculée à partir la table de Cramer-Von Mises.

programmation en code R

x=rexp(10, 5)

cvm.test(x)

Résultat de la commande :

Cramer-von Mises normality test

data : x

CM=0.07, p-value =0.76

Commentaire :

L'exemple ci-dessus renvoie une p-value non significative (p-value=0.76>0.01). Alors l'échantillon accepte la normalité pour la loi expontielle (on accepte l'hypothèse H_0).

3.1.5 Test d'Anderson-Darling

*Le test **d'Anderson-Darling (AD)** est une modification du test de **Cramer-von Mises**. Il diffère du test CVM de telle sorte qu'il accorde plus de poids aux queues de la distribution (Farrel et Stewart, 2006). Selon Arshad et al. (2003), ce test est le plus puissant des tests EDF. La statistique du test AD appartient à la classe quadratique de la statistique EDF dans la quelle elle est basée sur la différence au carré.

1. La statistique AD d'Andarson-Darling est définie par :

$$AD = -n - \frac{1}{n}(\sum_{i=1}^{n}(2i-1)[\ln(F_i) + \ln(1 - F_{n-i+1})]) \tag{3.10}$$

où F_i : est la fréquence théorique de la loi de répartition normale centrée réduite associée à la valeur standardisée,

$$z_i = \frac{x_{(i)} - \bar{x}}{s}$$

2. La région critique : les valeurs critiques A_{crit} pour différents niveaux de risques sont résumées dans le tableau suivant, ils ont été produits par simulation et ne dépendent pas de l'effectif de l'échantillon.

α	A_{crit}
0.1	0.631
0.05	0.752
0.01	1.035

TAB. 3.2 – les valeurs critiques de test d'Andarson darling

L'hypothèse de normalité est rejetée lorsque la statistique AD prend des valeurs trop élevées :

$$AD \geq A_{crit}$$

les étapes de calculs de ce test : sont présentés dans les points suivants :

1. (a) On ordonne les observations pour obtenir les $x_{(i)}$.
   (b) On calcule $\bar{x}$ et s.
   (c) On calcule les données centrées réduites z_i.
   (d) On utilise la fonction de répertition de la loi $N(0,1)$ pour calculer F_i et aprés $ln(F_i)$.
   (e) De la même manière,nous formons (1-F_{n-1+i}) puis $ln(1 - F_{n-i+1})$.
   (f) On calcule la somme $S = \sum_{i=1}^{n}(2i-1)[\ln(F_i) + \ln(1 - F_{n-i+1})]$.
   (g) On calcule la statistique $AD = (-n - \frac{1}{n})S$.

(h) Pour faire la décision, on compare la valeur **AD** avec $\mathbf{A}_{crit}$.

3. Calcul de P-value : la p-value est calculé à partir de la statistique A_m définie par :

$$A_m = AD(1 + \frac{0.75}{n} + \frac{2.25}{n})$$

programmation en code R

x=rexp(10, 5)

ad.test(x)

Résultat de la commande :

Anderson-Darling normality test

data : x

AD=0.377, p-value = 0.5932

Commentaire :

L'exemple ci-dessus renvoie une p-value non significative (p-value=0.5932>0.01). Alors l'échantillon suit donc une loi normale (on accepte l'hypothèse H_0).

3.1.6 Test de Jarque-Bera

*Avant de représenter ce test, on va définir les coefficients d'asymétrie et d'aplatissement, car le test de Jarque-Bera et basé sur les deux derniers.

1. Coefficients d'asymétrie : le coefficiant d'asymétrie correspond à une mesure de l'asymétrie de la distribution d'une variable aléatoire réelle c'est le premier des paramètre de forme donné. On a 3 cas de forme de distribution :

1. (a) Un coefficiant positive indique une distribution décalée à gauche de la moyenne et donc une queue de distribution étalée vers la droite.

   (b) Un coefficiant négatif indique une distribution décalée à droit de la moyenne et donc une queue de distribution étalée vers la gauche.

   (c) Un coefficiant nul indique une distribution symétrique :comme la loi normale ce coefficiant est définie par :

$$\alpha_3 = \frac{E\left[(X-\mu)^3\right]}{\sigma^3} = \frac{\mu^3}{\sigma^3} \tag{3.11}$$

Où : X variable aléatoire d'éspérence μ et variance σ^2.

L'éstimateur de ce coefficiant est définie par :

$$G_1 = \frac{\hat{\mu}^3}{\hat{\sigma}^3} = \frac{(\frac{1}{n}\sum_i (x_i - \bar{x}))^3}{(\frac{1}{n}\sum_i (x_i - \bar{x})^2)^{\frac{3}{2}}} = \frac{n}{(n-1)(n-2)} \sum_i \left(\frac{x_i - \bar{x}}{s}\right)^3 \tag{3.12}$$

2. Coefficient d'aplatissement : le coefficient d'aplatissement c'est le $2^{éme}$ des paramètre de forme. Les valeurs peuvent être distribution au voisinage de la moyenne et dans ce cas le pic réprésentatif est aigu. Dans d'autre distribution peuvent être plates, donc le coefficiant d'aplatissement donne une évaluation de l'importance du pic, il a pour expression.

$$\alpha_4 = \frac{E\left[(X-\mu^4)\right]}{\sigma^4} = \frac{\mu^4}{\sigma^4} \tag{3.13}$$

Où : X variable aléatoire normale, alors α_4 égale à 3.

L'éstimateur de ce coefficiant est définie par :

$$G_2 = \frac{\hat{\mu}^4}{\hat{\sigma}^4} = \frac{(\frac{1}{n}\sum_i (x_i - \bar{x}))^4}{(\frac{1}{n}\sum_i (x_i - \bar{x})^2)^2} = \frac{(n-1)n}{(n-1)(n-2)(n-3)} \sum_i \left(\frac{x_i - \bar{x}}{s}\right)^4 - \frac{3(n-1)}{(n-2)(n-3)} \tag{3.14}$$

La loi conjointe de l'estimateurs (G_1, G_2) est la loi normale bivariée, on écrit :

$$\sqrt{n}\begin{pmatrix} \mathrm{G}_1 \\ \mathrm{G}_2 \end{pmatrix} \overset{loi}{\rightsquigarrow} N\left[\begin{pmatrix} 0 \\ 3 \end{pmatrix}, \begin{pmatrix} 6 & 0 \\ 0 & 24 \end{pmatrix}\right]$$

La matrice de variance covariance présentée ici est une expression simplifiée valable pour les grandes valeurs de n. Il est possible de produire des expressions plus précises. Nous notons également que la covariance de $G_1 et\ G_2$ est nulle.

3. Principe de test de Jarque-Bera :

Le test de normalité de Jarque-Bera est également fondé sur les coefficients d'asymétrie et

d'aplatissement. Il éalue les écarts simultanés de ces coefficients avec les valeurs de référence de la loi normale.

***L'hypothèse de ce test** : elle basé sur les deux coefficient d'asymétrie et d'aplatissement

$$\begin{cases} H_0 : G_1 = 0 \text{ et } G_2 = 3 \\ H_1 : G_1 \neq 0 \text{ et } G_2 \neq 3 \end{cases}$$

***La statistique** T : la statistique de Jarque-Bera s'écrit :

$$T = \frac{n}{6}\left(G_1^2 + \frac{(G_2 - 3)^2}{4}\right) \tag{3.15}$$

***La région critique :** nous observons que la statistique T est une somme de deux variables aléatoires indépendantes de loi du $\mathcal{X}^2$ à 2 degrés de liberté. Alors la région critique pour un risque α du test est définie par :

$$P(T \geq \mathcal{X}^2_{1-\alpha}(2)) = \alpha$$

où $\mathcal{X}^2_{1-\alpha}(2)$ est une valeur théorique lu à partir la table de $\mathcal{X}^2$.

***La p-value :** Malgré la statistique T suit la loi de $\mathcal{X}^2$, on peut rejetter H_0 ou l'accepter à partir une comparaison entre la p-value et le niveau de risque α. Cette dernière calculé à partir des simulation utilisant des logiciels $(R, ...)$.

α	$\mathcal{X}^2_{1-\alpha}(2)$
0.05	5.99
0.01	9.21

TAB. 3.3 – exemples des valeurs critiques de test de Jarque Bera

programmation en code R

X=rexp(10,5)

jb.test(x)

Résultat de la commande :

Jarque-Bera normality test

data : x

T=9.094, p-value = 0.01

Commentaire :

L'exemple ci-dessus renvoie une p-value non significative (p-value=0.01$\leq$0.01). Alors on rejette l'hypothèse H_0 c-à-d l'échantillon rejette la normalité pour la loi expontielle.

Conclusion

Le but de notre mémoire est de présenter les différentes méthodes de tests de normalités le plus important qui nous permettent de vérifier si les donnés réelles suivent la loi normale ou non. Il existe deux méthodes pour vérifier la normalité. la première méthode est la méthode graphique, où elle nous donne une ideé sur la distribution de l'échantillon réel, telle que : la symétrie. Mais la forme symétrique de l'histogramme ne conduit pas à la normalité des données, car la loi normale n'est pas la seule à avoir une courbe symétrique, c'est également le cas des lois de Cauchy et de Student. Alors la méthode graphique n'est pas suffisante pour conclure la normalité, pour cela, on applique les tests d'hypothèses. On résume les étapes des tests de normalité dans les points suivants :

1. On détermine un test parmi les tests de normalité, telle que : anderson darling, shapiro,...etc.
2. On fixe une valeur du seuil critique α.
3. On calcul la statistique correspendant au test choisi.
4. On compare la $p-value$ calculé à partir la simulation avec le seuil α, si $p-value>\alpha$, on accepte l'hypothèse de normalité H_0, si non on le rejette.

Bibliographie

[1] Carpentie, F.(2012). statistique paramétrique et non paramétrique. Université de Brest.

[2] Chine Amel, Test de normalié (mémoire). Université Mohamed Kheider de biskra.

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

[4] D,Rhesse et A.B, Dufour.(2008). Décision et risque dérreur. S.penel. URL

[5] Gilbert Colletaz, statistique non paramétrique et Économétrie et statistique appliquée cours, 2009.

[6] Maumy-Bertrand, (M.Bertrand, (2010)). Initiation à la statistique avec R : Cours, exemples, problèmes corrigés. Dunod.

[7] Nacira Hadjaj Seddik, ≪Les test de normalité de lhoste≫ Mathématique et sciences humaines 2003.

[8] N AKAKPO-Note de cours issuses du module statistique...2017-Ipsm.paris.

[9] Odile Wolber, Les graphique dans R.

[10] Olivier Gaudoin. Principes et les Méthodes statistique, Notes cours. Université de Laguna.

[11] Olivier Torrés. (2010). Calculs de probabilités avec loi normal. Université.R.L.

[12] Ricco Rakotomalala. Comparaison de population et Tests non paramétriques, Université Lumière Lyon 2, 2008.

[13] Ricco Rakotomalala. Techniques empirique et tests statistiques. Manuel de cours, université Lumière Lyon, 2008 - eric.univ-lyon2.fr.

[14] V.Monbet, Tests statistique notes de cours, L2 S1-2009.

Annexe A : Logiciel *R*

R est un système, communément appelé langage et logiciel, qui permet de réaliser des analyses statistiques. Plus particulièrement, il comporte des moyens qui rendent possible la manipulation des données, les calculs et les représentations graphiques. *R* a aussi la possibilité d'exécuter des programmes stockés dans des fichiers textes et comporte un grand nombre de procédures statistiques appelées paquets. Ces derniers permettent de traiter assez rapidement des sujets aussi variés que les modèles linéaires (simples et généralisés), la régression (linéaire et non linéaire), les séries chronologiques, les tests paramétriques et non paramétriques classiques, les différentes méthodes d'analyse des données,... Plusieurs paquets, tels *ade4, FactoMineR, MASS, multivariate, scatterplot3d et rgl* entre autres sont destinés à l'analyse des données statistiques multidimensionnelles.

*Il a été initialement créé, en 1996, par *Robert Gentleman* et *Ross Ihaka* du département de statistique de l'Université d'Auckland en Nouvelle Zélande. Depuis 1997, il s'est formé une équipe "*R Core Team*" qui développe *R*. Il est conçu pour pouvoir être utilisé avec les systèmes d'exploitation *Unix*, *Linux*, *Windows* et *MacOS*.

*Un élément clé dans la mission de développement de *R* est le *Comprehensive R Archive Network* (CRAN) qui est un ensemble de sites qui fournit tout ce qui est nécessaire à la distribution de R, ses extensions, sa documentation, ses fichiers sources et ses fichiers binaires.

*C'est aussi un outil trés puissant et trés complet, particuliérement bien adapté pour la mise en œuvre informatique de méthodes statistiques. Il est plus difficile d'accés que certains autres logiciels du marché (comme SPSS ou Minitab par exemple), car il n'est pas conçu pour être utilisé à l'aide de «clics» de souris dans des menus. L'avantage en est toutefois double :

1. l'approche est pédagogique puisqu'il faut maîtriser les méthodes statistiques pour parvenir à les mettre en œuvre.
2. l'outil est trés efficace lorsque l'on domine le langage R puisque l'on devient alors capable de créer ses propres outils, ce qui permet ainsi d'opérer des analyses trés sophistiquées sur les données.

Fonction nous permettent de définir les méthodes graphiques :

plot(X) : graphe des valeurs de X.

qqnorm(X): quantiles de X en fonction des valeurs attendues selon une loi normale.

hist : Histogramme des fréquences de X.

boxplot(X) : graphe boites à moustaches.

Fonction nous permettent de faire les tests de normalité :

ks.test : Test de kolmogorov-smirnov.

lillie.test : Test de lillieforse.

shapiro.test : Test de shapiro wilk.

cvm.test : Test de cramer von-mises.

ad.test : Test d'andarson-Darlling.

jb.test : Test de jarque-bera.

Annexe B : Abréviations et Notations

v.a : variable aléatoire.

R(X) : rang de l'observation X.

p − v : P-value.

iid : identiquement independent distribuée.

E(X) : esperance mathématique.

$\bar{\mathbf{x}}$: moyenne empirique.

$\sigma(\mathbf{x})$: écart type.

F(**X**) : fonction de répertition théorique.

$\mathbf{F}_n(X)$: fonction de répertition empirique.

$\mathbf{D}_n$: statistique de kolmogorov-smirnov.

K-S : kolmogorov-sminov.

$\mathbf{d}_{n,\alpha}$: quantille théorique de kolmogorov-smirnov.

Lill : lillifors.

D : statistique de lillifors.

CM : Statistique de cramer von mise.

AD : statistique d'Anderson Darling.

α_3 : Coefficient d'asymetrie.

α_4 : Coefficient d'aplatissement.

$\mathbf{G}_1$: estimateur de coefficient d'asymetrie.

$\mathbf{G}_2$: estimateur de coefficient d'aplatissement.

T : Statistique de Jarque-Bera.

var(X) : variance mathématique.

W : statistique de shapiro-wilk.

$\mathcal{X}^2$: loi de khi deux.

$\rightsquigarrow$: suit la loi.

EDP : Fonction de Distribution Empirique.

Annexe C : Tables statistiques

Table de la loi normale :

Soit $Z \sim N(0,1)$. La table ci-dessous donne les valeurs de $F(x) = P(Z \leq x)$, avec $x \in [0; 0,99]$ et $x = x_1 + x_2$.

x_1 \ x_2	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

FIG. 3.1 – Table de la loi normale

Table de loi de chi deux :

Soit $K \sim \mathcal{X}^2(v)$. La table ci-dessous donne, pour un α et un v choisis, la valeur $k_\alpha(v)$ telle que $P(K \geq k_\alpha(v)) = \alpha$.

Table des valeurs de Shapiro-Wilk :

Les valeurs intérieures du tableau ci-dessous donnent les coefficient $w_{\alpha,n}$ utilisé dans le test de Shapiro-Wilk. Ici, n est la taille de l'échantillon et α est la valeur du risque.

ν \ α	0,99	0,975	0,95	0,90	0,10	0,05	0,025	0,01	0,001
1	0,0002	0,001	0,004	0,016	2,71	3,84	5,02	6,63	10,83
2	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21	13,82
3	0,11	0,22	0,35	0,58	6,25	7,81	9,35	11,34	16,27
4	0,30	0,48	0,71	1,06	7,78	9,49	11,14	13,28	18,47
5	0,55	0,83	1,15	1,61	9,24	11,07	12,83	15,09	20,51
6	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81	22,46
7	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,48	24,32
8	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09	26,12
9	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67	27,88
10	2,56	3,25	3,94	4,87	15,99	18,31	20,48	23,21	29,59
11	3,05	3,82	4,57	5,58	17,28	19,68	21,92	24,73	31,26
12	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22	32,91
13	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69	34,53
14	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14	36,12
15	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58	37,70
16	5,81	6,91	7,96	9,31	23,54	26,30	28,85	32,00	39,25
17	6,41	7,56	8,67	10,09	24,77	27,59	30,19	33,41	40,79
18	7,01	8,23	9,39	10,86	25,99	28,87	31,53	34,81	42,31
19	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19	43,82
20	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57	45,31
21	8,90	10,28	11,59	13,24	29,62	32,67	35,48	38,93	46,80
22	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29	48,27
23	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64	49,73
24	10,86	12,40	13,85	15,66	33,20	36,42	39,36	42,98	51,18
25	11,52	13,12	14,61	16,47	34,38	37,65	40,65	44,31	52,62
26	12,20	13,84	15,38	17,29	35,56	38,89	41,92	45,64	54,05
27	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96	55,48
28	13,56	15,31	16,93	18,94	37,92	41,34	44,46	48,28	56,89
29	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59	58,30
30	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89	59,70

FIG. 3.2 – Table de la loi Chi-deux

n \ α	0,05	0,01
3	0,767	0,753
4	0,748	0,687
5	0,762	0,686
6	0,788	0,713
7	0,803	0,730
8	0,818	0,749
9	0,829	0,764
10	0,842	0,781
11	0,850	0,792
12	0,859	0,805
13	0,856	0,814
14	0,874	0,825
15	0,881	0,835
16	0,837	0,844
17	0,892	0,851
18	0,897	0,858
19	0,901	0,863
20	0,905	0,868
21	0,908	0,873
22	0,911	0,878
23	0,914	0,881
24	0,916	0,884
25	0,918	0,888
26	0,920	0,891

n \ α	0,05	0,01
27	0,923	0,894
28	0,924	0,896
29	0,926	0,898
30	0,927	0,900
31	0,929	0,902
32	0,930	0,904
33	0,931	0,906
34	0,933	0,908
35	0,934	0,910
36	0,935	0,912
37	0,936	0,914
38	0,938	0,916
39	0,939	0,917
40	0,940	0,919
41	0,941	0,920
42	0,942	0,922
43	0,943	0,923
44	0,944	0,924
45	0,945	0,926
46	0,945	0,927
47	0,946	0,928
48	0,947	0,929
49	0,947	0,929
50	0,947	0,930

FIG. 3.3 – Table des valeurs de Shapiro-Wilk