Université Djilali Liabes de Sidi Bel Abbes
Faculté des Sciences Économiques, Commerciales et Sciences de Gestion

---

Analyse des données I
Cours et travaux pratiques

Par : Ezzine Abdelmadjid

Polycopié destiné aux étudiants de la Première Année Master :
Marketing Bancaire & Marketing de services

Année universitaire : 2017-2018

Analyse des données I
Cours et travaux pratiques

Dedicaces

À mon père Ahmed, mon beau-père Kacem, à Djeloul.

Sommaire

Avant-Propos

Ce Polycopié "Analyse des données I" s'adresse spécialement aux étudiants de la Première année Master Marketing et à tous qui veulent apprendre les méthodes statistiques via SPSS.

Notre démarche est basée sur la stratégie suivante :

– Introduire le principe de la méthode statistique étudiée.

– Donner un exemple d'application (en marketing) afin de mieux comprendre la méthode.

– Détailler la procédure à suivre sous SPSS pour la réalisation de cette méthode.

Le chapitre 1 : Travaillez avec SPSS, permet aux étudiants de se familiariser avec Le logiciel SPSS.

Le chapitre 2 : Statistiques descriptives, fournit les analyses de base de données.

Le chapitre 3 : Les tests univariés, expose un panorama de tests statistiques qu'un chercheur en marketing doit les maitriser.

Le chapitre 4 : Analyse de la variance, est la suite du chapitre 3, il présente quelque techniques statistiques liées au design de recherche expérimentale.

Les travaux pratiques à la fin de chaque chapitre permettent aux étudiants de tester leur capacité à passer de la théorie à la pratique.

Chapitre 1

Travailler avec SPSS

L'objet de ce chapitre est présenter le logiciel SPSS, d'indiquer comment ce logiciel peut être mis en oeuvre et de décrire quelques fonctions élémentaires de SPSS.

1.1 La mise en oeuvre

Lorsque vous lancez SPSS version 18,19 ou 20, une boite de dialogue apparait qui vous permet de sélectionner ce que vous voulez faire [6] (voir figure1.1)

FIGURE 1.1 – Première boite de dialogue de SPSS

Cette boite de dialogue : **Que voulez-vous faire ?** vous propose plusieur options et c'est à vous de décider :

- **Ouvrir une source de donnée existante** (pour ouvrir un fichier de données). il vous faudra alors sélectionner le fichier approprié et cliquer sur **OK**.
- **Ouvrir un autre type de données** (Importer les données d'un autre programme(Excel))
- **Saisir des données** (Pour entrer de nouvelles données)

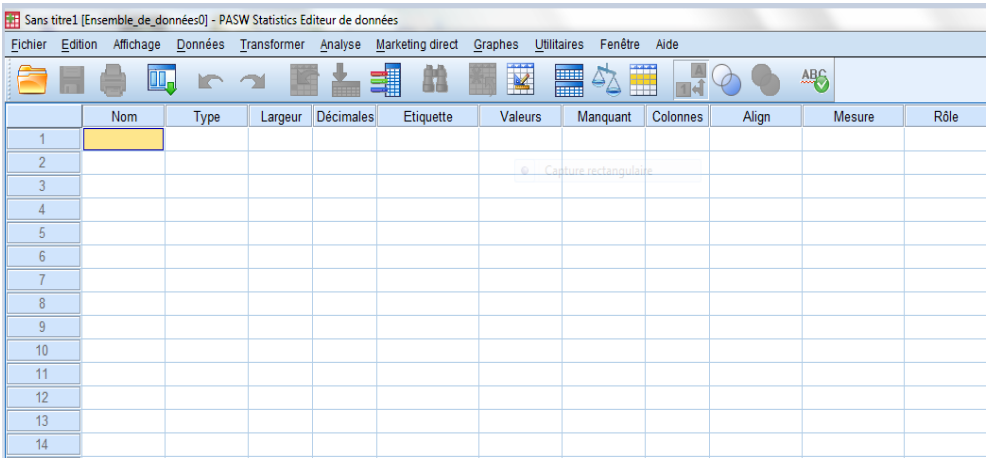


FIGURE 1.2 – Editeur des variables

Si vous avez choisi la dernière option (revient à cliquer tout simplement sur **Annuler**) l'écran de la figure1.2 apparaitra :

Si l'écran n'est pas semblable à celui présenté ci-dessus (figure1.2), cliquez sur l'onglet

Affichage des variables

en bas à gauche de la fenêtre

1.1.1 SPSS : fonctions assurées et commandes principales

Les fonctions assurées par SPSS sont[14] :

- **La gestion des données**
- **La mise en forme des informations**
- **Le traitement statistiques des données**

Les principales commandes des menus SPSS figurent ci-dessous [1].

- **Fichier :** est le menu qui concerne le fichier de travail.
- **Édition :** contient les commandes servant à couper, copier et coller du texte.
- **Affichage :** pour afficher notamment les noms de variables ou les données.
- **Données :** il permet de définir des variables et d'insérer de nouvelles informations et de nouvelles variables
- **Transformer :** joue aussi un rôle essentiel, qui est de transformer les variables selon les besoins de l'analyse des données.
- **Analyser :** renferme les principales procédures statistiques, les plus connues et les plus utilisées.
- **Graphes :** permet de créer des graphiques de toutes les formes possibles.
- **Utilitaires :** propose deux façons d'afficher les informations : par le nom des variables ou par leur contenu.
- **Fenêtre :** donne un accès facile et rapide aux fenêtres d'applications, de définition des variables et aux fenêtres des résultats de l'application des commandes.
- **Aide :** fournit des indications sur les façons d'utiliser les commandes de SPSS et sur les diverses procédures statistiques.
-

1. Pour plus de détaille vous pouvez consulter[14] page et [16] page 39

1.2 Taper directement les données sur SPSS

Afin de discuter les différents éléments importants pour entrer les données, on utilisera l'exemple suivants. Supposons qu'on veut taper le tableau ci-dessous dans SPSS :

Prénom	**Genre**	**Taille(cm)**	**Poids(kg)**
Ahmed	1	185	80
Khadija	0	170	72
Ibrahim	1	180	85
Fatima	0	175	75
Kacem	1	190	80

TABLE 1.1 – Exemple pour la mise en oeuvre de SPSS

On commence par déclarer les variables. Sur l'écran de visualisation des variables, une ligne représente une variable et les colonnes donnent les caractéristiques fondamentales des variables. Faites entrer les (un par un) noms des variables dans la colonne *Nom* ; cliquez sur la première ligne de cette colonne et entrer le nom de la première variable :**Prénom**, la deuxième ligne est faite pour la deuxième variables : **Genre** et ainsi de suite.

Remarque 1 *Le nom de la variable ne doit contenir aucune espace ou caractère spécial.*

Maintenant c'est le moment pour déclarer les caractéristiques des variables.

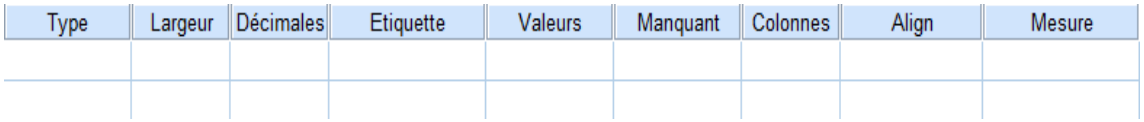


FIGURE 1.3 – Caractéristiques fondamentales des variables

- **Type :** Il y a plusieurs types de données. le type **Numérique** (le plus utilisé) est indiqué pour les variables dont les valeurs contiennent des nombres (taille, poids). Par contre le type **Chaîne** est réservé pour les variables textuelles (Prénom) (voir figure1.4).
- **Etiquette :** Pour donner aux variables des titres plus clairs (peuvent contenir des espaces). Par exemple, pour la variable taille on peut écrire comme étiquette : *Taille des étudiants première année master* .
- **Valeurs :** Les groupes d'individus sont représentés par des numéros (codage)(voir figure 1.5).
- **Manquant :** Cette colonne sert à désigner des numéros aux valeus manquantes (on préconise le numéro 99)(voir figure 1.6).
- **Mesure :** Pour déterminer le niveau de mesure d'une variable : Nominal Ordinal ou Echelle. pour plus de détaille sur les échelles de mesure primaires voir Repère1.

Repère 1 *:(* **Les échelles de mesure)** *On distingue quatre types d'échelles. Nous les présenterons de la plus générale à la plus restrictive. Les opérations autorisées seront rares dans les premières et plus nombreuses dans les suivantes* [a].

L'échelle nominale *Il s'agit de répartir les individus en catégories. Les modalités jouent le role d'étiquettes. L'exigence de base et que chaque individu doit pouvoir recevoir une affectation et une seule[6].*

L'échelle ordinale *est une échelle de classement comme l'échelle nominale, dans laquelle les nombres attribués à chaque modalité ont une relation d'ordre avec un continuum sous-jacent. On peut, par exemple, utiliser une échelle ordinale pour classer les préférence de marques[3].*

L'échelle d'intervalles *Les échelles d'intervalles représentent le premier niveau des échelles métriques. Elle permet de tenir compte de la différence entre deux valeurs d'une variable.*

L'échelle de rapports *Elle est sans conteste l'échelle la plus riche en propriétés. Elle possède un zéro "naturel" qui indique l'absence du phénomène étudié[6].*

a. Le chapitre 8 du livre de Malhotra[12] Présente une introduction rigoureuse, riche en exemples pour ceux qui veulent approfondir leur connaissances en ce qui concerne la notion de "mesures et échelles"

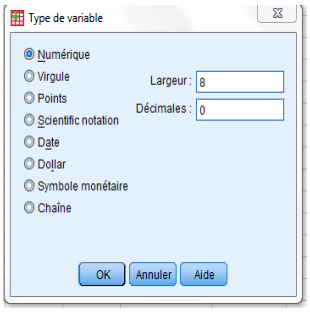


FIGURE 1.4 – Types des variables

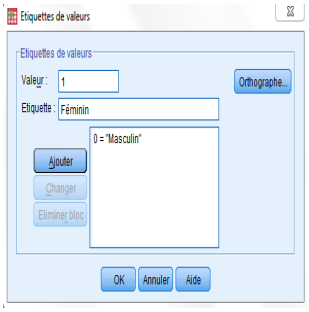


FIGURE 1.5 – Etiquettes de valeurs

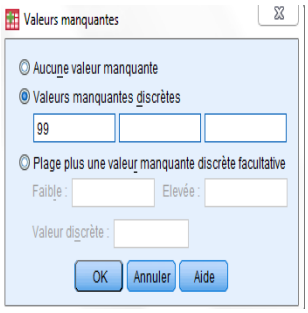
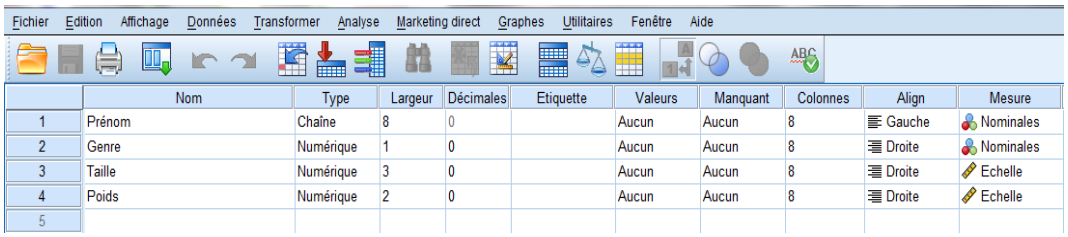


FIGURE 1.6 – Valeurs Manquantes

Quand vous avez fini, l'écran doit être semblable à ceci :



	Nom	Type	Largeur	Décimales	Etiquette	Valeurs	Manquant	Colonnes	Align	Mesure
1	Prénom	Chaîne	8	0		Aucun	Aucun	8	Gauche	Nominales
2	Genre	Numérique	1	0		Aucun	Aucun	8	Droite	Nominales
3	Taille	Numérique	3	0		Aucun	Aucun	8	Droite	Echelle
4	Poids	Numérique	2	0		Aucun	Aucun	8	Droite	Echelle
5										

FIGURE 1.7 – Déclaration des variables

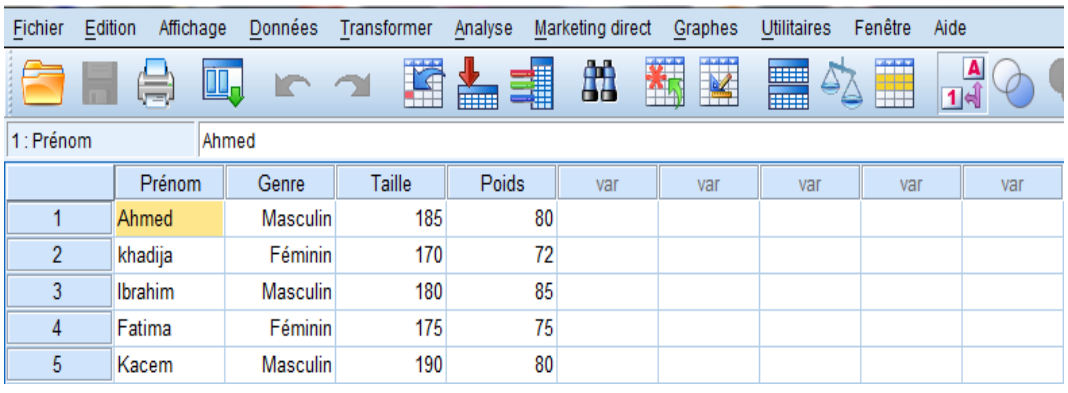
Vous pouvez maintenant saisie les données. Pour cela, cliquez sur l'onglet

Affichage des données

Remarque 2

- *Les colonnes sont nommées Prénom, Genre, Taille et Poids.*
- *Chaque colonne correspond à une variable et chaque ligne à un individu.*
- *Toutes les données correspondant à la variable* **Prénom** *sont dans la première colonne et celles correspondant à la variable* **Genre** *dans la seconde....*

Maintenant faire entrer les données du tableau 1.1. Quand vous avez fini, l'écran doit ressembler à :



	Prénom	Genre	Taille	Poids	var	var	var	var	var
1	Ahmed	Masculin	185	80					
2	khadija	Féminin	170	72					
3	Ibrahim	Masculin	180	85					
4	Fatima	Féminin	175	75					
5	Kacem	Masculin	190	80					

FIGURE 1.8 – Déclaration des données

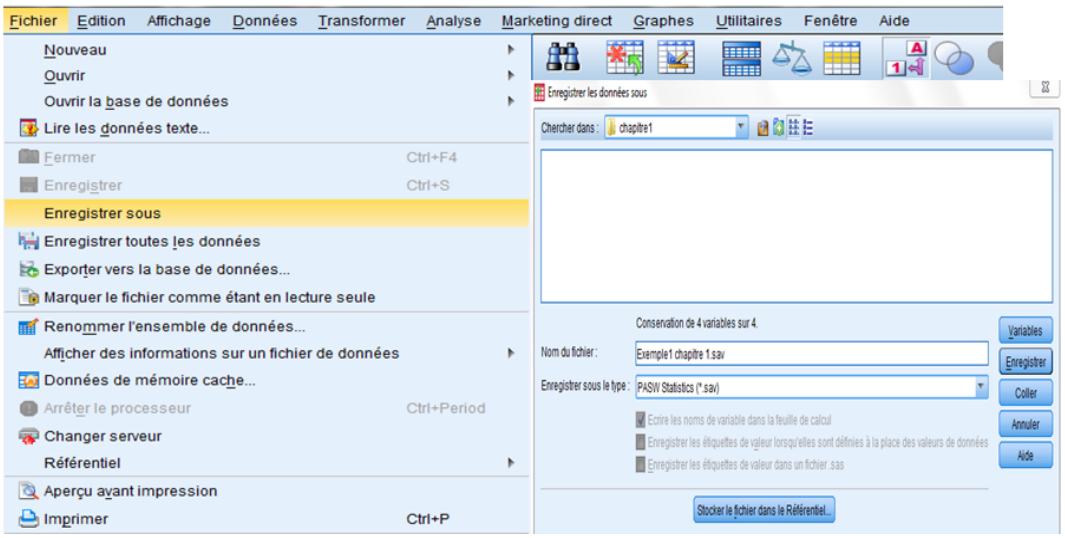


FIGURE 1.9 – Enregistrer les données

1.3 Enregistrer les données

Après la saisie des données, n'oubliez pas de les enregistrer dans un fichier. Pour ce faire, voir la figure 1.9 : Dans la boite de dialogue apparait, il suffit de taper le nom que vous souhaitez attribuer à votre fichier (par exemple Exemple1 chapitre1), puis de cliquer sur **Enregistrer**.

Bravo !

Vous avez créer votre fichier et les données sont enregistrées.

Remarque 3 *: Il n'est pas de taper l'extension* **.sav**, *elle sera faite automatiquement par SPSS.*

1.4 Création ou calcul d'une nouvelle variable

Supposons que nous voulons inclure une colonne suplémentaire dans notre exemple indiquant le BMI (Body-Mass Index),indice de mass corporel ; Le BMI est défini par la mass corporelle en kilogramme divisée par le carré de la taille en mettre. :

$$BMI = \frac{Poids_{kg}}{Taille/100} \tag{1.1}$$

la figure1.10 montre comment ajouter cette nouvelle variable.

a) La figure 1.10 indique le chemin suivi pour le calcul d'une nouvelle variable :

Transformer ⇒ Calculer la variable

b) La case **Variable cible** dans la boite de dialogue est réservée pour le nom de la nouvelle variable.

c) Dans la case **Expression numérique** On tapera la formule 1.1.

d) L'icone ↪ de la boite de dialogue facilite le glissement des variables vers la case **Expression numérique**

La variable BMI est maintenant affichée sur l'écran **Affichage des données** (voir figure 1.11)

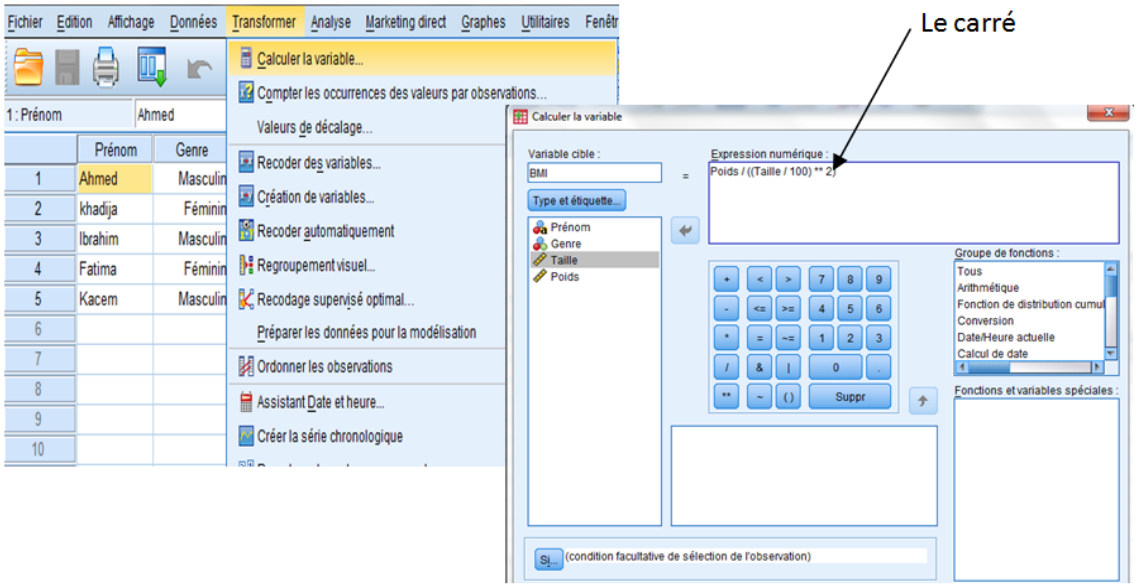


FIGURE 1.10 – Calculer une nouvelle variable

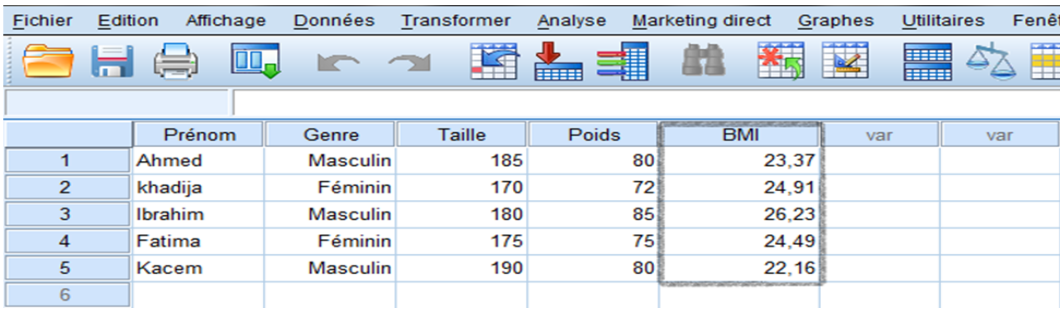


FIGURE 1.11 – L'ajout de la variable BMI

1.5 Etudier un sous ensemble d'observation

Pour réaliser cet objectif SPSS propose deux techniques, l'une (**Sélection des observations**)permetant d'écarter un sous ensemble d'individus en selectionnant un sous échantillon suivant des critères bien précis. L'autre technique (**Scinder un fichier de données**) garde l'ensemble des individus en le subdivisant en plusieurs groupes.

1.5.1 Sélection des observations

Si on veut mener une étude sur un nombre spécifique d'observations (des cas). Il est possible de créer un fichier séparé en supprimant de façon temporaire les individus non inclus dans l'étude. Par exemple, si on veut que notre étude se focalise uniquement sur les hommes, il suffit donc de passer par les étapes suivantes :(voir figure 1.12)

1. Suivre le chemin suivant Données ⇒ Selectionner des observations
2. Cocher l'obtion **Selon une condition logique** et cliquer sur **Si**
3. Glisser la variable **Genre** et mettre la égale à 0 (0 est le code pour genre masculin).
4. Cliquer sur **Poursuivre** puis sur **Ok**.

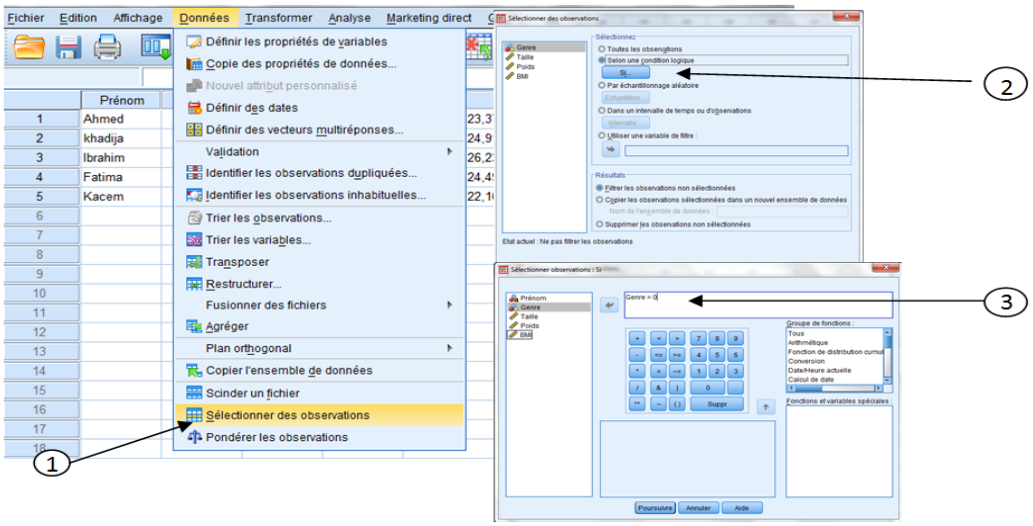
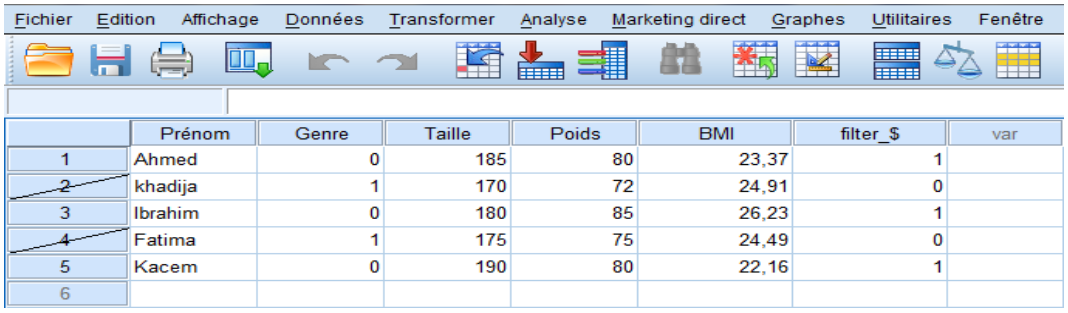


FIGURE 1.12 – Le processus de sélection d'un sous ensemble d'individus

La figure 1.13 mettre en évidence la fin de ce processus.



	Prénom	Genre	Taille	Poids	BMI	filter_$	var
1	Ahmed	0	185	80	23,37	1	
2	khadija	1	170	72	24,91	0	
3	Ibrahim	0	180	85	26,23	1	
4	Fatima	1	175	75	24,49	0	
5	Kacem	0	190	80	22,16	1	
6							

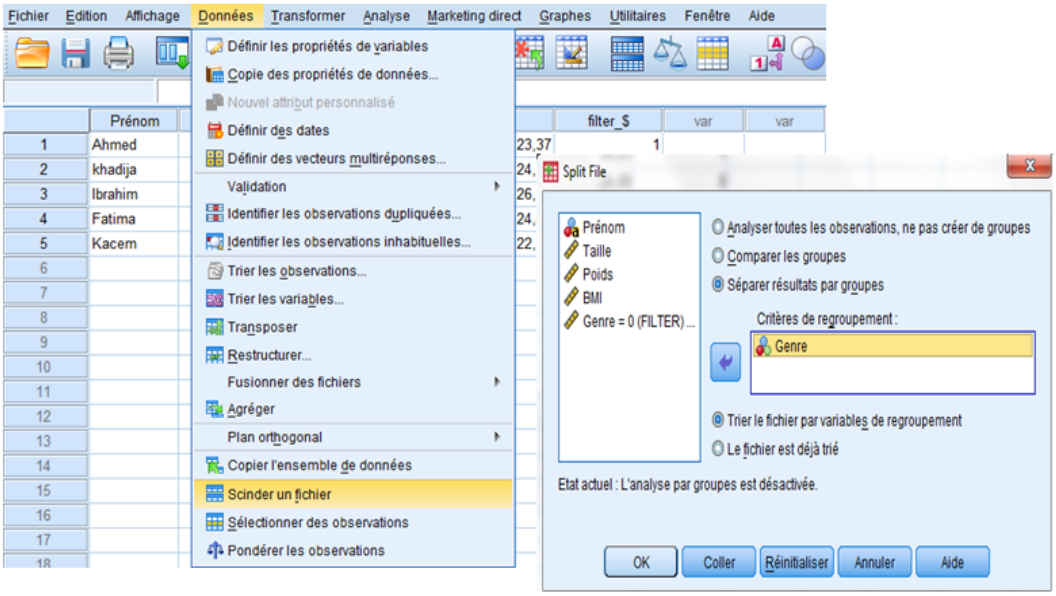
FIGURE 1.13 – Suppression des femmes

Remarque 4 *:*

- *Vous remarquez sur la figure 1.13 qu'une nouvelle variable (filter_$) était crée, cette variable indique si un individu a été sélectionné (=1) ou non (=0).*
- *Si vous voulez travailler sur tous les individus ; répéter l'étape 1) et cocher l'option* **Toutes les observations** *puis cliquez sur* **Ok**. *La variable (filter_$) ne va pas disparaitre, vous pouvez l'utiliser à n'importe quel moment.*

1.5.2 Scinder un fichier de données

Dans cette option, une variable qualitative (Genre) est utilisée pour faire la subdivision. Si on lance une analyse statistique, celle ci sera faite sur les données de chacun des sous groupes en parallèle.
Supposons que vous voulez lancer (en même temps) deux analyse statistiques séparées, une pour les femme et la deuxième pour les hommes, alors la variable **Genre** sera utilisée comme variable se séparation. La figure 1.14 met en évidence le chemin et la procédure à suivre pour réaliser cette opération.



,

FIGURE 1.14 – Suppression des femmes

1.6 Recodage des variables

Supposons qu'on a trois variables formulées comme des déclarations à propos desquelles les répondants indiquent leurs degré d'accord ou désaccord. Ces variables sont évaluées sur une échelle de 7 points (1= pas tout à fait d'accord jusqu'au 7= tout à fait d'accord).

Q1 : Un dentifrice doit renforcer les gencives.

Q2 : Un dentifrice doit rendre les dents brillantes.

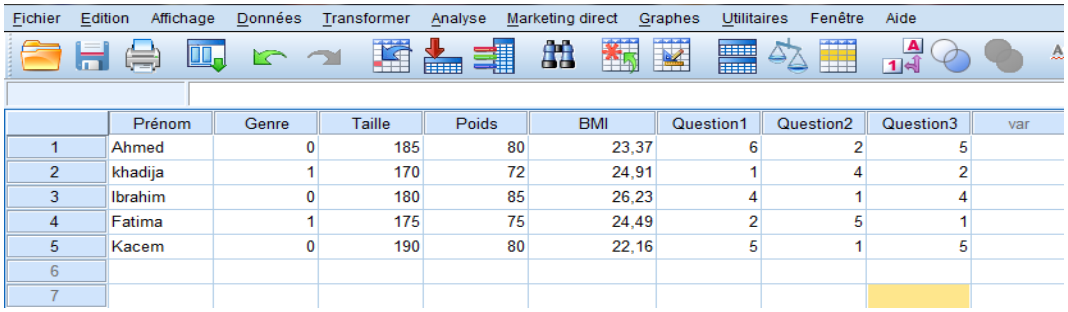
Q3 : La prévention contre les caries n'est pas un avantage important du dentifrice.

Ces questions portent sur trois avantages fondamentaux recherchés par les individus lors de l'achat d'un dentifrice Les deux premières sont exprimées de façon affirmative, par contre la troisième prend une forme négative. Les réponses sont affiché dans Le tableau 1.2

Prénom	Question1	Question2	Question3
Ahmed	6	2	5
Khadija	1	4	2
Ibrahim	4	1	4
Fatima	2	5	1
Kacem	5	1	5

TABLE 1.2 – Réponses sur les avantages du dentifrice

Maintenant c'est à vous de jouer pour saisir les données comme le montre le figure 1.15. Sinon voir la section 1.2 pour réaliser cette tache.



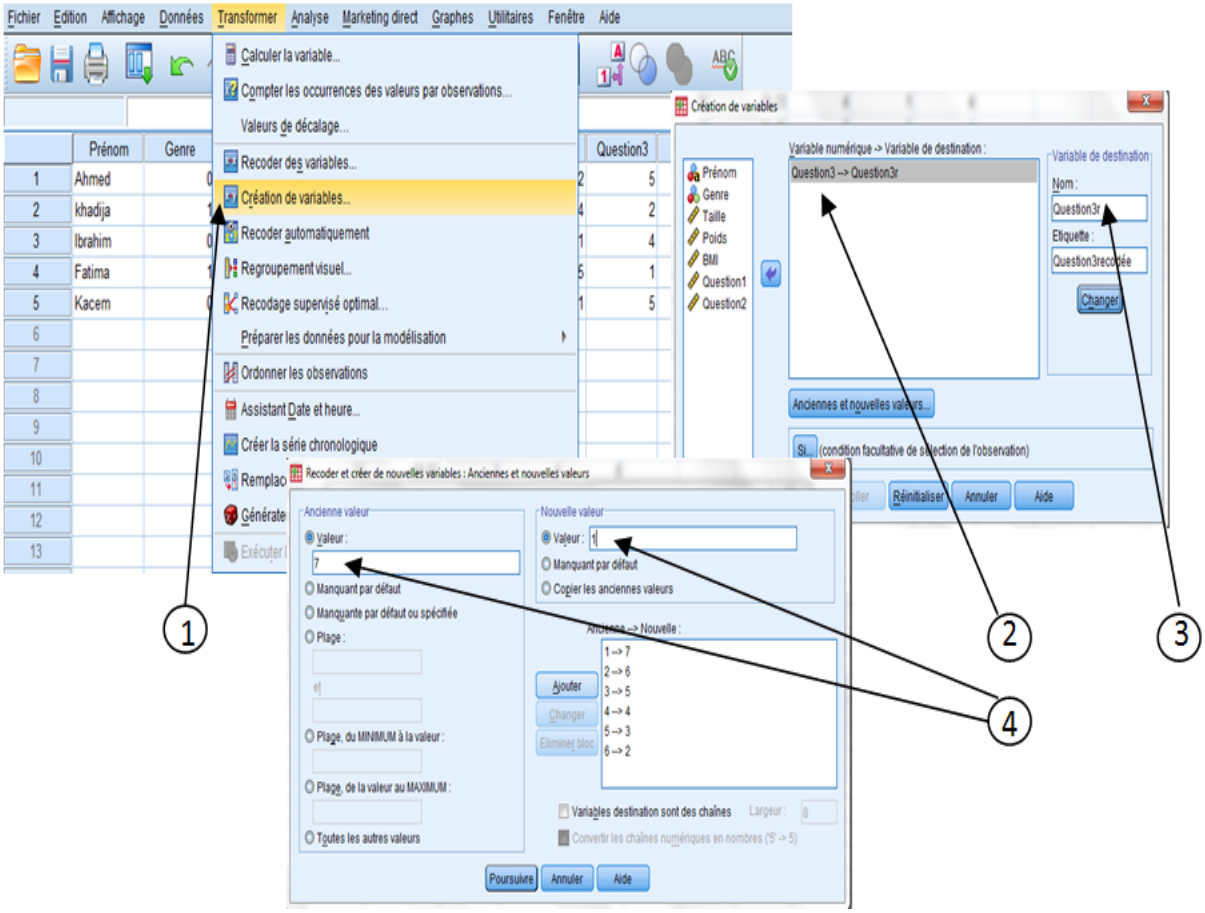
	Prénom	Genre	Taille	Poids	BMI	Question1	Question2	Question3	var
1	Ahmed	0	185	80	23,37	6	2	5	
2	khadija	1	170	72	24,91	1	4	2	
3	Ibrahim	0	180	85	26,23	4	1	4	
4	Fatima	1	175	75	24,49	2	5	1	
5	Kacem	0	190	80	22,16	5	1	5	
6									
7									

,

FIGURE 1.15 – Saisir les données dentifrice

Si on veut attribuer à chaque individu un score total mesurant son attitude global envers les avantages du dentifrice via les trois variables **Q1, Q2, Q3**, on doit s'assurer premièrement que ces trois variables sont scalées dans la même direction. Hors la question 3 prend une forme négative d'où le recours au recodage de cette variable. Pour ce faire on doit suivre les étapes suivantes comme l'indique la figue 1.16.

1. Suivre le chemin Transformer ⇒ Création de variables
2. Faire glisser la variable **Question3** dans la case **Variable numérique** → **Variable de destination**
3. Sous **Variable de destination** taper le nom de la nouvelle variable **Question3r** dans la case **Nom** puis cliquer sur **Changer**.
4. Cliquer sur le bouton **anciennes et nouvelles valeurs**, par exemple pour recoder la dernière valeur 7, mettre la dans la case **Valeur** sous **Ancienne valeur** et la valeur 1 dans la case **Valeur** sous **Nouvelle valeur**, puis cliquer sur le bouton **Ajouter**
5. Enfin cliquer sur **Poursuivre** puis sur **Ok**.



,

FIGURE 1.16 – Recodage de variable

Remarque 5 *Pour ce type de recodage, on peut utiliser la formule*

$$Question3r = 8 - Question3$$

pour calculer cette nouvelle variable voir section 1.4

Chapitre 2

Les statistiques descriptives

Introduction

Ce chapitre décrit les techniques élémentaires (Statistiques descriptives) auxquelles le chargé d'études peut avoir recours pour réduire les données en une forme graphique (Représentation graphique de données)ou numérique (Mesures de position centrale et mesures de dispersion) et ceci dont le but d'avoir une bonne interprétation de ses données.
L'exemple suivant va nous accompagner tout au long de ce chapitre afin d'illustrer certaines étapes de ce processus.

Exemple 1 *(voir [12])*

Dans le cadre d'un pré-test, certaines données relatives à la marque Nike ont été recueillies auprès de 45 clients. Ces données figurent dans le tableau 2.1. Elles concernent le niveau d'utilisation, le sexe, la notoriété, l'attitude, la préférence, l'intention et la fidélité vis-à-vis de la marque Nike. Le niveau d'utilisation a été codé 1, 2 ou 3, selon qu'il était faible, moyen, ou important. Le sexe a été codé 1, pour les femmes, et 2 pour les hommes. La notoriété, l'attitude, la préférence, l'intention et la fidélité ont été mesurées sur une échelles de type Likert en sept points (1=très défavorable, 7= très favorable). On notera que cinq répondants présentent des valeurs manquantes, notées 9.

Numéro	Utilisation	Sexe	Notoriété	Attitude	Préférence	Intention	Fidélité
1	3	2	7	6	5	5	6
2	1	1	2	2	4	6	5
3	1	1	3	3	6	7	6
4	3	2	6	5	5	3	2
5	3	2	5	4	7	4	3
6	2	2	4	3	5	2	3
7	2	1	5	4	4	3	2
8	1	1	2	1	3	4	5
9	2	2	4	4	3	6	5
10	1	1	3	1	2	4	5
11	3	2	6	7	6	4	5
12	3	2	6	5	6	4	4
13	1	1	4	3	3	3	3
14	3	2	6	4	5	3	2
15	1	2	4	3	4	5	6
16	1	2	3	4	2	4	2
17	3	1	7	6	4	5	3
18	2	1	6	5	4	3	2
19	1	1	1	1	3	4	5
20	3	1	5	7	4	1	2
21	3	2	6	6	7	7	5
22	2	2	2	3	1	4	2
23	1	1	1	1	3	2	2
24	3	1	6	7	6	7	6
25	1	2	3	2	2	1	1
26	2	2	5	3	4	4	5
27	3	2	7	6	6	5	7
28	2	1	6	4	2	5	6
29	1	1	9	2	3	1	3
30	2	2	5	9	4	6	5

TABLE 2.1 – Données relatives à la marque Nike

Numéro	Utilisation	Sexe	Notoriété	Attitude	Préférence	Intention	Fidélité
31	1	2	1	2	9	3	2
32	1	2	4	6	5	9	3
33	2	1	3	4	3	2	9
34	2	1	4	6	5	7	6
35	3	1	5	7	7	3	3
36	3	1	6	5	7	3	4
37	3	2	6	7	5	3	4
38	3	2	5	6	4	3	2
39	3	2	7	7	6	3	4
40	1	1	4	3	4	6	5
41	1	1	2	3	4	5	6
42	1	1	1	3	2	3	4
43	1	1	2	4	3	6	7
44	1	1	3	3	4	6	5
45	1	1	1	1	4	5	3

TABLE 2.2 – Données relatives la marque Nike(suite)

Sourse :[12] page 410

2.1 Représentation graphique de données

L'information chiffrée, une fois recueillie, doit être transmise de façon rigoureuse sur un support afin de faire passer le message chiffré. Le premier support est le **tableau statistique**. Vous pouvez jeter un coup d'oeil sur le tableau 2.1.

1. Il est plus précis que le support **diagramme** (sous section2.1.1) mais moins direct et moins convivial.
2. Il est plus riche en information que le support **calcul de caractéristiques** (sous section 2.2), mais plus lourd à manier (il comprend trop de nombres). Tout court il est moins synthétique.

Repère 2 :**La consigne de présentation d'un tableau :***([15] page 51) Elle implique le respect de quatre principes fondamentaux.*

1. **La source :** *c'est l'origine exacte du tableau.*
2. **Le titre :** *ils doit être suffisamment complet et précis pour qu'on ne puisse pas avoir un seul questionnement sur la définition du phénomène étudié.*
3. **Les intitulés des lignes et des colonnes :***Ils relèvent de la même précision*
4. **Les unités utilisées :** *elles doivent êtres précisées ainsi que leur ordre de grandeur.*

L'une des méthodes les plus simples de réorganiser les données pour les rendre plus intelligibles est d'en faire un graphique. Pour ce faire nous évoquerons tour à tour les distributions de fréquences, les histogrammes et les diagrammes en tiges et feuilles[2]

2. Les grandes lignes de ce chapitre sont inspirées de ceux du chapitres 2 du livre de D.Howell [9]

		Effectifs	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	Très défavorable	5	11,1	11,4	11,4
	2	5	11,1	11,4	22,7
	3	6	13,3	13,6	36,4
	4	7	15,6	15,9	52,3
	5	7	15,6	15,9	68,2
	6	10	22,2	22,7	90,9
	Très favorable	4	8,9	9,1	100,0
	Total	44	97,8	100,0	
Manquante	9	1	2,2		
Total		45	100,0		

FIGURE 2.1 – Distribution de fréquences relative au niveau de connaissance de la marque Nike

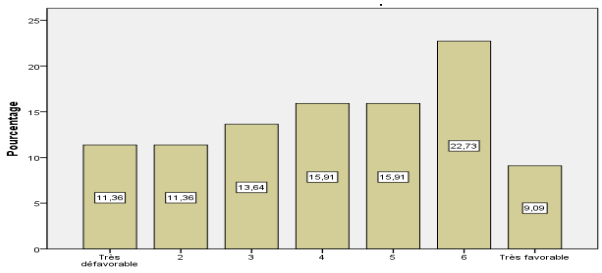


FIGURE 2.2 – Diagramme en barres des fréquences des valeurs de la variable Notoriété

2.1.1 Distributions de fréquences

Un chargé d'études se posent souvent des questions portant sur une seule et unique Variable, par exemple (voir[9] page 365) :

- Parmi les clients d'une marque, combien d'entre eux peuvent-ils être qualifiés de fidèles ?
- Quel est le Pourcentage d'utilisateurs assidus, d'utilisateurs moyens, d'utilisateurs occasionnels et non-utilisateurs ?
- Quelle est la distribution des revenus des clients de la marque ? Cette distribution est-elle orientée vers la tranche inférieure ?

Les réponses à ce genre de questions peuvent êtres obtenues par l'examen des distributions des fréquences. La distribution de fréquences d'une variable se représente sous la forme d'un tableau répertoriant les effectifs,Fréquences et fréquences cumulées de toutes ses valeurs, ou un graphique ayant en abscisse les valeur de la variable et en ordonnée les effectifs, ou fréquences de ces valeurs [1].

Prenons à titre d'exemple la variable **Notoriété** :

- La distribution de fréquences permet d'évaluer plus facilement l'importance des non réponses (9 sur 45 dans la figure 2.1).
- La distribution de fréquences (figure 2.2) indique aussi l'allure de la distribution empirique de la variable, on peut facilement vérifier la cohérence de la distribution observé en regard de celle que l'on attendait (une distribution normale, par exemple)

1. On parle de diagramme en barres(ou en tuyaux) si la variable est qualitative et diagramme en bâtons si la variable est quantitative (voir [6])

Procédure sous SPSS

1. Suivre le chemin suivant : **Analyse** $\Longrightarrow$ **Statistiques descriptives** $\Longrightarrow$ **Effectifs**.
2. Glisser la variable **Notoriété** dans la case **Variable(s)** et cocher la case **Afficher les tableaux des effectifs**.
3. Cliquer sur **Diagrammes** et sélectionner **Diagramme en bâtons** avant de cliquer sur **Poursuivre**

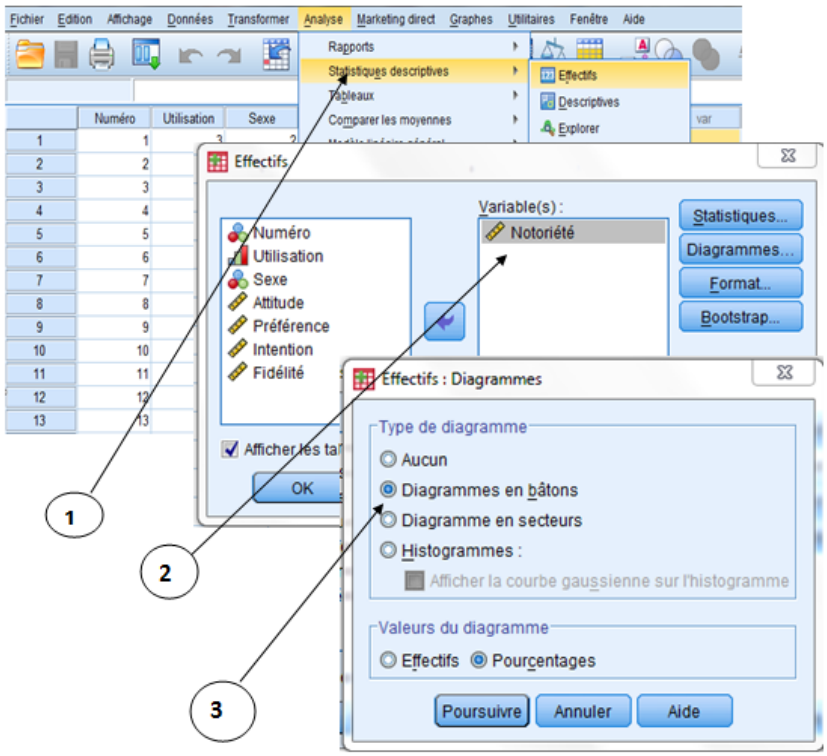


FIGURE 2.3 – Distribution de fréquences sous SPSS

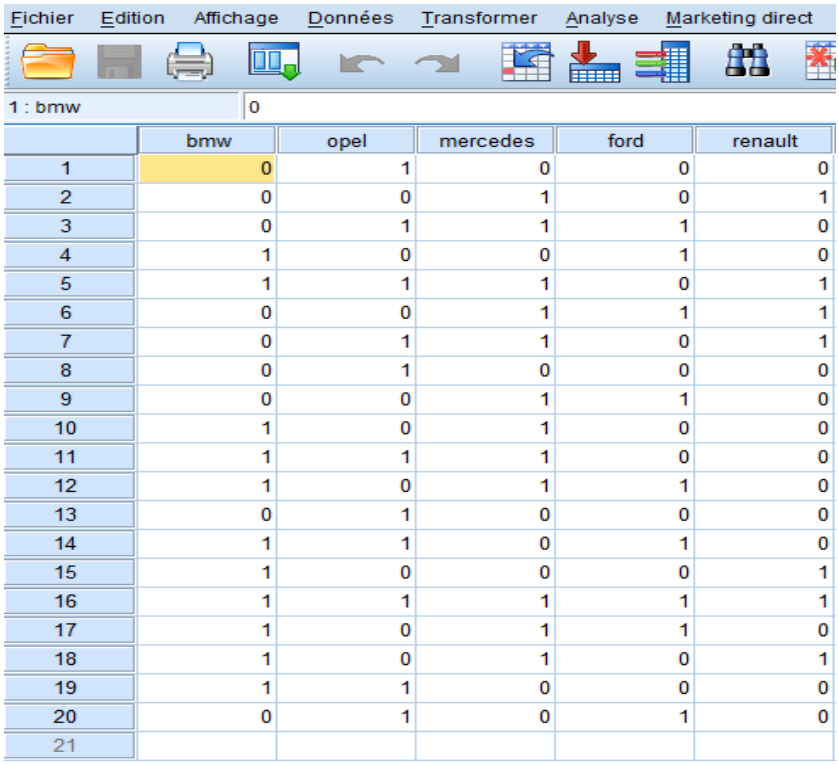
Questions à choix multiples

Dans les questions liées aux variables proposées pour l'étude de la marque Nike, une seule modalité doit être choisie parmi une liste de réponses proposées. Dans cette section on va étudier un autre type de questions (Fermées multiples) où le répondant peut choisir une ou plusieurs réponses. deux cas se présentent. Soit on limite le nombre de réponses possibles, Soit on ne le limite pas. Pour plus de détaille sur les différents formats de réponses voir [8]

Exemple 2 *[10]***Question avec nombre de réponses non limité :** *Voici un nombre de marques de voitures. Indiquer la ou les marque(s) que vous trouvez attractive(s)*

- *BMW*
- *Opel*
- *Mercedes*
- *Ford*
- *Renault*

Pour chaque marque, une variable est crée avec :
Code 0 si la marque est non indiqué Code 1 sinon.(voir figure2.4)



	bmw	opel	mercedes	ford	renault
1	0	1	0	0	0
2	0	0	1	0	1
3	0	1	1	1	0
4	1	0	0	1	0
5	1	1	1	0	1
6	0	0	1	1	1
7	0	1	1	0	1
8	0	1	0	0	0
9	0	0	1	1	0
10	1	0	1	0	0
11	1	1	1	0	0
12	1	0	1	1	0
13	0	1	0	0	0
14	1	1	0	1	0
15	1	0	0	0	1
16	1	1	1	1	1
17	1	0	1	1	0
18	1	0	1	0	1
19	1	1	0	0	0
20	0	1	0	1	0
21					

FIGURE 2.4 – Procédure

La procédure ci-dessous vous montre comment déclarer ce genre de questions sous SPSS.

Procédure sous SPSS

1. Suivre le chemin suivant **Analyse** $\Longrightarrow$ **Réponses multiples** $\Longrightarrow$ **Définir des groupes de variables**
2. Glisser les variables dichotomiques (BMW $\longrightarrow$ Renault) dans la case **Variables de l'ensemble**
3. Dans la case **Nom** saisir le nom de la variable : Marques et Marques attractives dans la case Étiquette.
4. Cocher sur **Dichotomies** et mettre 1 dans la case **Valeur comptée**. Cliquer sur **Ajouter** puis sur **Fermer** (Voir la figure2.5).

Pour afficher la distribution des fréquences de cette variable (voir figure 2.6), il suffit de suivre le chemin suivant :

Analyse $\Longrightarrow$ **Réponses multiples** $\Longrightarrow$ **Effectifs**.

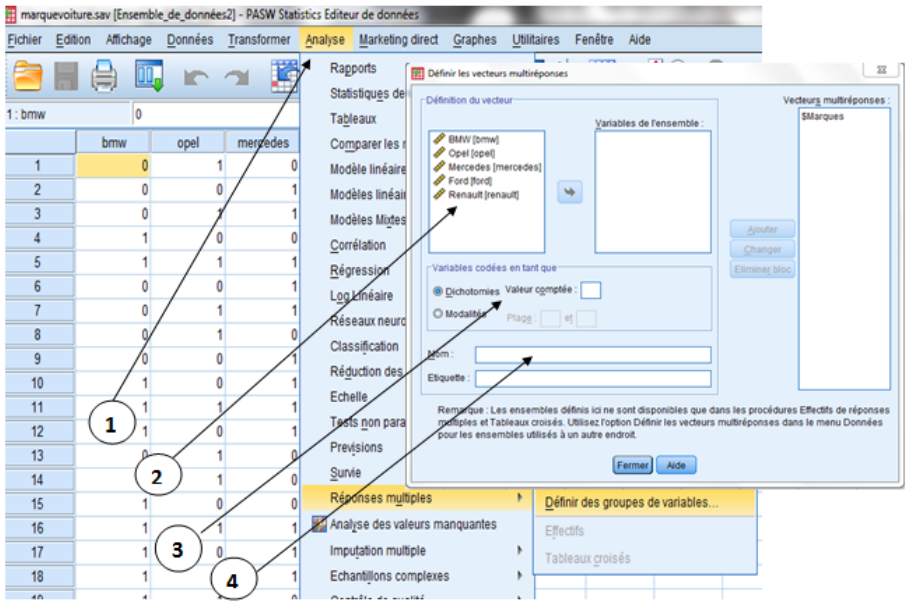


FIGURE 2.5 – Procédure de déclaration d'une question fermé multiple : nombre de réponses non limité

		Réponses		Pourcentage d'observations
		N :	Pourcentage :	
Marques attractives[a]	BMW	11	22,0%	55,0%
	Opel	11	22,0%	55,0%
	Mercedes	12	24,0%	60,0%
	Ford	9	18,0%	45,0%
	Renault	7	14,0%	35,0%
Total		50	100,0%	250,0%

a. Groupe de dichotomies tabulé à la valeur 1.

FIGURE 2.6 – Distribution de fréquences de la variable : Marques attractives

Exemple 3 *[10]***Nombre de réponses limité** *Voici un nombre de marques de voitures. Indiquer la ou les marque(s) que vous trouvez attractive(s) (deux réponses au maximum)*

- *BMW(1)*
- *Opel(2)*
- *Mercedes(3)*
- *Ford(4)*
- *Renault(5)*

Dans ce cas on crée seulement deux variables/ Choix1 et Choix2.

Procédure sous SPSS

La même procédure que celle de l'exemple2 sauf dans le menu **Définir les vecteurs multiréponses** cocher dans la case **modalités**, voir figure2.7

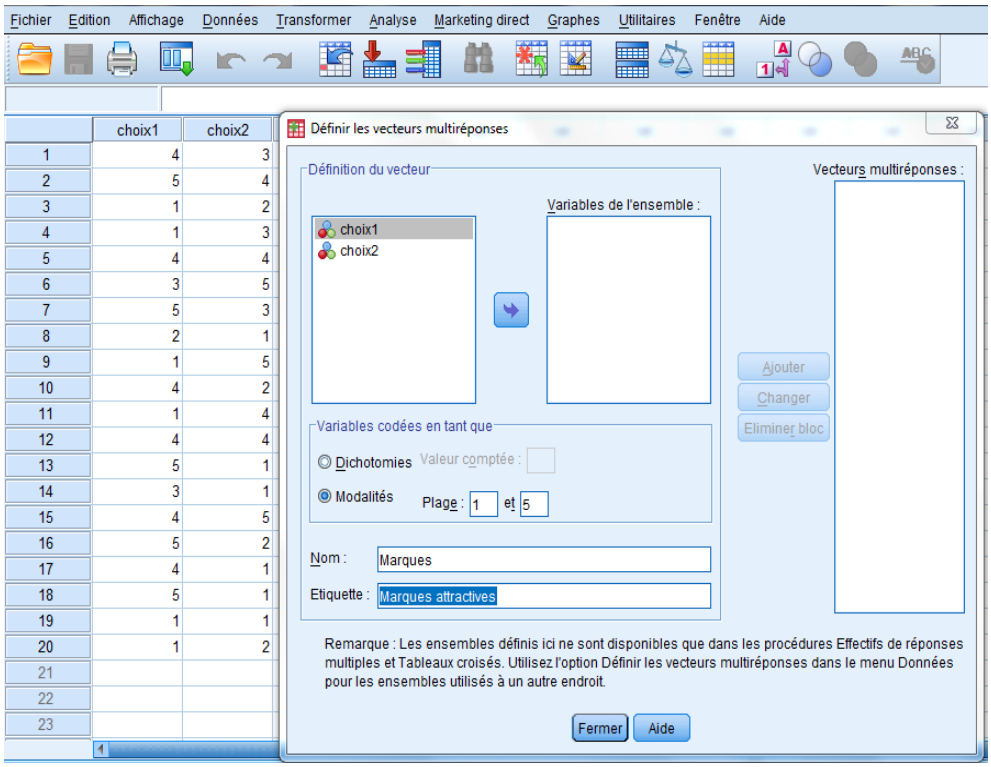


FIGURE 2.7 – Procédure de déclaration d'une question fermé multiple : nombre de réponses limité

Pour afficher la distribution des fréquences de cette variable (voir figure 2.8), il suffit de suivre le meme chemin que dans l'exemple 2 :

Analyse $\Longrightarrow$ **Réponses multiples** $\Longrightarrow$ **Effectifs**.

		Réponses		Pourcentage d'observations
		N :	Pourcentage :	
Marques attractives[a]	BMW	12	30,0%	60,0%
	Opel	5	12,5%	25,0%
	Mercedes	5	12,5%	25,0%
	Ford	10	25,0%	50,0%
	Renault	8	20,0%	40,0%
Total		40	100,0%	200,0%

a. Groupe

FIGURE 2.8 – Distribution de fréquences de la variable : Marques attractives

2.1.2 Histogrammes

L'étude empirique d'une variable statistique discrète dont le nombre des valeurs possibles est très élevée s'appuie sur le principe d'étude d'une variable continue Elle nécessite le groupement des unités statistiques par tranche ou classe de valeurs afin d'établir une présentation numérique puis graphique (histogramme)simple est lisible([2].

Il s'agit de construire, dans un système d'axes coordonnées, une suite de rectangles associés à chacune des classes et dont la surface est égale à l'effectif de cette classe. Il est important de préciser que cette construction n'est pas équivalente à celle qui définit la hauteur des rectangles par l'effectif, sauf évidemment quand toutes les classes ont la même longueur[6].
Le graphique présenté (de la variable notoriété) dans la figure2.9 est appelé un histogramme des pourcentages [2]

FIGURE 2.9 – Histogramme des pourcentages de la variable Notoriété

Procédure sous SPSS

Pour obtenir un histogramme à trois classes de la variable Notoriété (figure2.9), il suffit de suivre les étapes suivantes.

1. Dans le menu **Graphes** sélectionner **Générateur de diagrammes**, puis cliquer sur **Ok**.
2. Cliquer sur **Galerie** et choisir **Histogramme** parmi la liste.
3. Quatre modèles sont disponibles, cliquer deux fois sur le modèle **Histogramme :simple** .
4. Déplacer la variable **Notoriété** dans la zone **Axe de X ?** votre écran sera similaire à celui de la figure 2.10.
5. Cliquer sur **Propriétés des éléments** puis sur le bouton **Définir les paramètre** Vous pouvez choisir diverses options grâce à la dernière boite de dialogue qui s'affiche alors (voir la figure 2.11)

2. On a supposé que les classes sont de même longueur

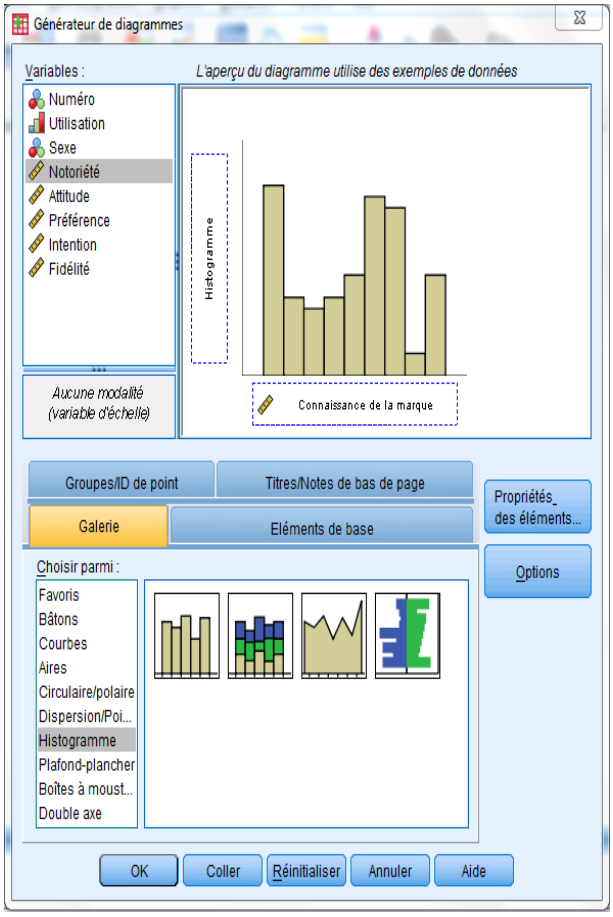


FIGURE 2.10 –

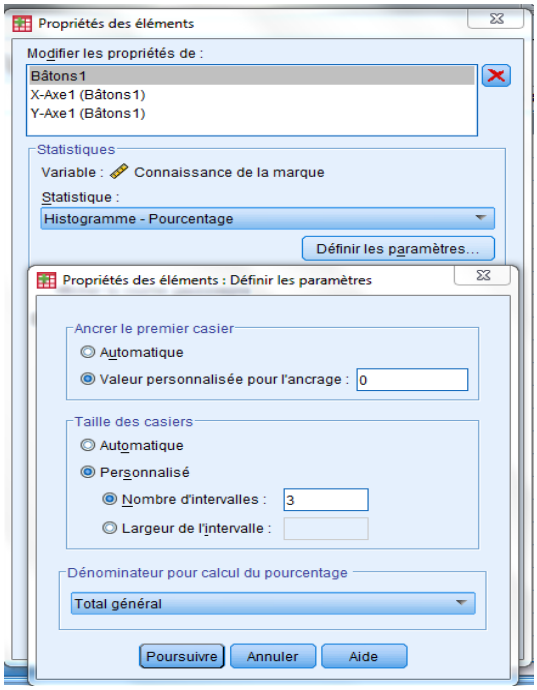


FIGURE 2.11 –

2.1.3 Diagrammes en tiges et feuilles

Les histogrammes et distributions de fréquences, méthodes couramment utilisées pour présenter les données, ne sont pas pour autant dépourvus d'inconvénients. Les histogrammes représentent des données groupées ; ils masquent donc les valeurs numériques réelles situées dans chaque intervalle. Quant aux distributions de fréquences, elles conservent les valeurs des différentes observations, mais peuvent s'avérer difficiles à utiliser lorsqu'elles ne résument pas suffisamment les donnés. Il existe une autre méthode alternative permettant d'éviter ces deux inconvénients : le diagramme en tiges et feuilles [3][9].

Supposons que le questionnaire destiné aux clients de la marque Nike comporte une question sur leur taille. Les valeurs de cette variable, arrondies au centimètre le plus proche, ont été saisies dans l'ordre du dépouillement de l'enquête. On dispose ainsi d'une série statistique d'effectif total 45 de nombres à trois chiffres :$\{170, 181, 169, ..., 152\}$ Les nombres contenus dans

170	181	169	180	169	155	190	167	174	194	162	165	175	174	157
180	183	177	190	188	197	195	156	183	182	160	165	181	179	151
173	169	190	171	161	156	160	161	164	177	184	198	195	170	152

TABLE 2.3 – Série observée des tailles

le tableau possèdent tous trois chiffres. Les deux premiers d'entre eux sont identiques pour plusieurs valeurs, par exemple 180 183 188 182 181. On peut différencier ces observations vis-à-vis de cette propriété commune (qui constitue **la tige**) par l'intermédiaire du troisième chiffre (que l'on dénomme la feuille). la deuxième observation 181 se situe sur la tige 18 et est présentée par la troisième feuille 1. La dernière observation 152 est attachée à la tige 15 et est représentée par la la feuille 2. On obtient ainsi le diagramme en tiges et feuilles de la figure

```
Taille en centimètres Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     2,00       15 .  12
     4,00       15 .  5667
     6,00       16 .  001124
     6,00       16 .  557999
     6,00       17 .  001344
     4,00       17 .  5779
     8,00       18 .  00112334
     1,00       18 .  8
     4,00       19 .  0004
     4,00       19 .  5578

 Stem width:        10
 Each leaf:       1 case(s)
```

FIGURE 2.12 – Diagramme en tiges et feuilles

3. Ce type de présentation était proposée par Tukey (1977)

Procédure sous SPSS

On va maintenant voir les étapes à suivre sous SPSS pour produire un diagramme en tige et feuilles. (voir figure2.13)

1. Suivre le chemin suivant : **Analyse** $\Longrightarrow$ **Statistiques descriptives** $\Longrightarrow$ **Explorer**.
2. Glisser la variable **Taille** dans la case **Liste des variables dépendante)**.
3. Cliquer sur **Diagrammes** et sélectionner **Tige et feuille** avant de cliquer sur **Poursuivre**

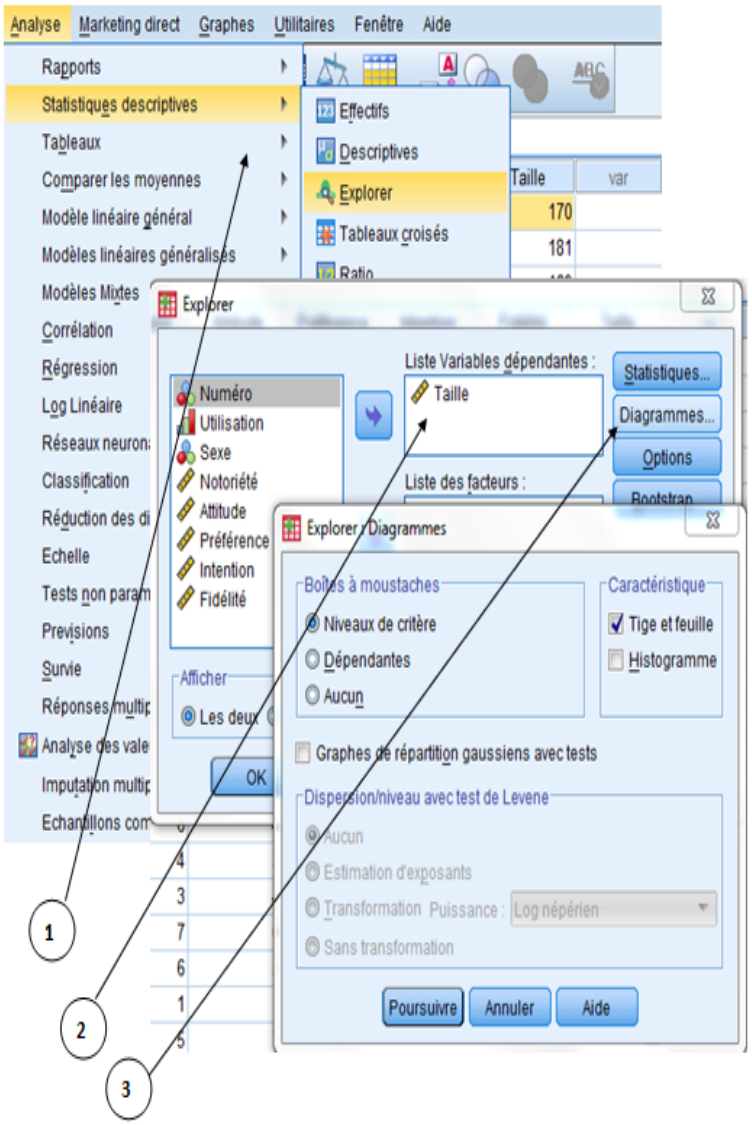


FIGURE 2.13 – Diagramme en tige et feuilles sous SPSS

2.2 Statistiques associées

La distribution de fréquences, les tableaux de fréquences fournissent des informations parfois trop détaillées, et le chargé d'étude doit alors les synthétiser au moyen de statistiques descriptives[12].
Soit X une variable statistique.

2.2.1 Mesures de position centrale

Cette expression fait référence à l'ensemble des mesures liées à l'endroit où la distribution est centrée sur l'échelle. Les trois principales mesures de tendance centrale sont le mode, qui se base sur quelques points de données seulement, la médiane, qui fait abstraction de la plupart des données, et la moyenne, qui se calcule à partir de toutes les données[9].

- **Mode :** M_o Définit la valeur présentant la plus grande fréquence d'occurrence. Il représente le pic de la distribution. Le mode offre une mesure de position centrale lorsque la variable étudiée est qualitative, ou qu'elle a été transformée en variable qualitative.
- **Médiane :** $M_{\acute{e}}$ La médiane d'un échantillon désigne la valeur centrale d'un ensemble de données classées par ordre ascendant ou descendant.
- **Moyenne :** $\bar{X}$ Constitue la mesure de la tendance centrale la plus fréquemment employée. Elle sert à évaluer la moyenne des données collectées à l'aide d'une échelle d'intervalles ou de rapport. En l'absence de valeurs extrêmes, la moyenne offre une mesure solide.

2.2.2 Mesures de dispersion

Calculées sur des données d'intervalles ou de rapport, ces mesures caractérisent la répartition des observations les unes par rapport aux autres, ou encore autour d'une valeur centrale. On pourrait utiliser différents mesures, qui seront évoquées l'une après l'autre, en commençant par les plus simples (pour plus de détailles voir[9] ou [6] pour ceux qui adorent les formules mathématiques)

- **L'étendue :** E Mesure la dispersion des données. Il correspond à la différence entre la plus forte et la plus faible valeur de l'échantillon.
- **Écart interquartile :** Q Représente la différence entre le 75^e et le 25^e centile (correspondant au 3^e et au 1^{er} quartile notés Q_3 et Q_1 respectivement).
- **Variance et écart-type :** S^2 **et** S On appelle variance (ou moment centré d'ordre 2) de X, la moyenne des carrés des écarts des valeurs de X à la moyenne :

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

  L'écart-type se définit comme la racine carrée positive de la variance.
- **Coefficient de variation :** CV Correspond au rapport de l'écart-type sur la moyenne, exprimé en pourcentage. Il sert à comparer des distributions de moyennes fort différentes.

$$CV(X) = \frac{S}{\bar{X}}$$

2.2.3 Mesures de forme

Permettent de mieux comprendre la nature de la distribution. Pour évaluer la forme d'une distribution, on examine son asymétrie et son aplatissement [4]. Nous ne considérerons ici que le cas de variables quantitatives ou ordinales pour lesquelles le problème a un sens.

4. les définitions des concepts ci-dessus prennent leur origine du référence[12]

– **Symétrie et asymétrie**

1. **Distribution asymétrique :** Exprime la tendance d'écarts à se montrer plus importants dans une direction que dans l'autre, comme si l'une des extrémités de la distribution possédait un poids plus grand.
2. **Distribution normale :** Utilisée pour le calcul de la taille de l'échantillon et base des principales analyses statistiques. De nombreux phénomènes continus suivent une loi normale ou s'en approchent. La loi normale peut être utilisée comme approximation de nombreuses distributions de probabilités discrètes. Elle est en forme de cloche symétrique. Ses mesures de tendance centrale (moyenne, médiane et mode) sont identiques. Sa variable aléatoire associée X varie de $-\infty$ à $+\infty$.
3. **Distribution symétrique :** Les valeurs sont les mêmes de part et d'autre du centre de la distribution. La moyenne, le mode et la médiane sont égaux. Les écarts positif et négatif par rapport à la moyenne sont eux aussi identiques.

La figure 2.14 met en évidence la différence entre distribution symétrique et asymétrique négative (ou à gauche) :

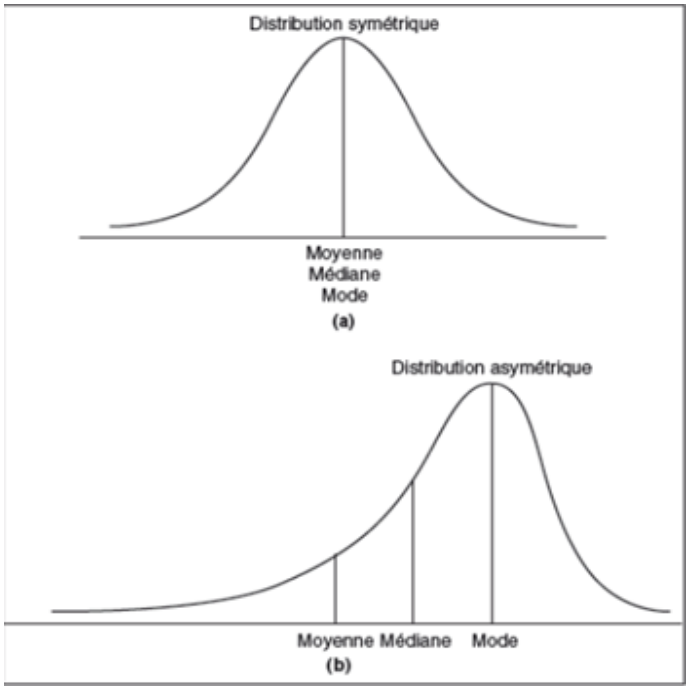


FIGURE 2.14 – Asymétrie d'une distribution,Source :[12]

– **Aplatissement :** Mesure la platitude ou le relief relatifs de la courbe dessinée par la distribution de fréquences. L'aplatissement d'une distribution normale est égal à 0. Un coefficient d'aplatissement (Kurtosis) positif caractérise une distribution plus concentrée que la distribution normale. À l'inverse, un aplatissement négatif indique une distribution plus plate que la distribution normale.

2.2.4 La boite à moustaches

Les représentations en tiges et feuilles permettent permettent de représenter les données de plusieurs facons intéressantes simultanément. Elles combinent les données en un shéma très proche de l'histogramme, tout en conservant les valeurs particulières des observations. outre la représentation en tiges et feuilles, John Tukey a élaboré d'autre moyens d'examiner les données ; L'un de ces moyens souligne davantage la dispertion des données : il s'agit de la méthode appelée

diagramme en forme de boite (boxplot) ou parfois **diagramme en forme de boite avec moustaches** [9].

Une boite à moustache nous indique de façon simple quelques trais marquants de la série. La médiane (notée aussi Q_2 : deuxième quartile) nous renseigne sur le milieu de la série. Les largeurs des deux parties de la boite nous rendent compte de la dispersion des valeurs situées au centre de la série (la boite contient 50% de l'ensemble des observations : 25% à gauche de la médiane et 25% à sa droite)[6]. Pour finir la définition présenté ci-dessus on introduit les deux concepts suivants :

a) **Les valeurs pivots** : Ces valeurs sont définies par les relations suivantes :

$$p_g = Q_1 - 1,5(Q_3 - Q_1) \text{ et } p_d = Q_1 + 1,5(Q_3 - Q_1)$$

Elles sont donc situées de part et d'autre de la boite et en sont distantes d'une fois et demie sa longueur.

b) **Les valeurs adjacentes** : Contrairement aux valeurs pivots, les valeurs adjacentes doivent être des valeurs observées de la série statistique. Elles correspondront aux extrémités des moustaches gauche et droite du diagramme en boite. On les définit comme suit. la valeur adjacente gauche-notée x_g- est la plus petite observation supérieure ou égale p_g, tandis que la valeur adjacente droite-désignée par x_d- est la plus grande observation inférieure ou égale à p_d.

c) **Les valeurs extérieures** : Toutes les observations situées en dehors de $[p_g, p_d]$ sont dites extérieures.

la figure 2.15 met en évidence les différents concepts relatifs à boxplot.

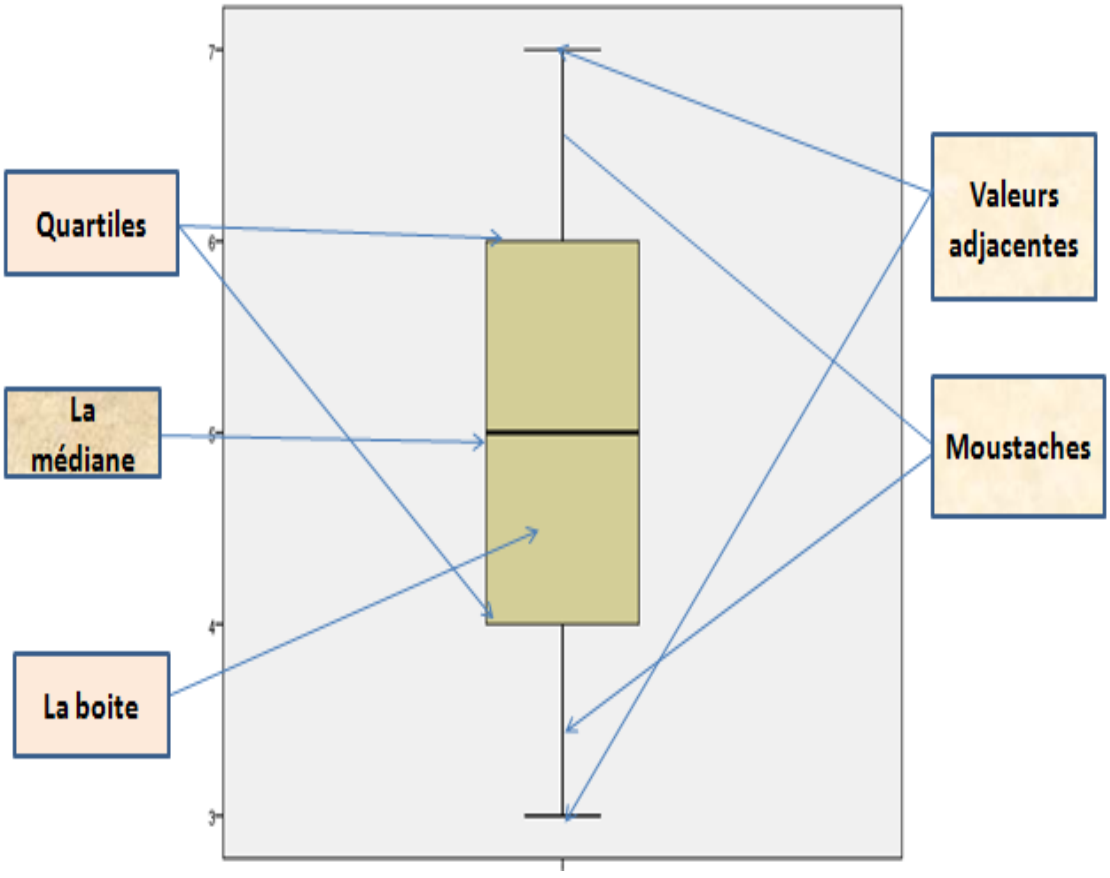


FIGURE 2.15 – La boite à moustaches

Descriptives

			Statistique	Erreur standard
Taille en centimètres	Moyenne		174,00	1,947
	Intervalle de confiance à 95% pour la moyenne	Borne inférieure	170,08	
		Borne supérieure	177,92	
	Moyenne tronquée à 5%		173,94	
	Médiane		174,00	
	Variance		170,545	
	Ecart-type		13,059	
	Minimum		151	
	Maximum		198	
	Intervalle		47	
	Intervalle interquartile		20	
	Asymétrie		,113	,354
	Aplatissement		-,931	,695

FIGURE 2.16 –

Traitement sous SPSS

Pour obtenir les différentes mesures indiquées ci-dessus, on préconise de suivre les les mêmes étapes indiquées pour la construction des Diagrammes **en tiges et feuilles** (voir2.1.3), sauf pour l'étape 3, on doit cliquer sur l'option **Statistiques** de la zone Afficher pour choisir les statistiques descriptives qui nous intéressent. La figure 2.16 met en évidence les statistiques descriptives de la variable **Notoriété** calculée avec SPSS.

2.3 Travaux pratiques

TP2

Vous avez pour mission de déterminer si le remplacement d'une lumière blanche par une lumière rouge dans un bureau aura une influence sur la qualité du travail de personnes qui doivent rentrer des données dans un ordinateur. On pense que la lumière rouge augmentera peut-être la vigilance des employés qui feront ainsi moins d'erreurs. En remplaçant la lumière blanche par une lumière rouge, vous trouvez sur un échantillon que le nombre d'erreurs par jour a baissé des quantités suivantes :

$$22, 22, 12, 10, 42, 19, 20, 19, 20, 21, 21, 20, 30, 28, 26, 18, 18, 20, 21, 19$$

1. Quelle est la variable indépendante ?
2. Quelle est la variable dépendante ?
3. Utiliser SPSS pour dessiner une boite à moustaches Pour les scores observés.
   - Les données sont-elles distribuées normalement ?
   - Y a-t-il des valeurs extrêmes ?
   - Utiliser SPSS pour trouver la moyenne et l'écart-type de la série de scores.

****

****

Le service qualité d'un opérateur téléphonique a interroger 2500 clients et leur a demandé d'exprimer, sur une échelle de 1 à 7, leur degré de satisfaction quant au service délivré (1 pour pas du tous satisfait, 7 pour très satisfait). Les résultats obtenus figurent dans le tableau ci-dessous

Valeurs	1	2	3	4	5	6	7
Effectifs	180	520	512	388	315	385	200

1. Représenter le diagramme en bâton de la série.
2. Déterminer le mode, la moyenne arithmétique, la médiane et les premier et troisième quartiles de la série.
3. Évaluer l'écart-type, l'étendu et l'intervalle inter-quartile de la série

****

Questionnaire Automobile

1. Possédez-vous une automobile?

○ 1. Non ○ 2. Oui une ○ 3. Oui2 ○ 4. Oui+de2

2. Pouvez-vous me citer les différentes marques automobiles que vous avez déjà possédées ?

☐ 1. Renaut ☐ 2. Peugeot ☐ 3. Citroen ☐ 4. Tablot ☐ 5. Ford ☐ 6. Mercéd ☐ 7. Wolksw ☐ 8. BMW
☐ 9. Opel ☐ 10. Fiat ☐ 11. Volvo ☐ 12. Toyota ☐ 13. Honda ☐ 14. Autre

(si plus de cinq, citer les plus récentes)

3. Parmi les qualités suivantes, ctez les 3 plus importantes à vos yeux.

1. Puiss 2. Vitesse 3. Confort 4. Sécurité 5. Esthétique 6. Consom

☐ ☐ ☐

Classez-les par ordre décroissant

4. Quel kilométrage effectuez-vous mensuellement? ☐

5. Pour vous, l'automobile est un équipement:

○ 1. Très utile ○ 2. Utile ○ 3. Peu utile ○ 4. Inutile ○ 5. Nuisible

1. Traduire le questionnaire ci-dessus dans le format SPSS.
2. Faire la saisie des données (fictifs) pour 10 observation.
3. Produire les distributions de fréquences pour chaque variable figurant dans le questionnaire.

****

Chapitre 3

Les tests univariés

L'analyse de base des données implique obligatoirement les tests d'hypothèses. Les exemples d'hypothèses générées par les études marketing sont légion [12] :

- Un grand magasin est fréquenté par plus de 10% des foyers.
- Les clients assidus et occasionnels d'une marque se différencient par leurs caractéristiques psycho-graphiques.
- Un hôtel possède une image plus haut de gamme que son concurrent direct.
- Le fait de connaitre un restaurant se traduit par une préférence plus marquée en sa faveur.

La validation des hypothèses s'appuie sur un ensemble de techniques, qui toutes visent à établir la signification statistique d'un résultat, afin d'en généraliser la portée à l'ensemble de la population dont est extrait l'échantillon.

Deux catégories de tests sont traitées dans ce chapitre

- Les tests paramétriques portent sur la valeur d'un paramètre de la variable statistique parente.

  Les tests paramétriques sont de deux types.

  1. Les tests de conformité à une norme portent sur un seul paramètre. Ils sont construits sur la base d'une valeur connue que doit ou devrait avoir ce paramètre. On teste, par exemple, si la durée moyenne des grossesses est égale à 41 semaines.
  2. Les tests de comparaison (ou d'homogénéité) tranchent entre deux hypothèses concernant le même paramètre mesuré dans deux populations. Ils répondent ainsi à des questions du type : la dispersion des salaires des femmes est-elle supérieure à celle des hommes ?

- Les tests sont dits non paramétriques lorsqu'ils ne portent pas sur la valeur d'un paramètre. Le test non paramétriques le plus courant est le test du Khi-deux : le test du Khi-deux appliqué à la détermination de l'indépendance, ou non, de deux variables statistiques et le test du Khi-deux d'ajustement d'une distribution observée par une distribution théorique.

3.1 Procédure générale des tests d'hypothèses

Un test d'hypothèses implique les étapes suivantes (voir figure 3.1)

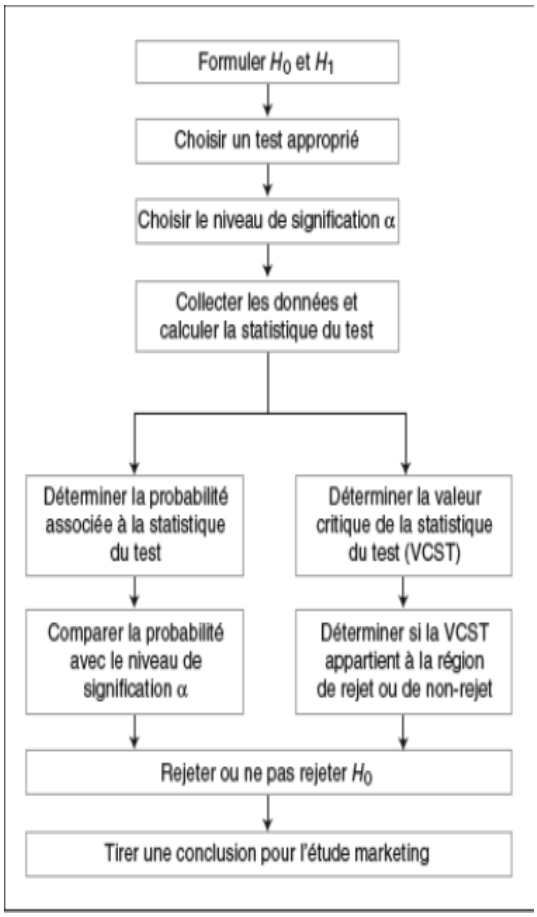


FIGURE 3.1 – Procédures générales d'un test d'hypothèses. Source :[12]

Étape 1 : Formuler les hypothèses L'objectif d'un test statistique est toujours de trancher entre deux hypothèses antagonistes. La première est appelée hypothèse nulle, et notée H_0. La seconde, appelée hypothèse alternative, est notée H_1. Les deux hypothèses sont asymétriques. L'hypothèse H_0 est l'hypothèse de statu-quo ou de stabilité. En revanche l'hypothèse alternative H_1, est l'hypothèse de recherche ou d'évolution.

En études marketing, l'hypothèse nulle se trouve formulé de telle sorte que son rejet aboutisse à l'adoption de la conclusion souhaité. L'hypothèse alternative présente la conclusion que l'on cherche à motiver. On peut imaginer par exemple qu'un magasin soit en train d'étudier la mise en place d'un service d'achat par Internet et ne se décide à le lancer qu'à condition que plus de 40% des internautes effectuent leurs achats par ce biais-là.

Repère 3 *:***(Hypothèses simples et composites)** [a] *Les hypothèses d'un test statistique sont simples ou composites. Les définitions de ces concepts diffèrent selon que les tests sont, ou non, paramétriques.*

Tests paramétriques

Pour les tests paramétriques de conformité, le paramètre testé est noté θ et la norme qui sert de référence est notée θ_0. Pour les tests paramétriques de comparaison, les valeurs du paramètre testé sont notées θ_A et θ_B.

◇ *Une hypothèse simple s'exprime sous la forme d'une égalité soit :*
- *Test de conformité : $\theta = \theta_0$*
- *Test de comparaison $\theta_A = \theta_B$*

◇ *Une hypothèse composite se traduit par une ou plusieurs inégalités*
- *Test de conformité : $\theta > \theta_0$ ou $\theta < \theta_0$ ou $\theta \neq \theta_0$.*
- *Test de comparaison $\theta_A > \theta_B$ $\theta_A < \theta_B$ $\theta_A \neq \theta_B$*

◇ *L'hypothèse nulle d'un test paramétrique peut toujours être ramenée à une hypothèse simple :*

$$H_0 : \theta = \theta_0 \text{ ou bien } H_0 : \theta_A = \theta_B$$

En revanche, l'hypothèse alternative d'un test peu être simple ou prendre les trois formes composites ci-dessus. Il y a ainsi :
- *quatre manières de formuler un test de conformité*

$$1)\begin{cases} H_0, & \theta = \theta_0 \\ H_1 & \theta = \theta_1 \end{cases} \text{ ou } 2)\begin{cases} H_0, & \theta = \theta_0 \\ H_1 & \theta > \theta_0 \end{cases} \text{ ou } 3)\begin{cases} H_0, & \theta = \theta_0 \\ H_1 & \theta < \theta_0 \end{cases} \text{ ou } 4)\begin{cases} H_0, & \theta = \theta_0 \\ H_1 & \theta \neq \theta_0 \end{cases}$$

- *trois manières de formuler un test de comparaison (le dernier est bilatéral)*

$$1)\begin{cases} H_0, & \theta_A = \theta_B \\ H_1 & \theta_A > \theta_B \end{cases} \text{ ou } 2)\begin{cases} H_0, & \theta_A = \theta_B \\ H_1 & \theta_A < \theta_B \end{cases} \text{ ou } 3)\begin{cases} H_0, & \theta_A = \theta_B \\ H_1 & \theta_A \neq \theta_B \end{cases}$$

La difficulté principale d'un test réside dans le choix de sa formulation. Afin d'éclairer le raisonnement à tenir, des exemples concrets sont traités dans les sections suivantes.

Tests non paramétriques

Ces tests portent généralement sur les lois de probabilité de variables aléatoires. Par exemple, on teste, à partir de la distribution observée d'une variable statistique, si l'histogramme peut être ajusté par la densité de probabilité d'une loi de Gauss. Ce test ne concerne donc pas un paramètre, mais une loi de probabilité. Pour ces tests, une hypothèse est simple lorsqu'elle spécifie complètement la loi de probabilité, elle est composite dans le cas contraire. Ainsi, l'hypothèse $H_0 : X \sim (10; 0,25)$ est simple alors que $H_0 : X \sim (10; \pi)$ est composite.
Enfin, pour le cas particulier d'un test d'indépendance, le test se formule toujours sous la forme :

$$\begin{cases} H_0 & X \text{ et } Y \text{ sont indépendantes} \\ H_1 & X \text{ et } Y \text{ ne sont pas indépendantes} \end{cases}$$

Dans ces tests d'indépendance, les expressions " hypothèse simple " ou " hypothèse composite " ne sont pas utilisées puisque la formulation du test est toujours la même.

a. Pour plus de détaille vous pouvez consulter la référence [17]

Étape 2 : Choisir un test approprié (Variable de décision) La statistique du test mesure la proximité de l'échantillon vis-à-vis de l'hypothèse nulle. Elle s'aligne généralement sur une distribution classique : Normale, Student, ou encore khi-deux (Voir l'annexe). Pour notre exemple (**Test de conformité d'une proportion à une norme**), la va-

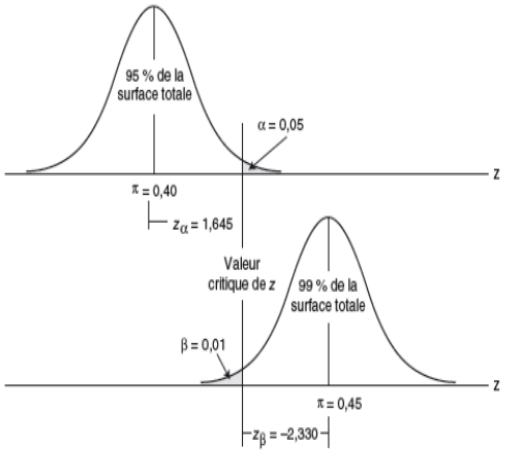


FIGURE 3.2 – Erreur de type I et de type II

riable de décision est :

$$T_n = \frac{F_n - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

F_n est la variable aléatoire indiquant la fréquence d'individus faisant leurs achats par internet et $\pi_0 = 0,4$.

Repère 4 (Test de conformité sur la proportion), *Source :[17]*

		Distribution d'échantillonnage de T_n	
Paramètres Connus/inconnus σ^2	**Variable de décision sous $H_0 : \pi = \pi_0$**	**Gaussien taille quelconque**	**Non gaussien grande taille $n > 30$**
Variance connue sous $H_0 : \pi = \pi_0$	$\downarrow$ $T_n = \frac{F_n - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$ *Régions critiques* $\downarrow$		$T_n \to \mathcal{N}(0;1)$
$\begin{cases} H_0 : \pi = \pi_0 \\ H_1 : \pi = \pi_1 \\ \pi_1 > \pi_0 \end{cases}$	$\to \to$	$\mathcal{C}_r = [z_{(1-\alpha)};\ +\infty[$	
$\begin{cases} H_0 : \pi = \pi_0 \\ H_1 : \pi > \pi_0 \end{cases}$	$\to \to$	$\mathcal{C}_r = [z_{(1-\alpha)}\ ^a;\ +\infty[$	
$\begin{cases} H_0 : \pi = \pi_0 \\ H_1 : \pi < \pi_0 \end{cases}$	$\to \to$	$\mathcal{C}_r =]-\infty;\ z_\alpha]$	
$\begin{cases} H_0 : \pi = \pi_0 \\ H_1 : \pi \neq \pi_0 \end{cases}$	$\to \to$	$\mathcal{C}_r =]-\infty;\ z_{1-\alpha/2}] \cup [z_{1-\alpha/2};\ +\infty[$	

a. $z_{(1-\alpha)}$ est le quantile d'ordre $(1-\alpha)$ de la loi normale standard

Étape 3 : Choisir le niveau de signification α . Dès que l'on cherche à dégager des inférences par rapport à une population, on prend le risque d'aboutir à une conclusion erronée. Deux types d'erreurs peuvent survenir. : Erreur de type I Erreur de type II. Les probabilités de l'erreur de type I(α) et de l'erreur de type II(β) sont représentées à la figure 3.2

Repère 5 *:***(Erreurs et risque, associés à un test)** [a]

A l'issue du test, on aboutit à l'une des deux décisions suivantes :

– *non rejet de* H_0

– *rejet de* H_0

La prise de décision est liée au caractère aléatoire des échantillons susceptibles d'être retenus. Les erreurs possibles, inhérentes it tous les tests, sont définies ci-dessous.

Définition 1

1. *L'erreur de première espèce est la décision de rejeter* H_0*. alors que* H_0 *est vraie. La probabilité de commettre cette erreur, notée* α *s'appelle le risque de première espèce ou le seuil du test . Notation :* $\alpha = P(H_1/H_0)$
2. *L'erreur de seconde espèce est la décision d'accepter* H_0 *alors que* H_1 *est vraie. La probabilité de commettre cette erreur, notée* β*, s'appelle le risque de deuxième espèce. Notation* $P(H_0/H_1)$*.*
3. *La puissance d'un test, notée* $\gamma = 1 - \beta$*, est la probabilité de rejeter* H_0 *quand cette hypothèse est fausse.*

Les notations $\alpha = P(H_1/H_0)$ *et* $\beta = P(H_0/H_1)$ *méritent une attention particulière. Ces deux probabilités ne sont pas des probabilités conditionnelles.*

		Réalité inconnue	
		H_0 *est vraie*	H_1 *est vraie*
Décisions possibles	*Décider du non rejet de* H_0	$1-\alpha$	β
	Décider du rejet de H_0	α	$\gamma = 1-\beta$

a. Voir [17]. L'ouvrage [13] est une excellence référence pour ceux qui sont intéressé par la logique des tests

Étape 4 : Collecter les données et calculer la statistique du test

1	2	1	1	1	1	2	2	1	1	2	2	2	2	1
2	1	1	1	2	2	2	2	1	1	1	1	2	1	1

TABLE 3.1 – Données : Achat sur Internet, 1=Oui et 2=Non

La variable F_n prend comme réalisation la valeur f_n =, par suite T_n sera réalisé par $t = \dfrac{f_n - \pi_0}{\sqrt{\dfrac{\pi_0(1-\pi_0)}{n}}} = 1,863$

Étape 5 : déterminer la probabilité (valeur critique) La p-value (ou seuil critique de signification en anglais)est la probabilité de sélectionner un échantillon dans lequel la valeur de t soit au moins en désaccord avec H_0 que celle observé dans notre échantillon, ceci si H_0 est vraie. si cette valeur est inférieure à α on décide H_1(voir figure3.3) :

Étape 6 et 7 : Comparer la probabilité (valeur critique) et prendre une décision. Au seuil de signification $\alpha = 0,050$ on peut rejeter l'hypothèse nulle d'égalité des proportions. Autrement dit, l'hypothèse alternative selon laquelle $\pi > 0,4$ est significative.

Étape 8 : Conclusion pour l'étude marketing On peut conclure que la proportion d'internautes effectuant leurs achats en ligne s'avère de toute évidence très supérieur à 0,40. On conseillera donc au magasin d'ouvrir son nouveau service d'achats Internet.

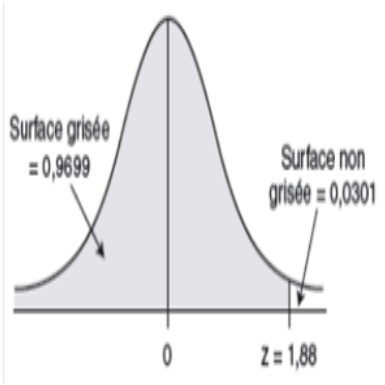


FIGURE 3.3 – Probabilité associée à z pour un test unilatéral

Fréquence : 17 / Fréquence totale : 30
Proportion test : 0,4

Test Z pour 1 proportion / Test unilatéral à droite :

Remarque : la loi binomiale est approximée par la loi normale

Z (valeur observée)	1,863
Z (valeur critique)	1,645
p-value unilatérale	0,031
Alpha	0,05

FIGURE 3.4 – Comparaison d'une proportion avec XLSTAT 7.5.2

3.2 tests univariés :Principes et applications

La figure 3.5 englobe un panorama des tests univariés classés selon trois dimensions :

1. L'échelle de mesure de la variable étudiée(Nominale, Ordinale, Intervalle ou Ratio)(voir Repère1, chapitre1).
2. Le nombre d'échantillons (1, 2 ou plus de deux échantillons).
3. Les échantillons (indépendants, Appariés).

Tests Univariés

Echelle de mesure	Un échantillon	deux échantillons		k échantillons	
		Indépendants	Dépendants	Indépendants	Dépendants
Nominal	Test Binomial (Z-test de proportion) x^2	x^2	McNemar	x^2	Cochran Q
Ordinal	Kolmogorov-Smirnov	Mann-Whitney	Wilcoxon	Kruskal-Walis	Friedman
Intervalle ou Ratio	Test-t Test-Z	Test-t Test-Z	Test-t de différences	Analyse de Variance (ANOVA)	Analyse de Variance à mesures répétés

FIGURE 3.5 – Panorama (non exhaustif) des tests univariés

Notre démarche est basée sur la stratégie suivante :

· **Principe :** justifiant l'utilité et si c'est nécessaire les procédures décrivant les étapes de ce test.

· **Exemple d'application :** met en évidence l'application concrète du test dans le domaine du marketing.

· **Procédure sous SPSS :** nous facilite la réalisation du test sous SPSS.

Pour le dernier point de cette stratégie, on va utiliser les données d'une étude consacrée à l'utilisation personnelle (non professionnelle) d'Internet.

Exemple 4 *Le tableau ci-dessous contient les données relatives à 30 répondants, lesquelles permettent de connaitre*

- *le sexe (1=masculin, 2 féminin)*
- *le niveau de connaissance d'Internet (1=très faible connaissance, 2=très bonne connaissance)*
- *le nombre d'heures d'utilisation par semaine*
- *l'attitude vis-à-vis d'Internet et de la technologie, mesurée sur une échelle à sept degrés (1=très défavorable, 7= très favorable)*
- *Ces données permettent aussi de savoir si les répondants ont déjà effectué des achats ou des opérations bancaires sur Internet (1=Oui, 2=Non).*

N°	**Sexe**	**Connaissance**	**Utilisation**	**Attitude Internet**	**Attitude Technologie**	**Achats Internet**	**Opérations Bancaires Internet**
1	*1*	*7*	*14*	*7*	*6*	*1*	*1*
2	*2*	*2*	*2*	*3*	*3*	*2*	*2*
3	*2*	*3*	*3*	*4*	*3*	*1*	*2*
4	*2*	*3*	*3*	*7*	*5*	*1*	*2*
5	*1*	*7*	*13*	*7*	*7*	*1*	*1*
6	*2*	*4*	*6*	*5*	*4*	*1*	*2*
7	*2*	*2*	*2*	*4*	*5*	*2*	*2*
8	*2*	*3*	*6*	*5*	*4*	*2*	*2*
9	*2*	*3*	*6*	*6*	*4*	*1*	*2*
10	*1*	*9*	*15*	*7*	*6*	*1*	*2*
11	*2*	*4*	*3*	*4*	*3*	*2*	*2*
12	*2*	*5*	*4*	*6*	*4*	*2*	*2*
13	*1*	*6*	*9*	*6*	*5*	*2*	*1*
14	*1*	*6*	*8*	*3*	*2*	*2*	*2*
15	*1*	*6*	*5*	*5*	*4*	*1*	*2*
16	*2*	*4*	*3*	*4*	*3*	*2*	*2*
17	*1*	*6*	*9*	*5*	*3*	*1*	*1*
18	*1*	*4*	*4*	*5*	*4*	*1*	*2*
19	*1*	*7*	*14*	*6*	*6*	*1*	*1*
20	*2*	*6*	*6*	*6*	*4*	*2*	*2*
21	*1*	*6*	*9*	*4*	*2*	*2*	*2*
22	*1*	*5*	*5*	*5*	*4*	*2*	*1*

Données sur l'utilisation d'Internet

N^o	**Sexe**	**Connaissance**	**Utilisation**	**Attitude Internet**	**Attitude Technologie**	**Achats Internet**	**Opérations Bancaires Internet**
23	*2*	*3*	*2*	*4*	*2*	*2*	*2*
24	*1*	*7*	*15*	*6*	*6*	*1*	*1*
25	*2*	*6*	*6*	*5*	*3*	*1*	*2*
26	*1*	*6*	*13*	*6*	*6*	*1*	*1*
27	*2*	*5*	*4*	*5*	*5*	*1*	*1*
28	*2*	*4*	*2*	*3*	*2*	*2*	*2*
29	*1*	*4*	*4*	*5*	*3*	*1*	*2*
30	*1*	*3*	*3*	*7*	*5*	*1*	*2*

Données sur l'utilisation d'Internet (Suite), Source :[12] page 366

Essayer de créer un fichier SPSS pour ces données, nommer le **Utilisation Internet**

3.2.1 Cas un seul échantillon

Variable Nominale

Test Binomial (Test-Z de proportion)

↪ **Principe :** Le test binomial est un test extrêmement simple. La distribution binomiale est la distribution d'échantillonnage des proportions que l'on peut observer dans des échantillons aléatoires tirés d'une population composée de deux catégories. Cette distribution permet de tester l'hypothèse nulle selon laquelle les proportions dans l'échantillon ne sont pas différentes de celles de la population dont il est tiré[7].

↪ **Exemple d'application :** Prenons l'exemple de produits sans marque, L'étude consiste à examiner la connaissance des produits sans marque par une population bien spécifique. L'hypothèse nulle sera que, parmi les personnes ayant répondu à l'enquête, la proportion de ceux qui connaissent des produits sans marque est de 50%. Le test binomial va consister à calculer la probabilité d'obtenir les valeurs observées dans l'échantillon ainsi que des valeurs plus extrêmes.On pourra ensuite, pour un seuil de confiance donné, accepter ou rejeter l'hypothèse nulle. Supposons que l'expérience menée sur les produits sans marque le soit sur un tout petit échantillon : 16 individus, dont seulement deux individus ne connaissent pas les produits sans marque. La probabilité d'obtenir la valeur observée de 2, ainsi que des valeurs plus extrêmes, à savoir 1 ou 0, est égale $0,002$ (pour un test unidirectionnel ou unilatéral). La valeur obtenue est inférieure au seuil classique $0,05$ et même au seuil $0,01$. On peut donc rejeter l'hypothèse nulle à $p \leq 0,01$.

Remarque 6 *Pour le test binomial, si la taille de l'échantillon dépasse la valeur* 30*, la loi binomial est approximée par une loi normale centrée réduite, d'où la nomination "Test-Z de proportion".*

↪ **Procédure sous SPSS :** (Voir figure 3.6)

1. Cliquer sur **Analyse**, **tests non paramétriques** et **test binomial**
2. Glisser la variable **Achats Internet** dans **Liste des variables à tester**
3. Cliquer sur l'option **Selon les données** dans le groupe **Définir la dichotomie**
4. Mettre la valeur $0,4$ dans la case **Proportion testée**
5. Cliquer sur **Options** pour définir quelques statistiques descriptives ou pour choisir la méthode de traitement des valeurs manquantes.

Les résultats de cette analyse se trouve dans la figure 3.7.

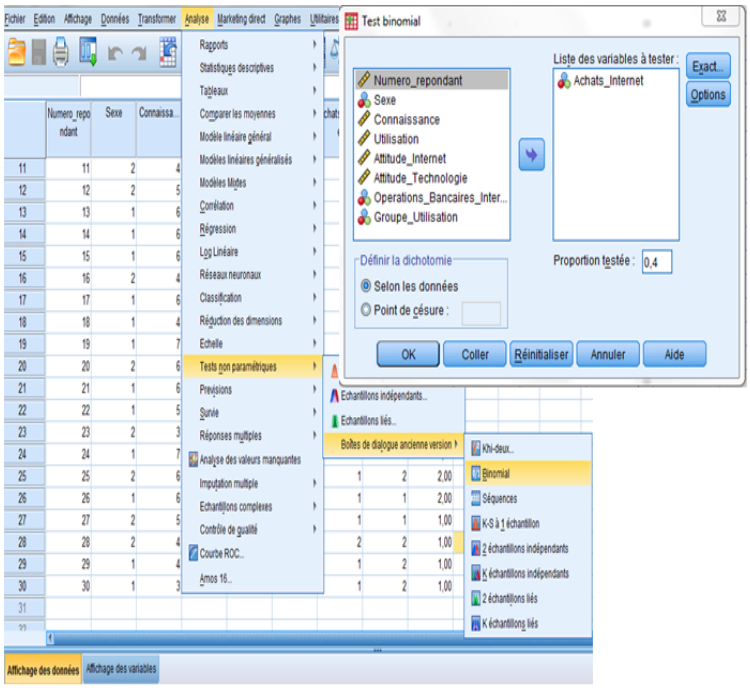


FIGURE 3.6 – Procédures du test binomial sous SPSS

Test binomial

		Modalité	N	Proportion observée.	Test de proportion	Signification asymptotique (unilatérale)
Achats sur Internet	Groupe 1	Yes	17	,6	,4	,048[a]
	Groupe 2	No	13	,4		
	Total		30	1,0		

a. Basée sur l'approximation de Z.

FIGURE 3.7 –

Les test de χ^2

[1] Les tests de χ^2 peuvent être appliqués sur tous types de variables : qualitative nominale, ordinale, qualitative binaire, quantitative discrète ou continue discrétisé. Selon la situation on distingue :

- test de χ^2 de **conformité (ou d'ajustement)**. Il sert à comparer une distribution observée sur un échantillon à une distribution une distribution connue dans une population ou à une distribution théorique : binomiale, Poisson, normale, etc.
- le test de χ^2d'**homogénéité**. Il sert à comparer deux ou plusieurs distribution observées sur des échantillons.
- le test de χ^2 d'**indépendance**. Il sert à étudier sur un même échantillon la liaison entre les distributions de deux variables (nominales ou binaire). Quelle que soit la situation, le principe et le calcul du test sont identiques.

Test de χ^2 d'ajustement

↪ **Principe :** On cherche à savoir si les fréquences (ou proportions) observées dans les différentes classes de réponses sont significativement différentes de celles estimées ou attendues dans la population. En d'autres termes, on cherche à vérifier s'il existe une différence entre

des proportions observées et des proportion théoriques (ou attendues). La statistique du Chi-deux repose sur le calcul des écarts entre une distribution théorique, afin de tester la probabilité qu'ils se produisent sous l'hypothèse nulle H_0 qui postule l'égalité des distributions. Le rejet de H_0 qui s'observe pour des valeurs fortes du χ^2 permet de conclure que la distribution observée est différente de la distribution théorique avec un seuil de confiance élevé. La valeur du χ^2 tabulé est celle donnée par la table de distribution du Chi-deux pour $(k-1)$ degrés de liberté (k étant le nombre de classes ou de catégories) et la formule de calcul du χ^2 est égale à :

$$\chi^2_{calc} = \sum_{i=1}^{i=k} \frac{(N_i - T_i)^2}{T_i}$$

N_i : fréquence observée dans la catégorie i.

T_i : fréquence théorique (attendue) dans la catégorie i.

k : nombre de catégories.

↪ **Exemple d'application :** Pour étudier l'éventuelle influence de la couleur de l'emballage d'un nouveau savon, **SAVOLIVE** S.A. a procédé à une étude sur un échantillon de 200 ménages. On a remis à chacun d'entre eux quatre savons de même composition mais emballés dans des boites de couleurs différentes (rouge, blanc, bleu, vert), en leur affirmant qu'il s'agit de savons de formules différentes (ce qui est faux). On leur a indiqué en même temps qu'il leur sera demandé, un mois plus tard, vers lequel va leur préférence, **SAVOLIVE** leur offrant alors une caisse de 24 savons de leur choix. A la fin de l'expérience, les choix observés ont été les suivant :

Couleur(X)	**Rouge**	**Blanc**	**Bleu**	**Vert**	**Ensemble**
Effectif	51	74	30	45	200

Le problème posé est de savoir si la couleur de l'emballage a de l'importance. les quatre étapes suivantes suffiront pour pour répondre à notre problème.

1. **Les hypothèses**

$$\begin{cases} H_0 : X \sim \text{la loi Uniforme (le choix se fait au hasard)} \\ H_1 : X \text{ ne suit pas la loi Uniforme (la couleur a une influence)} \end{cases}$$

2. **Variable de décision**

$$T = \sum_{i=1}^{4} \frac{(N_i - np_i)^2}{np_i} \sim \chi^2_{4-1}.$$

3. **Région critique**

$$\delta_\alpha = [7,81;\ +\infty[\quad \text{avec } \alpha = 5\%$$

4. **Décision**
   Z prend la valeur

$$z = \sum_{i=1}^{4} \frac{(n_i - np_i)^2}{np_i} = \frac{(51-50)^2}{50} + \frac{(74-50)^2}{50} + \frac{(30-50)^2}{50} + \frac{(45-50)^2}{50} = 20,04$$

$20,04 \in \delta_\alpha$ par suite on décide H_1

Conclusion :la couleur de l'emballage a une influence sur les choix des individus

↪ **Procédure sous SPSS :**

Si les données sont disponibles sous forme d'un tableau statistique (effectifs) comme dans notre exemple d'application, on peut les entrer dans SPSS directement de cette manière.
Les données se présentent alors ainsi : figure3.8

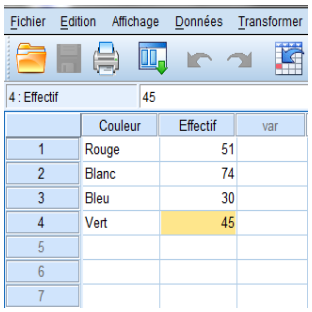


FIGURE 3.8 –

Pour la réalisation du test sous SPSS il suffit de suivre les étapes suivantes : (figure 3.9)

1. Pour informer SPSS qu'on se trouve dans le cas d'un tableau statistique : on sélectionne **Données** ⇒ **Pondérer les observations**
2. Déplacer effectif dans la zone "**variable de pondération**"
3. Cliquer ensuite sur **OK** pour revenir à l'écran principal.
4. Choisir **Analyse**, **Tests non paramétriques** puis **khi-deux :**
5. Cliquer si vous le désiriez sur **Options**, et finissez par **OK**

Les résultats apparaissent dans la fenêtre des résultats (figure 3.10) Le χ^2 de confirmité (χ^2, $ddl = 3$) correspond à significativité $p = 0.000$ inférieure à 5%. On est donc fondé à rejeter l'hypothèse nulle d'équivalence des quatre couleurs ; les différences de fréquences étant très significatives. Le mode observé est "**Blanc**". La couleur Blanche est donc globalement préférée aux trois autres.

Variable Ordinale

Test de Kolmogorov-Smirnov

↪ **Principe :** Il s'agit d'un test d'ajustement au même titre que celui du Chi-deux. Ce test consiste toujours à comparer une distribution observée à une distribution théorique (ou attendue). L'hypothèse nulle H_0 postule que la distribution des réponses dans l'échantillon n'est pas significativement différente d'une distribution théorique (qui peu être celle de la population). Pour procéder au test, il est nécessaire de calculer la distribution des proportions de fréquences cumulées observées et de la comparer avec celle des mêmes valeurs calculées à partir de la distribution théorique. Kolmogorov et Smirnov ont montré que, si l'hypothèse nulle s'applique, les écarts entre les cumuls de proportions observées et théoriques sont minimes. Ils ont montré que l'écart maximum (D) suit une loi de distribution connue. Pour un échantillon qui comporte plus de 35 observations, les valeur de rejet de l'hypothèse nulle sont données ci-dessous pour différents seuils de signification. Lorsqu'on dispose de moins de 35 observations, il convient de se référer à la table de la distribution de D([11]) (voir table en l'annexe)

Formule de calcul $D = Max$ | Prop.cumulée observée-Prop.cumulée théorique |

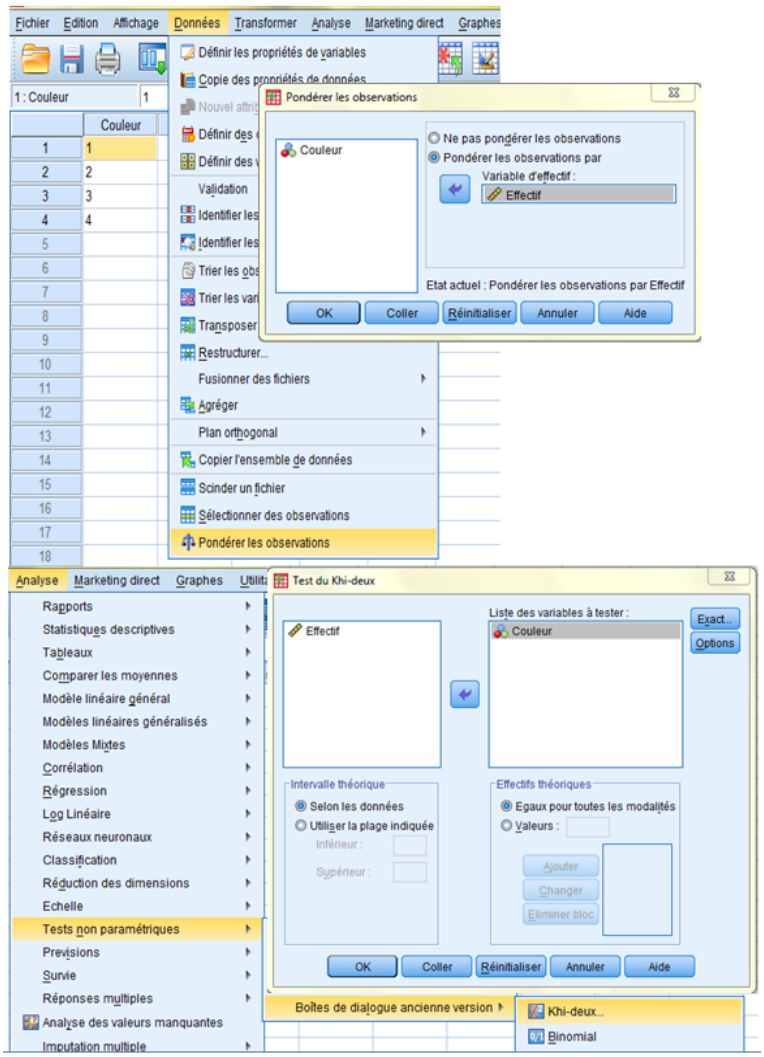


FIGURE 3.9 –

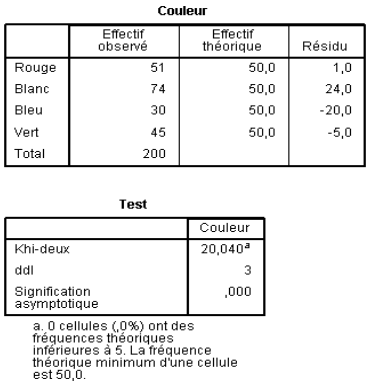
Seuil de signicativité : $\alpha = 10\%$ $\alpha = 5\%$ $\alpha = 1\%$

Valeur de rejet de H_0 $D \geq \dfrac{1,22}{\sqrt{n}}$ $D \geq \dfrac{1,36}{\sqrt{n}}$ $D \geq \dfrac{1,63}{\sqrt{n}}$

↪ **Exemple d'application :** ([7]) Supposons par exemple qu'un échantillon de 100 consommatrices ait noté la tonalité d'un nouveau cosmétique selon une échelle à 4 catégories allant de plus foncé à moins foncé : la répartition des réponses obtenues est donnée dans le tableau 3.2 ci-dessous. L'objectif du test est de comparer la distribution des réponses à celle qui serait obtenue selon une hypothèse nulle définie a priori ; on peut par exemple prendre comme hypothèse nulle l'indifférence des jugements se traduise par une équirépartition ("théorique") des réponses.

Le test consiste à calculer, dans le cas d'un test bilatéral, pour chaque modalité de la variable, la valeur absolue de la différence entre les pourcentages cumulés observés et théoriques ; la valeur testé (notée D) est la plus grande de ces différences, soit, dans l'exemple présenté, $0,30$. Cette quantité est comparée à une valeur dans une table spéciale si la taille de l'échantillon (n) est inférieure à 35 ; si $n > 35$; on applique les valeurs-seuil indiquées ci-dessus.

Dans notre exemple ($n = 100$), la valeur-seuil de 1% est égale à $0,163$. Elle est inférieure à la valeur observé $D_{obs} = 0,30$; l'hypothèse nulle (indifférence des jugements) est donc



Couleur

	Effectif observé	Effectif théorique	Résidu
Rouge	51	50,0	1,0
Blanc	74	50,0	24,0
Bleu	30	50,0	-20,0
Vert	45	50,0	-5,0
Total	200		

Test

	Couleur
Khi-deux	20,040[a]
ddl	3
Signification asymptotique	,000

a. 0 cellules (,0%) ont des fréquences théoriques inférieures à 5. La fréquence théorique minimum d'une cellule est 50,0.

FIGURE 3.10 –

Catégorie de réponse	Distribution observée			Distribution théorique		Différence des cumulés (D) Obs-Théo (valeur absolue)
	Nbre	Prop	Proportion cumulée	Prop	Proportion cumulée	
1	50	0,50	0,50	0,25	0,25	0,25
2	30	0,30	0,80	0,25	0,50	0,30
3	15	0,15	0,95	0,25	0,75	0,20
4	5	0,05	1	0,25	1	0

TABLE 3.2 –

rejetée

↪ **Procédure sous SPSS :** Reprendre les données de l'exemple 4. On souhaite savoir si le niveau d'utilisation suit une distribution normale, pour cela on suit les étapes suivantes (voir figure3.11) :

1. Ouvrir le fichier **Utilisation Internet**
2. On clique su **Analyse** puis **Tests non paramétriques** et **K-S à 1 échantillon**.
3. Dans le menu **Test de Kolmogorov-Smirnov à un échantillon** Glisser la variable **Utilisation** dans le quadrant **liste des variables à tester**, cocher la case devant **Normale** puis cliquer sur **OK**. Les résultats apparaissent dans la fenêtre des résultats (figure 3.12)

La figure 3.12 montre que la probabilité d'observer une valeur de K égale à $0,222$ déterminée par la statistique Z normalisée, est de $0,103$. Cette probabilité étant supérieur au niveau de signification $0,05$, l'hypothèse nulle ne peut être rejetée, la distribution de l'utilisation d'Internet ne présente donc pas de déviation significative par rapport à la distribution normale.

Variable de type :Intervalle ou Ratio

Test-t et test-Z (Tests de conformité d'une moyenne à une norme)

↪ **Principe :** Ces deux tests permettent de comparer la moyenne des réponses de l'échantillon à celle estimée dans la population-mère (ou à une valeur théorique μ_0). La connaissance de la loi de probabilité de la variable de décision est indispensable pour mener un test à son terme. De ce fait, les tests de conformité sur une moyenne imposent de distinguer les cas suivants :

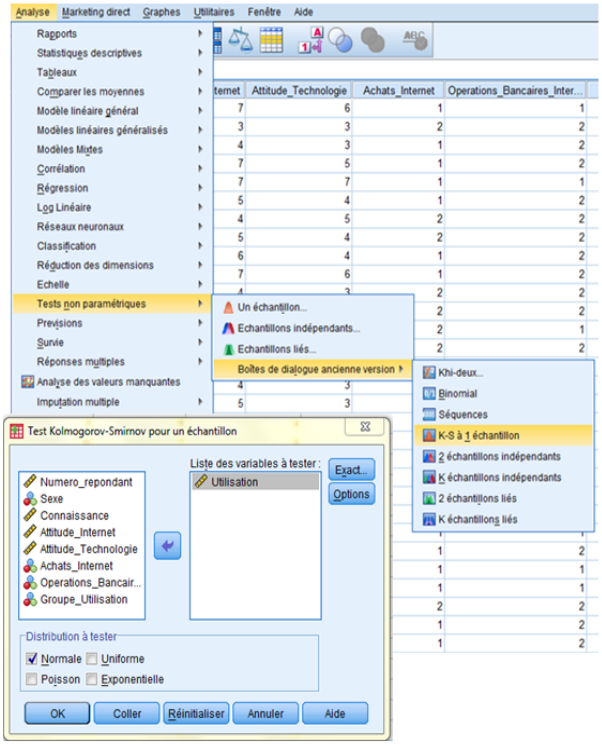


FIGURE 3.11 –

Test de Kolmogorov-Smirnov à un échantillon

		Utilisation d'Internet Hrs/Week
N		30
Paramètres normaux[a,b]	Moyenne	6,60
	Ecart-type	4,296
Différences les plus extrêmes	Absolue	,222
	Positive	,222
	Négative	-,142
Z de Kolmogorov-Smirnov		1,217
Signification asymptotique (bilatérale)		,103

a. La distribution à tester est gaussienne.

b. Calculée à partir des données.

FIGURE 3.12 –

- échantillon gaussien de taille quelconque,
- échantillon non gaussien de grande taille.

Pour la même raison, deux sous-situations doivent être prises en compte

- variance connue,
- variance inconnue [1].

Lorsqu'ils sont également applicables, les deux tests t et Z sont équivalents et leur mode de calcul est identique. La distribution des valeurs de référence est cependant issue de deux lois différentes : loi de Student pour le test t et loi normale centrée réduite pour le test Z. Dans l'hypothèse d'un échantillon d'une taille supérieure à 30 observations, aucune contrainte de distribution ne s'oppose à l'utilisation de l'un ou l'autre des deux tests. Lorsque l'échantillon est d'une taille inférieure à 30 observations et sous réserve

1. les logiciels statistique (comme SPSS) supposent toujours que la variance est inconnue

d'une distribution normale des réponses, le test t est le plus approprié. Pour de petites tailles d'échantillons, dans le cas où l'hypothèse de normalité de la distribution est rejetée, aucun des deux tests ne s'applique correctement : il faut alors recourir à un test non paramétrique. Le repère 6 regroupe tous les cas de figure évoqués liés au tests de conformité d'une moyenne à une norme , i compris leurs variables de décision à utiliser, leurs lois de probabilité et la forme des régions critiques.

Repère 6 (Tests de conformité sur la moyenne), *Source :[17]*

<table>
<tr><th rowspan="2">Paramètres
Connus/inconnus σ^2</th><th rowspan="2">Variable de décision
sous $H_0 : \mu = \mu_0$</th><th colspan="2">Distribution d'échantillonnage deT_n</th></tr>
<tr><th>Gaussien
taille quelconque</th><th>Non gaussien
grande taille
$n > 30$</th></tr>
<tr><td>Variance connue</td><td>$\downarrow$
$T_n = \dfrac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$
Régions critiques
$\downarrow$</td><td>Cas 1
$T_n \sim \mathcal{N}(0;1)$</td><td>Cas 1 bis
$T_n \to \mathcal{N}(0;1)$</td></tr>
<tr><td>$\begin{cases} H_0 : \ \mu = \mu_0 \\ H_1 : \ \mu = \mu_1 \\ \mu_1 > \mu_0 \end{cases}$</td><td>$\to \to$</td><td colspan="2">$\mathcal{C}_r = [z_{(1-\alpha)};\ +\infty[$</td></tr>
<tr><td>$\begin{cases} H_0 : \ \mu = \mu_0 \\ H_1 : \ \mu > \mu_0 \end{cases}$</td><td>$\to \to$</td><td colspan="2">$\mathcal{C}_r = [z_{(1-\alpha)};\ +\infty[$</td></tr>
<tr><td>$\begin{cases} H_0 : \ \mu = \mu_0 \\ H_1 : \ \mu < \mu_0 \end{cases}$</td><td>$\to \to$</td><td colspan="2">$\mathcal{C}_r =]-\infty;\ z_\alpha]$</td></tr>
<tr><td>$\begin{cases} H_0 : \ \mu = \mu_0 \\ H_1 : \ \mu \neq \mu_0 \end{cases}$</td><td>$\to \to$</td><td colspan="2">$\mathcal{C}_r =]-\infty;\ z_{1-\alpha/2}] \cup [z_{1-\alpha/2};\ +\infty[$</td></tr>
<tr><td rowspan="2">Variance inconnue</td><td>$T_n = \dfrac{\bar{X}_n - \mu_0}{S/\sqrt{n}}$</td><td>Cas 2
$T \sim S_{t_{n-1}}$</td><td>Cas 2 bis
$T_n \to \mathcal{N}(0;1)$</td></tr>
<tr><td>Régions critiques</td><td>Idem Cas1 et Cas1bis
en remplaçant z_α par t_α</td><td>Idem Cas1 et Cas1bis</td></tr>
</table>

↪ **Exemple d'application :** Un fabricant de cordes prétend que la tension de rupture de sa corde haut de gamme, spécialement conçue pour le canyoning, est de 2 800 kg. Un institut spécialisé achète 41 cordes pour vérifier les spécifications du produit. Les tensions de rupture des 41 cordes sont consignées dans le tableau ci-dessous :

2793	2802	2796	2799	2806	2801	2800	2793
2800	2795	2795	2796	2801	2799	2797	2800
2800	2800	2796	2802	2800	2799	2804	2804
2800	2802	2800	2798	2797	2802	2805	2802
2802	2800	2802	2804	2801	2801	2802	2796
2802							

TABLE 3.3 – Tension de rupture en kg de 41 cordes

Il s'agit de tester au seuil $\alpha = 10\%$ l'affirmation du fournisseur de cordes sachant qu'il

est avéré que la tension de rupture suit une loi de Gauss.

1. **Les hypothèses**
   Pour des raisons de sécurité envers les utilisateurs, il est important de vérifier que la tension de rupture n'est pas inférieure à 2 800 kg. L'hypothèse de recherche H_1, doit donc être $\mu < 2800$. Le test à résoudre s'écrit :

$$\begin{cases} H_0 : & \mu = 2800 \\ H_1 & \mu < 2800 \end{cases}$$

   Le risque de première espèce α est fixé à une valeur relativement élevée (10%) car il est important ici de diminuer le risque β de ne pas rejeter L'hypothèse nulle alors qu'en réalité la tension de rupture est inférieure à 2 800 kg.

2. **Variable de décision**
   Le test porte sur l'espérance de la variable aléatoire parente, la moyenne empirique X_n servira donc pour décider entre les deux hypothèses. La loi de probabilité de X, dépend de l'écart type σ de X ; comme cet écart type est inconnu, la variable de décision adaptée est $\dfrac{X_n - 2800}{S/\sqrt{n}}$. L'échantillon étant gaussien par suite T_n suit une loi de Student [2] à $(n-1)$ degrés de liberté.

3. **La région critique**
   L'idée est de rejeter l'hypothèse nulle $\mu = 2800$ si la moyenne empirique prend une valeur qui s'éloigne (trop) par valeur inférieure, de 2 800 kg. Autrement dit, le rejet de H_0 inter vient si la différence $X_n - 2800$ devient (trop) négative. Au seuil de 10% le rejet de H_0 a donc lieu si : $T_{41} = \dfrac{X_{41} - 2800}{S/\sqrt{41}} \leq t_{0,10}$ ou $t_{0,10}$ est le quantile d'ordre 10% de la loi de Student à 41 ddl. La région critique est ainsi de la forme :

$$\mathcal{C}_r =]-\infty;\ t_{0,10}].$$

   Pour 40 ddl, la valeur critique $t_{0,10}$ vaut -1,303, d'où la région critique

$$\mathcal{C}_r =]-\infty;\ -1,303].$$

4. **La décision**
   D'après les données du tableau 3.3 $\bar{X} = 2799,85$ et $S = 3,10$ d'où : $t_{obs} = \dfrac{2799,8 - 2800}{3,10/\sqrt{41}} = -0,310$.
   La réalisation de la statistique d'échantillon n'appartient pas à la zone de rejet de l'hypothèse nulle, on ne peut donc pas rejeter cette hypothèse. On ne Peut pas contester au seuil de 10%, l'affirmation du fabricant selon laquelle la tension moyenne de résistance des cordes est de 2 800 kg. Le problème est qu'il n'est pas possible de déterminer la probabilité que cette décision soit mauvaise. En effet, le risque encouru n'est pas le risque de première espèce de 10% ; le risque encouru est le risque de deuxième espèce β. Or, ce risque se calcule avec la loi de la variable de décision sous l'hypothèse H_1, et les paramètres de cette loi (en particulier l'espérance) ne sont pas connus

↪ **Procédure sous SPSS :** (figure3.13)

On souhaite vérifier l'hypothèse d'une moyenne de connaissance dépassant 4. Le niveau de signification choisi est de 0,05. La procédure sous SPSS comprend les étapes suivantes :

2. comme la taille de l'échantillon dépasse 30 le test Z est aussi applicable

1. On choisit sur SPSS le chemin suivant : **Analyse** > **Comparer les moyennes** > **Test T pour échantillon unique** .
2. On sélectionne la variable Connaissance dans la fenêtre **Variable(s) à tester**. On tape la valeur 4 dans **Valeur de test**. On peut cliquer sur **OK**. On obtient alors toutes les statistiques calculées (voir figure 3.14)
3. Si on veut changer le risque d'erreur on clique sur **Options** et on obtient le menu qui permet de modifier le niveau de confiance souhaité $(1-\alpha)$.

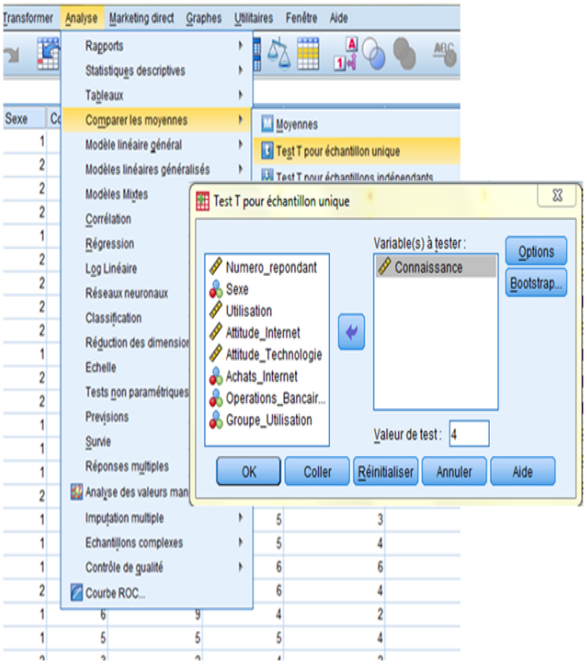


FIGURE 3.13 –

La statistique T présente $n-1=28$ degrés de liberté et a pour valeur $t=2,47$, la probabilité d'obtenir une valeur supérieure à $2,47$ est égale $0,02/2=0;01$ [3] inférieure à 0,05. L'hypothèse nulle est donc rejetée. Le niveau de connaissance n'excède pas 4.

Statistiques sur échantillon unique

	N	Moyenne	Ecart-type	Erreur standard moyenne
Connaissance	29	4,72	1,579	,293

Test sur échantillon unique

	Valeur du test = 4					
					Intervalle de confiance 95% de la différence	
	t	ddl	Sig. (bilatérale)	Différence moyenne	Inférieure	Supérieure
Connaissance	2,470	28	,020	,724	,12	1,32

FIGURE 3.14 –

3. SPSS affiche les sig pour des tests bilatéraux, pour avoir le sig d'un test unilatéral il suffit de diviser la valeur de SPSS par 2

3.2.2 Deux échantillons indépendants

Variable Nominale

Test d'indépendance de χ^2

↪ **Principe :** Les tris croisés présentent la distribution des fréquences de réponse pour deux ou plusieurs variables mises en relation mais ils ne permettent pas de démontrer l'existence de cette association du point de vue statistique. Pour mesurer véritablement la relation entre les variables, il est nécessaire de mettre en place des tests de signification statistique de l'association.

Le test le plus couramment utilisé est celui du khi-deux χ^2, car il consiste à tester la signification statistique d'une association de deux variables qualitatives (nominales ou ordinales). Plus précisément, il a pour objet de tester l'indépendance des variables dans un tableau croisé en comparant la distribution observée N_{ij} sur l'échantillon à une distribution théorique T_{ij} qui correspond à l'hypothèse que l'on veut rester.

Pour ce cas de figure, le test du Chi-deux permet de déterminer s'il existe une relation entre une ou plusieurs caractéristiques, que partagent les répondants appartenant à un même échantillon, et une variable de mesure commune aux deux échantillons. La formule de calcul du χ^2 est égale à :

$$\chi^2_{calc} = \sum_{i=1}^{i=r} \sum_{j=1}^{i=2} \frac{(N_{ij} - T_{ij})^2}{T_{ij}}$$

N_{ij} : fréquence observée dans la catégorie i.

T_{ij} : fréquence théorique (attendue) dans la catégorie i.

2 et r : nombre de colonnes et nombre de lignes du tableau de contingence.

↪ **Exemple d'application :**

Soit une variable nominale d'intention d'achat d'un véhicule de marque X relevée auprès de répondants se souvenant ou non d'avoir vue une publicité automobile. Nous cherchons à caractériser la relation entre le fait de se souvenir de la publicité testée et l'intention d'achat. Le tableau de contingence ci-dessous donne la répartition en effectifs et en pourcentage de la distribution observée et de la distribution théorique. La distribution théorique correspond à une absence d'association entre les deux variables. Le calcul de

En effectifs et % colonne	**Se souvenant de la publicité (n=314)**				**Ne se souvenant pas de la publicité (n=952)**			
	Observé		**Théorique**		**Observé**		**Théorique**	
	Effectif	%	**Effectif**	%	**Effectif**	%	**Effectif**	%
Intentionnistes (n=978)	**274**	**87,3**	**243**	**77,3**	**704**	**73,9**	**735**	**77,3**
Non-intentionnistes (n=288)	40	12,7	71	22,7	248	26,1	217	22,7
Total (n=1266)	314	100	314	100	952	100	952	100

TABLE 3.4 –

la statistique du Chi-deux à partir de la formule donnée ci-dessus nous donne une valeur calculée de 23,06. Le nombre de degré de liberté est $(r-1)(2-1) = 1$ (ce qui correspond au nombre minimum de cases qu'il suffit de connaitre pour compléter l'ensemble du tableau, étant donné que les totaux en ligne et colonne sont connus). La valeur du Chi-deux

pour 1 degré de liberté et avec un seuil de confiance de 99% est de $6,63$. Dans la mesure où la valeur calculée ici supérieur à la valeur critique, on doit donc rejeter l'hypothèse H_0. Les écarts entre la distribution observée et la distribution théorique sont trop grands pour qu'ils puissent se produire par hasard sous H_0. En conclusion, d'après les données de l'échantillon, on peut affirmer qu'il existe une relation entre le fait de se souvenir de la publicité et l'intention d'achat d'un véhicule de la marque X.

↪ **Procédure sous SPSS :** (voir figure 3.15) On souhaite tester l'hypothèse H_0 : Utilisation Internet [4] et Sexe sont indépendantes contre H_1 par la méthode du Chi-deux. Pour se faire on applique la procédure suivante

1. Nous avons réparti la variable utilisation en deux classes selon qu'elle est inférieure ou supérieure à 6. La nouvelle variable est **Groupe-utilisation**.
2. Pointer sur **Analyse** puis **Statistiques descriptives** puis **tableaux croisés**
3. Dans le menu **tableaux croisés** on sélectionnes les deux variables dont on veut tester l'indépendance : on sélectionne la variable **GroupeUtilisation** en ligne et variable **Sexe** en colonne.
4. Cliquant sur **Statistiques** on sélectionne le test du Chi-deux **Khi-deux** et clique sur **Poursuivre**.
5. Dans la rubrique **Cellules** on sélectionne **effectif observé** et **effectif théorique**

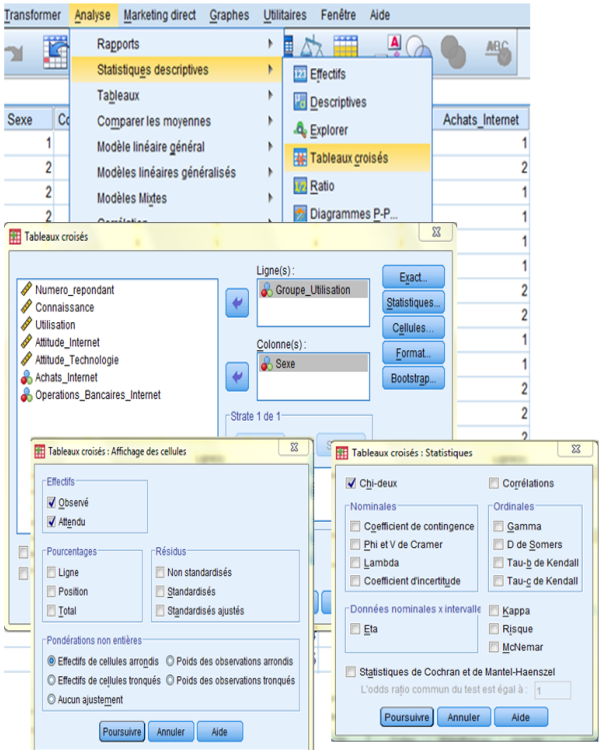


FIGURE 3.15 –

Comme le tableau comprend deux lignes et deux colonnes (tableau 2×2), on applique un facteur de correction [5]. On obtient la valeur de $2,133$ qui n'est pas significative pour un seuil $0,05$ (sig=0,114) (voir figure3.16).

4. Cette variable est de type ratio ; la variable **GroupeUtilisation** (dichotomique) est celle qui est introduite dans le test

5. Les statisticiens ne sont pas tous d'accord. Certains d'entre eux déconseillent d'appliquer une correction

Remarque 7

1. *Dans le cas d'un tableau 2×2, le Khi-deux est lié au coefficient phi (ϕ). Ce coefficient sert à mesurer l'intensité d'association. Il est proportionnel à la racine carré du khi-deux. Pour un échantillon de taille n, il s'obtient par la formule suivante :*

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

Pour notre exemple, ce coefficient ne représente aucun intérêt. Pour plus de détaille concernant ce genre de coefficients voir le **repère** *7.*

2. *Le test de Khi-deux présent deux limites. La première limite concerne le cas où certains effectifs théoriques sont trop faibles (<5) et la seconde peut se rencontrer si une (ou plusieurs) variable(s) est (sont) cachée(s) (voir [5] pour plus de détaille).*

3. *Lorsqu'un tableau de contingence comprend 4 cases, le test revient à comparer deux pourcentages, Nous avons vu que ce test pouvait être réalisé par le test de Khi-deux à 4 cases à condition que les effectifs théoriques soient supérieurs ou égaux à 5. Lorsque cette condition n'est pas remplie, il existe une façon exacte de tester l'homogénéité de 2 distributions de 2 variables binaires :* **le test exact de Fisher** *(pour plus de détaille voir [1])*

Tableau croisé Groupe d'utilisation d'Internet * Sexe

Effectif

		Sexe		
		Male	Female	Total
Groupe d'utilisation d'Internet	Light Users	5	10	15
	Heavy Users	10	5	15
Total		15	15	30

Tests du Khi-deux

	Valeur	ddl	Signification asymptotique (bilatérale)	Signification exacte (bilatérale)	Signification exacte (unilatérale)
Khi-deux de Pearson	3,333[a]	1	,068		
Correction pour la continuité[b]	2,133	1	,144		
Rapport de vraisemblance	3,398	1	,065		
Test exact de Fisher				,143	,072
Association linéaire par linéaire	3,222	1	,073		
Nombre d'observations valides	30				

a. 0 cellules (,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 7,50.

b. Calculé uniquement pour un tableau 2x2

FIGURE 3.16 –

Repère 7 *:***(Mesures d'association)***[5] Un Khi-deux observé χ^2_{obs} supérieur à la valeur critique signifie qu'il existe une liaison significative entre les deux caractères qualitatifs étudiés. Mais la valeur du Khi-deux dépend de l'effectif total n. Pour comparer de liaisons significatives de deux échantillons différents, il faut donc normaliser le Khi-deux. Plusieurs indices ont été proposés dans la littérature statistique. Nous présentons ci-dessous quelques indice classiques :*

Coefficient phi *On a déjà parler de ce coefficient dans la remarque ci-dessus.*

Coefficient de contingence *Ce coefficient noté C, permet d'évaluer l'intensité d'association pour un tableau de taille quelconque. Il est défini par*

$$C = \sqrt{\frac{\chi^2_{obs}}{\chi^2_{obs} + n}}$$

Le coefficient de contingence est toujours compris entre 0(indépendance) et 1 (qui n'est jamais atteint) $0 \leq C < 1$

V de Cramer *Il est définit par*

$$V = \sqrt{\frac{\phi^2}{min(I-1, J-1)}}, (I \text{ Nombre de lignes } J \text{ nombre de colonnes})$$

Le V de Cramer est toujours compris entre 0 (absence d'association) et 1 (Association parfaite)

T de Tschuprow *Il est défini par*

$$T = \sqrt{\frac{\phi^2}{\sqrt{(I-1)(J-1)}}}$$

Ce coefficient est toujours compris entre 0 et 1 (Liaison fonctionnelle)

Variable Ordinale

Test de Mann-Whitney

↪ **Principe :** ([11]) Lorsqu'on dispose de deux échantillons indépendants et que la variable traité est de niveau ordinal, le test de Mann-Whitney est un test approprié. Il s'assure qu'il existe une différence entre la distribution des fréquences d'une variable ordinale, mesurée auprès de deux échantillons indépendants. Le test repose sur le fait que, si la distribution d'une variable ordinale est identique auprès de deux échantillons indépendants, alors les rangs qu'occupent les observations, issues de l'un ou l'autre des des deux échantillons se répartissent aléatoirement, lorsqu'on réunit les répondants. Pour cela, on combine les observations des deux échantillons et on les numérote par ordre croissant de 1 à $(n_i + n_j)$, n_i, et n_j désigne les tailles des échantillons i et j respectivement. La statistique U_i calcule combien de fois le rang des observation issues de l'échantillon i dépasse le rang des observations j. On retient alors la plus petite valeur de U_i ou de U_j, qui sont liées par la formule : $U_i = n_i.n_j - U_j$.

$$U_i = n_i.n_j + \frac{n_i(n_i+1)}{2} - R_i$$

U_i **:** nombre de fois pour lesquelles le rang des observations de l'échantillon $[i]$ dépasse le rang des observations de l'échantillon $[j]$

R_i **:** la somme des rangs des observations de l'échantillon $[i]$.

Lorsque la taille du plus grand des deux groupes est supérieure à 20, la distribution d'échantillonnage de U se rapproche d'une distribution normale et l'hypothèse nulle H_0 peut être vérifiée en recourant à la distribution normale standardisée Z de moyenne μ_U et d'écart-type σ_U. Pour les autre cas il existe une table des valeur critiques de U (Voir l'annexe).

Test Z	**Moyenne μ_U**	**Ecar-type σ_U**
$Z = \frac{U - \mu_U}{\sigma_U}$	$\mu_U = \frac{n_i.n_j}{2}$	$\sigma_U = \sqrt{\frac{n_i.n_j.(n_i + n_j + 1)}{12}}$

↪ **Exemple d'application :**

On considère par exemple, les scores bruts observés dans un échantillon partagé en hommes (A : 10 observations)et femmes (B : 10 observations) représentés dans les deux premières colonnes du tableau 3.5, le classement joint (de 1 à 20) est représenté dans les troisième et quatrième colonnes.

Scores bruts		**Rangs combinés**	
A (hommes)	**B (femmes)**	**A**	**B**
69	82	6	2
76	32	4	15
30	44	16	12
54	49	8	11
85	73	1	5
28	20	18	20
77	24	3	19
63	52	7	9
51	29	10	17
38	39	14	13
		87	123

TABLE 3.5 – Exemple de liaison entre une variable ordinale et une variable nominale (Binaire), Source[7]

L'application de la procédure donne les résultats numériques suivants :

n_A=10 $n_B = 10$ $n = 20$
$R_A = 87$ $R_B = 123$ $R_A + R_B = 210$
$U_A = 100 + 55 - 87 = 68$
$U_B = 100 + 55 - 123 = 32$
$min(U_A, U_B) = 32$

En examinant une table de U de Mann-Whitney, on constate que $U_{obs} > U_\alpha$ dont la valeur est 27 pour $\alpha = 0,05$ (le test est unilatérale). Et ce résultat conduit à ne pas rejeter l'hypothèse nulle d'une égalité entre les deux classements. Cette conclusion est évidement identique à celle obtenue en considérant le seuil de signification observé.

↪ **Procédure sous SPSS :** On peut examiner les différences entre hommes et femmes vis-à-vis de l'utilisation d'Internet, en recourant au test U de Mann-Whitney. Pour ce faire on doit suivre les étapes suivantes (voir figure 3.17) :

1. On clique sur **Analyse**, puis **Tests non paramétriques**, puis **échantillons indépendants**
2. Faites glisser la variable à tester **Utilisation** dans le champs **Liste des variables à tester**.
3. Cliquer sur **Définir les groupes**. Il vous faut ensuite entrer les codes correspondant aux deux groupes : 1 pour **Male** et 2 pour **Female**. Cliquer sur **Poursuivre** afin de revenir à la fenêtre précédente.
4. Vous pouvez cliquer sur le bouton **Optios**, ce qui vous permet de modifier le niveau de confiance pour les estimations par intervalle si nécessaire.
5. Cliquer ensuite sur **Poursuivre**, puis sur **OK**.

La fenêtre des résultats (viewer) vous donne alors différentes indications voir figure 3.18)

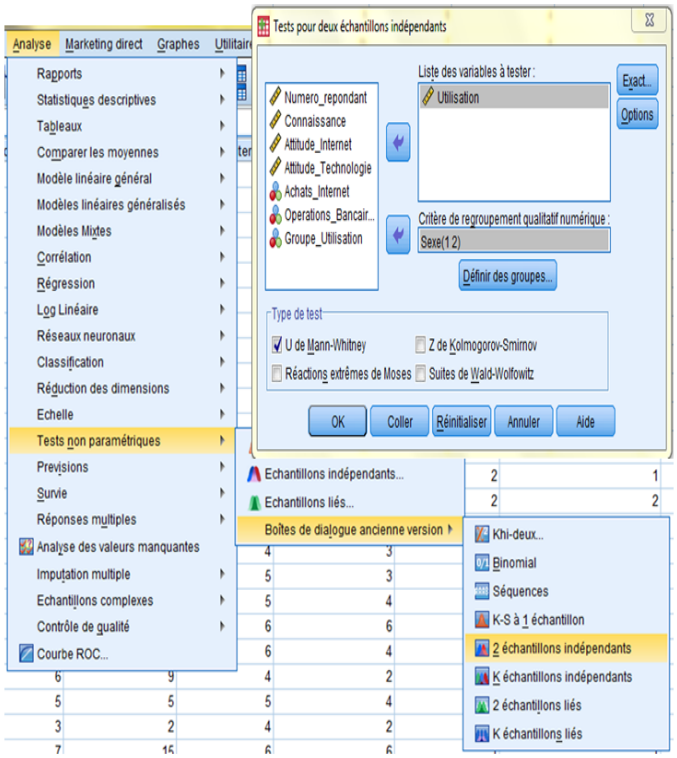


FIGURE 3.17 –

On constate une différence significative les deux groupes ($U = 31$ avec un sig=$0,001 < 5\%$). Les observations étant classées de la plus petite à la plus grande, le classement moyen des hommes (20,93) montre qu'ils utilisent plus Internet que les femmes (10,07).

Rangs

	Sexe	N	Rang moyen	Somme des rangs
Utilisation d'Internet Hrs/Week	Male	15	20,93	314,00
	Female	15	10,07	151,00
	Total	30		

Test[b]

	Utilisation d'Internet Hrs/Week
U de Mann-Whitney	31,000
W de Wilcoxon	151,000
Z	-3,406
Signification asymptotique (bilatérale)	,001
Signification exacte [2*(signification unilatérale)]	,000[a]

a. Non corrigé pour les ex aequo.

b. Critère de regroupement : Sexe

FIGURE 3.18 –

Variable de type :Intervalle ou Ratio

Test-t et Test-Z

↪ **Principe :** Dans le cas de deux échantillons indépendants, une adaptation du test t permet de vérifier si la différence de deux moyennes est significative ou imputable aux seuls aléas de l'échantillonnage. Le repère 8 regroupe tous les cas de figure liés au tests de comparaison de deux moyennes pour deux échantillons indépendants, i compris leurs variables de décision à utiliser, leurs lois de probabilité et la forme des régions critiques.

Repère 8 (Tests de comparaison de moyennes, échantillons indépendants), *Source :[17]*

		Distribution d'échantillonnage de T	
Paramètres Connus/inconnus σ_A^2 *et* σ_B^2	**Variable de décision T sous** $H_0 : \mu_A - \mu_B = 0$	**Gaussien taille quelconque**	**Non gaussien grande taille** $n_A > 30$ *et* $n_B > 30$
Variances connues	$\downarrow$ $T = \dfrac{\bar{X}_{n_A} - \bar{X}_{n_B}}{\sqrt{\dfrac{\sigma_A^2}{n_A} + \dfrac{\sigma_B^2}{n_B}}}$ *Régions critiques* $\downarrow$	**Cas 1** $T \sim \mathcal{N}(0;\ 1)$	**Cas 1bis** $T \to \mathcal{N}(0;\ 1)$
$\begin{cases} H_0 : & \mu_A = \mu_B \\ H_1 : & \mu_A > \mu_B \end{cases}$	$\to \to$	$\mathcal{C}_r = [z_{(1-\alpha)};\ +\infty[$	
$\begin{cases} H_0 : & \mu_A = \mu_B \\ H_1 : & \mu_A < \mu_B \end{cases}$	$\to \to$	$\mathcal{C}_r =]-\infty,\ z_\alpha]$	
$\begin{cases} H_0 : & \mu_A = \mu_B \\ H_1 : & \mu_A \neq \mu_B \end{cases}$	$\to \to$	$\mathcal{C}_r =]-\infty,\ z_{(1-\alpha/2)}] \cup \mathcal{C}_r = [z_{(1-\alpha/2)};\ +\infty[$	
Variances inconnues Supppées égales (test préliminaire)	$T = \dfrac{\bar{X}_{n_A} - \bar{X}_{n_B}}{\sqrt{\dfrac{S_p^2}{n_A} + \dfrac{S_p^2}{n_B}}}$ [a]	**Cas 2** $T \sim St_{(n_A+n_B-2)}$	**Cas 2bis** $T \to \mathcal{N}(0;\ 1)$
	Régions critiques	*Idem Cas1 et Cas 1bis en remplaçant* z_α *par* t_α	*Idem Cas1 et Cas 1bis*
Variances inconnues Supppées différentes	$T = \dfrac{\bar{X}_{n_A} - \bar{X}_{n_B}}{\sqrt{\dfrac{S_{n_A}^2}{n_A} + \dfrac{S_{n_B}^2}{n_B}}}$	**Cas 3** $T \sim St_n$ [b]	**Cas 3bis** $T \to \mathcal{N}(0;\ 1)$
	Régions critiques	*Idem Cas1 et Cas 1bis en remplaçant* z_α *par* t_α	*Idem Cas1 et Cas 1bis*

a. $S_p = \frac{(n_A-1)S_A^2+(n_B-1)S_B^2}{n_A+n_B-2}$ (**Variance de pool ou combinée**)

b. $n \cong \dfrac{(S_{n_A}^2/n_A + S_{n_B}^2/n_B)^2}{\dfrac{(S_{n_A}^2/n_A)^2}{n_A-1} + \dfrac{(S_{n_B}^2/n_B)^2}{n_B-1}}$ ou bien $n \cong min[(n_A-1);(n_B-1)]$ si $n_A \cong n_B$.

↪ **Exemple d'application :** ([17])

Les associations de consommateurs font appel à des organismes indépendants pour tester, pour de nombreux produits, les caractéristiques annoncées par les fabricants. Le sujet du mois de la publication d'une de ces associations concerne les laves linges.

Une des questions soulevée est relative à l'influence sur la consommation d'électricité de l'utilisation d'un adoucisseur d'eau pour alimenter la machine à laver. L'entartrage des

machines n'intervenant qu'après une utilisation prolongée, 60 machines TX100 âgées de 4 ans sont testées. La moitié d'entre elles ont toujours été branchées sur un adoucisseur d'eau, l'autre moitié a toujours fonctionné sans adoucisseur. le tableau suivant consigne les consommations relevées en kilowatts par heure pour le même programme de lavage.

Consommation avec adoucisseur (A)									
1,07	0,79	0,66	0,59	0,83	0,80	0,87	0,93	0,75	0,78
0,68	0,71	0,82	0,76	0,93	0,82	0,74	0,77	0,81	0,89
0,91	1,11	0,78	0,79	0,79	0,95	0,77	0,94	0,70	0,77
Consommation sans adoucisseur (B)									
1,01	0,91	0,81	0,93	0,89	0,90	0,88	1,07	1,04	0,77
0,92	0,82	0,67	0,90	0,81	0,93	0,88	0,87	0,95	0,91
0,92	0,83	0,91	0,93	0,75	0,87	1,06	0,89	0,90	0,68

TABLE 3.6 – Consommation électrique en kilowatts par heure du lave linge TX100, Source [17]

L'organisme de contrôle a déjâ validé l'hypothèse selon laquelle la consommation électrique de la TX100 suit une loi de gauss (loi normale). Pour décider de l'efficacité de l'adoucisseur l'organisme décide d'effectuer des tests de comparaison (consommations moyenne d'électricité)dans les deux situations d'alimentation. En d'autre terme l'adoucisseur d'eau est-il efficace pour réduire la consommation d'électricité ? Le seuil de signification du test est fixé à 5%.
Pour répondre à cette interrogation, il faut tester les consommations moyennes des deux populations. L'indice A se rapportant aux machines utilisées avec un adoucisseur, le test unilatéral posé est :

$$\begin{cases} H_0 : & \mu_A = \mu_B \\ H_1 : & \mu_A < \mu_B \end{cases}$$

En posant le test de la sorte, on peut fixer à 5% la probabilité de décider que l'adoucisseur est utile alors qu'en réalité il ne l'est pas.
Le travail de modélisation et de simulation effectué dans la première partie du défi a montré que les hypothèses de travail sont les suivants :

- deux échantillons gaussiens,
- où les variances sont inconnues mais égales.

La statistique qui sert de variable de décision est :

$$T = \frac{\bar{X}_{n_A} - \bar{X}_{n_B}}{\sqrt{\dfrac{S_p^2}{n_A} + \dfrac{S_p^2}{n_B}}} \text{ où } S_p^2 = \frac{n_A S_{n_A}^2 + n_B S_{n_B}^2}{n_A + n_B - 2}$$

En effet, comme les deux variances sont égales, il est pertinent de les estimer à l'aide du même estimateur. Celui-ci, appelé variance de pool. La loi de T est une loi de Student à $n_A + n_B - 2$ degrés de liberté (cas 2 du tableau du repère 8).
On sait que $S_{n_A}^2 = 0,0128$ et $S_{n_B}^2 = 0,0089$ 9 la réalisation de la variance de pool est donc :

$$S_p^2 = \frac{29 \cdot 0,0128 + 29 \cdot 0,008}{30 + 30 - 2}$$

Par ailleurs,les moyennes empiriques calculées dans les deux échantillons (tableau 1.1) sont $\bar{X}_{A30} = 0,8170$ et $\bar{X}_{B30} = 0,8870$.

On a alors $t_{obs} = \dfrac{0,8170 - 0,8870}{\sqrt{\dfrac{0,0109}{30}\dfrac{0,0109}{30}}}$.

La variable de décision suit une loi de Student à 58 degré de liberté, cette variable a une probabilité de 5% d'être inférieure à la valeur critique $t_{5\%} = -1,6716$ (LOI.STUDENT.INVERSE(0,10 ;58)=1,6716 sous Excel). La région critique est ainsi : $\mathcal{C}_r =]-\infty;\ -1,6716]$. La borne négative n'est pas surprenante. En effet, on doit rejeter l'hypothèse d'égalité des consommations moyennes si la différence $X_{A30}^{-} - X_{B30}^{-}$ est "trop" négative autrement dit si X_{A30}^{-} est "trop" inférieur à X_{B30}^{-}. Dans ce cas, c'est bien le modèle avec l'adoucisseur (le A) qui devrait être choisi.

La différence observée ($t_{obs} = 2,5921$) appartient ainsi à la zone de rejet de l'hypothèse nulle. On peut donc accepter l'hypothèse alternative $\mu_A < \mu_B$ avec un risque au plus égal à 5%. La valeur observée étant très éloignée de la valeur critique il est intéressant de calculer la probabilité critique pour calculer le risque exact. Celui-ci est égal $P[St_{58} < -2,5926]$. Le tableur Excel fournit cette probabilité : LOI.STUDENT (2,5926 ; 58 ;1) = 0, 0060. Cette étude permet de conclure, avec une probabilité de $0,60\%o$ de se tromper, que l'adoucisseur d'eau conduit à une diminution de la consommation moyenne d'électricité.

↪ **Procédure sous SPSS :**

À partir des données du tableau de l'exemple4 , on cherche à déterminer si l'utilisation d'Internet diffère en fonction du sexe. On effectue pour cela un test t sur deux échantillons indépendants, dont la figure3.19 présente les résultats. Le test F (**de Levene**) de la variance des échantillons possède une probabilité inférieure à $0,05$: H_0 est donc rejetée, et l'on doit utilisé un test t fondé sur une "hypothèse de non-égalité des variances". Les femmes utilisent en moyenne moins d'Internet (moyenne ± écart-type : $3,87 \pm 1,68$) que les hommes (moyenne ± écart-type : $9,33 \pm 4,40$). La différence moyenne entre les deux sexes (IC au niveau 95% : $[2,91; 8,02]$) correspond à un effet important de la variable sexe (d= 1,8) [6]. Le test de Student pour deux groupes indépendants montre que cette différence est significative ($t(18,01) = 4,492; p < 0,000$). On conclut donc que le sexe à un effet sur l'utilisation d'Internet.

Statistiques de groupe

	Sexe	N	Moyenne	Ecart-type	Erreur standard moyenne
Utilisation d'Internet Hrs/Week	Male	15	9,33	4,402	1,137
	Female	15	3,87	1,685	,435

Test d'échantillons indépendants

		Test de Levene sur l'égalité des variances		Test-t pour égalité des moyennes						
									Intervalle de confiance 95% de la différence	
		F	Sig.	t	ddl	Sig. (bilatérale)	Différence moyenne	Différence écart-type	Inférieure	Supérieure
Utilisation d'Internet Hrs/Week	Hypothèse de variances égales	15,507	,000	4,492	28	,000	5,467	1,217	2,974	7,960
	Hypothèse de variances inégales			4,492	18,014	,000	5,467	1,217	2,910	8,024

FIGURE 3.19 –

6. Ce score est plus facile à interprété. Il se nomme taille ou grandeur de l'effet, et se note d. d mesure l'écart entre les deux moyennes, exprimé en écarts type. On le calcule comme cela : $d = \dfrac{X_1 - X_2}{\text{écart-type moyen}} = \dfrac{X_1 - X_2}{\dfrac{S_1 + S_2}{2}}$

Pour obtenir les résultats inclus dans la figure 3.19 il suffit de suivre les étapes suivantes : (voir figure3.20).

1. On pointe sur **Analyse** puis sur **Comparer les moyennes** et **Test T pour échantillons indépendants**.
2. Dans le menu **test-T pour échantillons...** on retient la **variable à tester** Utilisation et la variable Sexe pour **variable de regroupement**
3. Cliquer sur **Définir groupes** on obtient le menu dans lequel on sélectionne les échantillons à comparer, ici om met **1** dans groupe 1 puis **2** dans groupe 2.
4. Cliquer sur **Options** on obtient le menu dans lequel on sélectionne le risque d'erreur ou inversement le niveau de confiance.

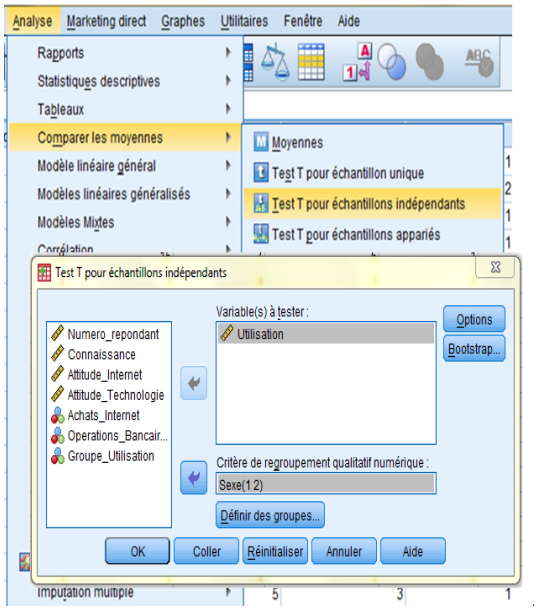


FIGURE 3.20 –

3.2.3 Deux échantillons dépendants (appariés)

Le fait de prendre des décisions à partir de données observées sur des échantillons pose toujours deux problèmes fondamentaux :

- Les différences observées dans les échantillons sont-elles liées justement à l'échantillon- nage ? Autrement dit, aurait-on observé ces différences si on avait eu l'opportunité d'interroger toute la population ? Un moyen de diminuer cette incertitude consiste à réduire l'erreur standard en augmentant, dans la mesure du possible, la taille des échantillons.
- Les différences observées dans les échantillons doivent-elles être seulement attribuées à la variable " explicative " prise en compte ? Par exemple, dans le cas du traitement à la pulmotrycine versus placebo, le fait que les patients traités se portent mieux n'est pas exclusivement lié au traitement mais peut-être au fait que les individus du groupe traité ont des caractéristiques propres (facteurs génétiques, moral, etc.). Pour limiter l'effet de tels facteurs exogènes, une solution consiste à effectuer les tests sur des échantillons appariés. Notons cependant que l'appariement n'est pas toujours envisageable (Demander aux mêmes individus d'être à la fois (ou successivement), sous traitement et sous placebo n'a guère de sens.)

Variable Nominale

Test de Mc Nemar

↪ **Principe :** [11] Cee test a une utilisation courante pour mesurer les changements d'attitude ou de comportement à la suite de l'exposition à un stimulus (offre promotionnelle, annonce publicitaire, nouveau concept, etc.)

La procédure fait appel à une matrice de 2 lignes et 2 colonnes, soit 4 cases respectivement numérotées A, B, C et D, dans lesquelles sont inscrites les fréquences d'opinions ou de préférence avant et après exposition au stimulus. En se reportant au schéma ci-dessous, on note aisément que les fréquences A et D correspondent aux sujets qui change d'avis entre les deux phases d'interrogation (inversement, les fréquences B et C correspondent aux répondants qui émettent le même avis).

		Après 2	**Après** 1
Avant	**1**	**A**	**B**
	2	**C**	**D**

$$\chi^2_{calc} = \frac{(\mid A - D \mid -1)^2}{A + D}$$

Le test statistique s'intéresse uniquement aux sujets qui change d'avis, étant entendu que, si les fréquences portées dans les cases A et D sont identiques, le résultat du test est à "somme nulle" : le nombre de répondants qui évoluent de l'opinion1 vers l'opinion2 est égale à celui des répondants qui adoptent l'opinion2, après avoir émis l'opinion1. L'hypothèse nulle H_0 revient à affirmer que le changement s'opère simultanément dans les deux sens. Pour tester cette hypothèse, on calcule un indice à l'aide de la formule ci-dessus [7]. Cet indice, sous l'hypothèseH_0, suit une distribution d'échantillonnage du Khi-deux à 1 degré de liberté. Le rejet de H_0, pour une valeur de Khi-deux calculé supérieure à sa valeur tabulée, conduit à admettre l'existence d'un effet induit par la manipulation du stimulus. Une précaution s'avère nécessaire : lorsqu'une des des fréquences mentionnées dans le tableau est inférieure à 5 répondants, il est préférable d'utiliser le test binomial

↪ **Exemple d'application :**[5] Un groupe de 100 personnes en age de voter participe à une étude politique. On leur demande tout d'abord leur opinion (favorable ou défavorable) au sujet d'un homme politique. On leur pose la même question 6 mois plus tard et on obtient les résultats du tableau suivant

Sondage1 \ Sondage2	favo	defa	Total
favo	42	20	62
defa	15	23	38
Total	57	43	100

7. Dans d'autres ouvrages on trouve la formule suivante $\frac{(A-D)^2}{A+D}$ voir par exemple [5] ou [1]

La formule du test de Mc Nemar donne :

$$\chi^2_{obs} = \frac{(\mid 20 - 15 \mid -1)^2}{20 + 15} = 0,457$$

et $\chi^2_{cal} = 3,84$. On conclut donc que ces données ne permettent pas de déceler une différence significative entre les résultats des deux sondages.

↪ **Procédure sous SPSS :** ([1]) On désire comparer deux techniques biologiques,l'ELISA et l'hémagllutination (IHAT) dans le diagnostic de l'hydatidose (kyste hydatique). Un total de 56 malades a été testé simultanément par chacune des 2 techniques. La performance (sensibilité) d'une technique est jugée par le nombre de résultats positifs observés (voir la table ci-dessous).

Résultat ELISA	**Résulta IHAT**	**Nombre de malades**
+	+	43
-	+	2
+	-	10
-	-	1

Pour l'application du test de Mc nemar sur ces doonées, il suffit de suivre les étapes suivantes (voir figure3.21) :

1. Pour informer SPSS qu'on se trouve dans le cas d'un tableau statistique : on sélectionne **Données** ⇒ **Pondérer les observations**
2. Déplacer la variable Nombre dans la zone "**variable de pondération**"
3. Cliquer ensuite sur **OK** pour revenir à l'écran principal.
4. Choisir **Analyse**, **Tests non paramétriques** puis **2 échantillons liés :** pour ouvrir le menu **Tests pour deux échantillons liés**
5. Glisser les deux variables **ELISA** et **IHAT** dans la case **Paires à tester**.
6. Choisir le test **Mc Nemar** dans le groupe **Type de test**
7. Cliquer si vous le désiriez sur **Options**, et finissez par **OK**

Les résultats apparaissent dans la fenêtre des résultats (figure 3.22). Le test est significatif (sig=0,039<0,05), les performances des deux techniques diffèrent significativement. L'ELISA est plus sensible que l'hémagglutination. Le χ^2 de confirmité (χ^2, $ddl = 3$) correspond à significativité $p = 0.000$ inférieure à 5%.

Variable Ordinale

Test de Wilcoxon

↪ **Principe :** ([11]) Dans la mesure où il s'applique à une variable ordinale, ce test permet de vérifier dans le cas d'une expérimentation à mesures répétées (test avant-après par exemple), la fréquence, la direction, mais aussi l'amplitude des variations de la variable ordinale mesurée. Pour chaque répondant, on dispose d'une paire d'observations pour laquelle on calcule la différence D_i des scores avant après exposition au stimulus. Lorsque les scores des réponses jumelées sont égaux, la différence D_i calculée étant nulle,la paire d'observations est écartée de la procédure. Ensuite, les valeurs absolus des différences pour tous les répondants sont ordonnées par ordre croissant et numérotées. Puis les rangs obtenus sont affectés d'un signe, tantôt positif, tantôt négatif, selon le sens de la différence à laquelle ils ont attachés. Le test porte sur une comparaison de la somme des rangs

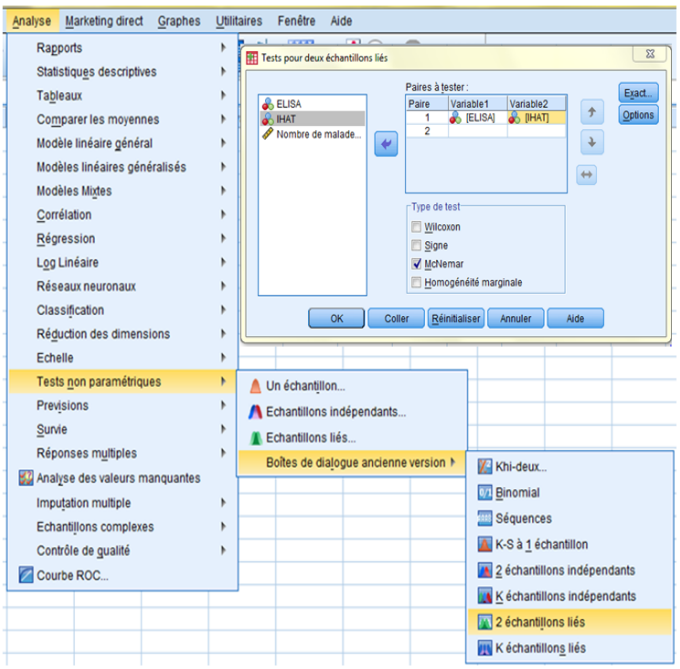


FIGURE 3.21 –

ELISA & IHAT

ELISA	IHAT	
	-	+
-	1	2
+	10	43

Test[b]

	ELISA & IHAT
N	56
Signification exacte (bilatérale)	,039[a]

a. Distribution binomiale utilisée.

b. Test de McNemar

FIGURE 3.22 –

positifs et de la somme des rangs négatifs, étant entendu que, sous l'hypothèse nulle H_0, les deux sommes sont égales. L'indice T de Wilcoxon est égal à la plus petite des deux sommes, négative ou positive. Lorsque le nombre d'observations est supérieur à 8, le test T de Wilcoxon suit approximativement une distribution normale dont les paramètres sont précisés ci-dessous :

Test Z	**Moyenne** μ_T	**Ecar-type** σ_T
$Z = \dfrac{T - \mu_T}{\sigma_T}$	$\mu_T = \dfrac{n(n+1)}{4}$	$\sigma_T = \sqrt{\dfrac{n(n+1)(2n+1)}{24}}$

↪ **Exemple d'application :** ([11]) 10 consommateurs interrogés sur leur intention d'achat d'une marque de biscuit qu'ils goûtent en "aveugle" (les répondants n'ont pas connaissance de la marque du fabriquant) à l'aide d'une note de 1 ("très incertain") à 10 ("tout à fait certain"). Les 10 consommateurs sont interrogés sur leur intention d'achat du même biscuit présenté avec sa marque. On cherche à savoir si le fait de connaitre la marque au

moment de goûter le produit à modifié leur intention d'achat. Les données recueillies se présente comme suit :

Consm-mateurs	**Intention d'achat**		**d**	**Rang de d**	**Plus petite des 2 sommes de rangs positifs (ou négatifs)**
	En aveugle	**Avec connaissance de la marque**			
1	3	10	7	9	
2	5	8	3	5	
3	9	7	-2	-2,5	2,5
4	5	5	0	-	
5	4	7	3	5	
6	4	8	4	7	
7	8	9	1	1	
8	6	4	-2	-2,5	2,5
9	3	6	3	5	
10	4	9	5	8	
					T=5,0

On cherche à tester l'hypothèse nulle H_0 selon laquelle la connaissance de la marque n'a pas d'effet sur l'intention d'achat du produit goûté. L'alternative H_1 est que l'intention d'achat diffère entre les deux phases- en aveugle et avec connaissance de la marque- du test. Le seuil de confiance est de 0,05 et le nombre de paires à considérer est de 9 puisque pour une paire la différence de rangs est nulle. La distribution d'échantillonnage de la statistique T est une distribution normale standardisée donnée par :

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} = \frac{5,0 - \frac{9(9+1)}{4}}{\sqrt{\frac{9(9+1)(18+1)}{24}}} = \frac{-17,5}{8,44} = -2,07$$

La valeur de Z est 1,96 au seuil de confiance 0,05 pour un test bilatéral : $t_{calc} < t_{\text{tabulé}}$; on ne peut pas rejeter l'hypothèse nulle au seuil de 5% et on conclut que la connaissance de la marque n'a pas d'effet sur l'intention d'achat du produit goûté.

↪ **Procédure sous SPSS :**([12]) On peut reprendre les données de notre exemple qui visait à déterminer si les répondants présentaient des différences quant à leurs attitudes vis-à-vis d'Internet et de la technologie. Imaginons que ces deux variables soient mesurées sur une échelle ordinale, et non sur sur une échelle d'intervalles. On applique par conséquent le test de Wilcoxon, dont les résultats figurent dans la figure 3.23. On constate à nouveau une différence significative entre les variables. Les différences négatives,témoignent d'une attitude moins favorable vis-à-vis de la technologie que vis-à-vis d'Internet, sont au nombre de 23. La moyenne de ces différences négatives s'élève à 12,72. On ne compte en revanche qu'une seule différence positive. La moyenne de cette différence est de 7,50. On dénombre enfin six égalités-des observations qui présentent les mêmes valeurs pour les deux variables. Ces données relèvent une attitude plus favorable par rapport à Internet que par rapport à la technologie. La probabilité associée à la statistique Z est en outre inférieure à 0,05, ce qui confirme bien que la différence est significative.

La réalisation du test de Wilcoxon passe par les étapes suivantes (voir figure3.24) :

1. Choisissez **Analyse** > **Tests non paramétriques** > **Deux échantillons liés**

Rangs

		N	Rang moyen	Somme des rangs
Attitude vis-à-vis de la technologie - Attitude vis-à-vis d'Internet	Rangs négatifs	23[a]	12,72	292,50
	Rangs positifs	1[b]	7,50	7,50
	Ex aequo	6[c]		
	Total	30		

a. Attitude vis-à-vis de la technologie < Attitude vis-à-vis d'Internet

b. Attitude vis-à-vis de la technologie > Attitude vis-à-vis d'Internet

c. Attitude vis-à-vis de la technologie = Attitude vis-à-vis d'Internet

Test[b]

	Attitude vis-à-vis de la technologie - Attitude vis-à-vis d'Internet
Z	-4,207[a]
Signification asymptotique (bilatérale)	,000

a. Basée sur les rangs positifs.

b. Test de Wilcoxon

FIGURE 3.23 –

2. Déplacez les deux variables (**Attitude-Internet** et **Attitude-Technologie**)dans la zone **Paire(s) à tester**. Cochez bien l'option **Wilcoxon**.

3. Si vous voulez faire en sorte que les statistiques descriptives sortent en même temps, choisissez **Options**, puis **Poursuivre** quand vos choix sont faits.

4. Cliquez maintenant **OK**.

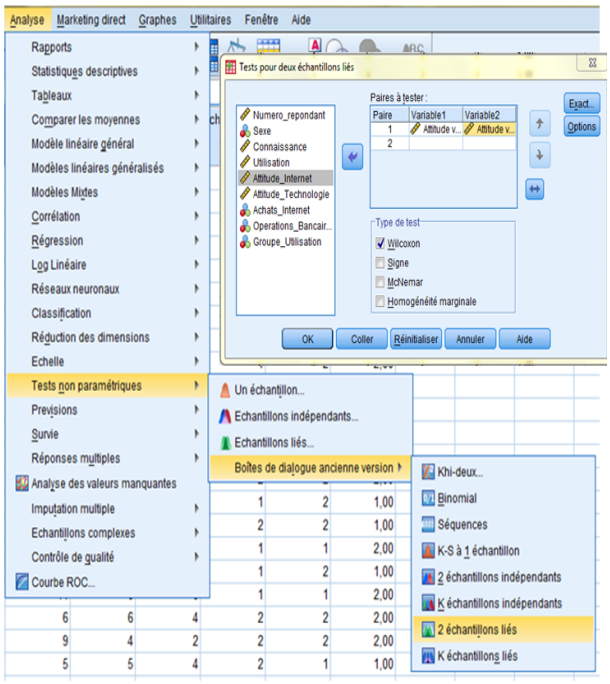


FIGURE 3.24 –

Variable de type :Intervalle ou Ratio

Test-t des différences

↪ **Principe :** ([12]) Un échantillon de répondants pourra par exemple noter deux marques concurrentes, indiquer l'importance relative de deux caractéristiques d'un même produit, ou encore évaluer une marque à deux moments différents. Dans ces cas-là, les différences sont analysées au moyen de l'extension d'un test **t**. Pour obtenir t, on définit d'abord la différence entre les observations d'une même paire, notée D, dont on calcule ensuite la moyenne et la variance. Les degrés de liberté sont au nombre de $n-1$, où n représente le nombre de paires. Le repère 9 regroupe tous les cas de figure liés au tests de comparaison de deux moyennes pour deux échantillons appariés, i compris leurs variables de décision à utiliser, leurs lois de probabilité et la forme des régions critiques.

Repère 9 (Tests de comparaison de moyennes, échantillons appariés), *Source :[17]*

<table>
<tr><td>$\mathbf{D}=X_A-X_B$</td><td>$\mu_D=\mu_A-\mu_B$</td><td colspan="2">Distribution d'échantillonnage de T</td></tr>
<tr><td>Paramètres
Connus/inconnus
σ_D^2</td><td>Variable de décision T
sous $H_0:\mu_D=0$</td><td>Gaussien
taille quelconque
$n=n_A=n_B$</td><td>Non gaussien
grande taille
$n=n_A=n_B>30$</td></tr>
<tr><td>Variances connues

(cas rare)</td><td>↓
$T=\dfrac{\bar{D}-\mu_D}{\sigma_D/\sqrt{n}}$
Régions critiques
↓</td><td>Cas 1
$T\sim\mathcal{N}(0;\ 1)$</td><td>Cas 1bis
$T\to\mathcal{N}(0;\ 1)$</td></tr>
<tr><td>$\begin{cases} H_0: & \mu_A=\mu_B \\ H_1: & \mu_A>\mu_B \end{cases}$</td><td>→ →</td><td colspan="2">$\mathcal{C}_r=[z_{(1-\alpha)};\ +\infty[$</td></tr>
<tr><td>$\begin{cases} H_0: & \mu_A=\mu_B \\ H_1: & \mu_A<\mu_B \end{cases}$</td><td>→ →</td><td colspan="2">$\mathcal{C}_r=]-\infty,\ z_\alpha]$</td></tr>
<tr><td>$\begin{cases} H_0: & \mu_A=\mu_B \\ H_1: & \mu_A\neq\mu_B \end{cases}$</td><td>→ →</td><td colspan="2">$\mathcal{C}_r=]-\infty,\ z_{(1-\alpha/2)}]\cup\mathcal{C}_r=[z_{(1-\alpha/2)};\ +\infty[$</td></tr>
<tr><td rowspan="2">Variances inconnues</td><td rowspan="2">$T=\dfrac{\bar{D}-\mu_D}{S_D/\sqrt{n}}$</td><td>Cas 2
$T\sim St_{(n_D-1)}$</td><td>Cas 2bis
$T\to\mathcal{N}(0;\ 1)$</td></tr>
<tr><td>Régions critiques</td><td>Idem Cas1 et Cas 1bis
en remplaçant z_α par t_α</td><td>Idem Cas 1bis</td></tr>
</table>

↪ **Exemple d'application :** Des chemises en soie haut de gamme sont cousues dans un atelier de confection. Deux méthodes (notées A et B) qui diffèrent par l'ordre des tâches effectuées ont été mises au point. il est demandé à huit couturières de confectionner des chemises en utilisant alteniative ment la méthode **A** puis la méthode **B**. Le tableau 1.7 consigne les temps de réalisations (Il s'agit plus précisément du temps moyen pour confectionner 20 chemises, pour chaque couturière et pour chaque méthode.)

Tableau 1.7 Temps de confection en minutes

Ouvrière	**Méthode A** **Variable X_A**	**Méthode B** **Variable X_B**	**Différence** $D = X_A - X_B$
1	21,0	20,2	0,8
2	20,4	19,7	0,7
3	19,0	19,1	-0,1
4	20,9	19,8	1,1
5	20,3	20,0	0,3
6	20,5	21,0	- 0,5
7	19,5	19,1	0,4
8	19,8	19,1	0,7
	$\bar{X}_A = 20,175$	$\bar{X}_B = 19,750$	$\bar{d} = 0,425$
			$S_D = 0,520$

Le problème est de tester au seuil de 5% , l'égalité des temps de confection moyens. Construirele test bilatéral $\begin{cases} H_0 : & \mu_A = \mu_B \\ H_1 : & \mu_A \neq \mu_B \end{cases}$ revient à construire le test $\begin{cases} H_0 : & \mu_D = 0 \\ H_1 : & \mu_D \neq 0 \end{cases}$ avec $D = X_A - X_B$ et $\mu_D = \mu_A - \mu_B$.
Grâce à l'appariement, le test de comparaison se ramène donc à un test de conformité sur une moyenne. En supposant que la différence D est gaussienne, et compte tenu du fait que sa variance est inconnue, on est conduit à utiliser la statistique d'échantillon (cas 2 tableau 1.5) $T = \dfrac{\bar{D} - \mu_D}{S_D/\sqrt{n}}$.
On sait que cette statistique suit une loi de Student à $(n-1) = 7$ ddl. Le test étant bilatéral, la région critique est de la forme : $\mathcal{C}_r =]-\infty;\ -t_{(1-\alpha/2)}] \cup [t_{(1-\alpha/2)};\ +\infty[$. D'après la table de Student , pour 7 ddl on trouve t $t_{(1-\alpha/2)} = 2,365$. Comme $\mu_D = 0$ sous l'hypothèse nulle, on a $T = \dfrac{\bar{D}}{S_D/\sqrt{n}}$ dont la réalisation est : $t_{obs} = \dfrac{0,425}{0,520/\sqrt{8}} = 2,312$. Cette réalisation n'appartient pas à $\mathcal{C}_r =]-\infty;\ -2,365] \cup [2,365;\ +\infty[$. La différence de temps de confection moyen observée dans l'échantillon (20, 175 pour A et 19, 750 pour B) ne permet donc pas de rejeter l'hypothèse d'égalité des temps moyens. En procédant de la sorte, on ne peut donc pas conclure à la supériorité d'une méthode de travail par rapport à l'autre et on ne connaît pas le risque pris puisque l'on ne rejette pas l'hypothèse nulle.

Dans la pratique, ce genre de comparaison de séquences de confection est rare. Par suite, il est plus rationnel de calculer en premier lieu les réalisations des temps moyens pour chaque méthode, puis de poser un test unilatéral qui tient compte des résultats. Ici, on a $\bar{d} = \bar{A} - \bar{B} = 0,425$. Si une des deux méthodes est meilleure que l'autre ce devrait être la méthode B. D'où le test unilatéral : $\begin{cases} H_0 : & \mu_D = 0 \\ H_1 : & \mu_D > 0 \end{cases}$.

La forme de la région critique est : $[t_{(1-\alpha)};\ +\infty[$. En maintenant un risque de première espèce de 5%, on trouve (table de Student), $t_{95\%} = 1,895$. La réalisation de la variable de décision est toujours, $t_{obs} = 2,312$ mais cette fois cette valeur appartient à la zone de rejet de l'hypothèse nulle. Autrement dit, avec ce test unilatéral (Avec un risque de première espèce de $2,5\%$, le test unilatéral ne permet pas de conclure non plus. En effet le test bilatéral à 5% " correspond " à un test unilatéral à $2,5\%$.), on peut conclure avec un risque d'erreur de 5% que la méthode B est plus rapide que la méthode A.

↪ **Procédure sous SPSS :** Dans l'exemple consacré à l'utilisation d'Internet, une extension du test t pourrait permettre de déterminer si les répondants diffèrent en fonction de leur attitude vis-à-vis de la technologie. La figure 3.25 en fournit les résultats. L'attitude moyenne s'établit à 5,167 vis-à-vis d'Internet et 4,10 vis-à-vis de la technologie. La différence moyenne entre les deux variables s'élève à 1,067, avec un écart-type de 0,828 et une erreur standard de 0,1511. On en déduit que les répondants présentent une attitude plus favorable vis-à-vis d'Internet que vis-à-vis de la technologie en général. Pour obtenir ces résultats à l'aide SPSS, il suffit de suivre les étapes suivantes (voir la figure 3.26).

Statistiques pour échantillons appariés

		Moyenne	N	Ecart-type	Erreur standard moyenne
Paire 1	Attitude vis-à-vis d'Internet	5,17	30	1,234	,225
	Attitude vis-à-vis de la technologie	4,10	30	1,398	,255

Corrélations pour échantillons appariés

		N	Corrélation	Sig.
Paire 1	Attitude vis-à-vis d'Internet & Attitude vis-à-vis de la technologie	30	,809	,000

Test échantillons appariés

		Différences appariées					t	ddl	Sig. (bilatérale)
					Intervalle de confiance 95% de la différence				
		Moyenne	Ecart-type	Erreur standard moyenne	Inférieure	Supérieure			
Paire 1	Attitude vis-à-vis d'Internet - Attitude vis-à-vis de la technologie	1,067	,828	,151	,758	1,376	7,059	29	,000

FIGURE 3.25 –

1. Ouvrez votre fichier de données, et choisissez, dans le menu **Analyse**, **Comparer les moyennes**, puis **Test T pour échantillons appariés...**
2. Déplacez les deux variables (**Attitude-Internet** et **Attitude-Technologie**)dans la zone **variables appariées**.
3. Cliquer sur **Options** si nécessaire pour changer le niveau de confiance des intervalles de confiance, fixé par défaut à 95%. Cliquez alors sur **Poursuivre**, puis sur **OK**

3.2.4 K échantillons indépendants

Variable Nominale

Test de χ^2 d'indépendance

↪ **Principe :** Ce test est déjà présenté dans le cas de deux échantillons

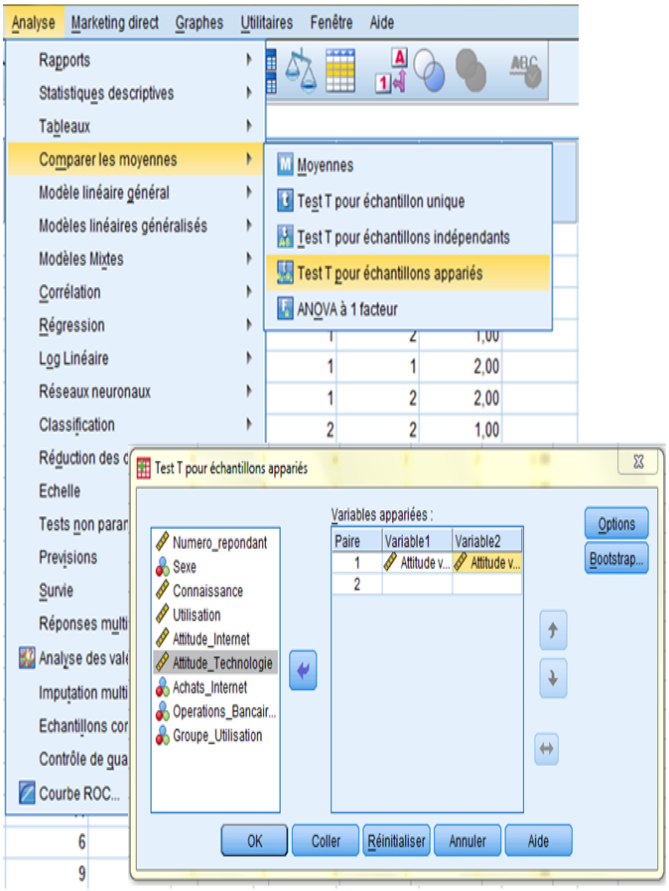


FIGURE 3.26 –

↪ **Exemple d'application :** Une compagnie d'assurances automobile se demande s'il y a indépendance entre **X** : l'âge de l'assuré (X exprimé en nombre d'années) et **Y** : le nombre d'accidents déclarés par le dit assuré au cours de l'année. Pour cela on considère le couple aléatoire $(\mathbf{X};\mathbf{Y})$ où X peut prendre n'importe quelle valeur entre 18 et 95 ans : $\triangle_X = [18\,;95]$ et Y n'importe quelle valeur naturelle entre : $\triangle_Y = \mathrm{N}$. Afin de tester

$$\begin{cases} H_0 : & X \text{ et } Y \text{ indépendants} \quad \text{contre} \\ H_1 : & X \text{ et } Y \quad \text{dépendants} \end{cases}$$

avec un niveau de signification de 5%, on prélève dans le fichier de la compagnie un échantillon de 100 couples de valeurs numériques prises par

$$(X;Y) : (x_1;y_1) = (19;2), (x_2;y_2) = (23;1), ..., (x_{100};y_{100}) = (76;0).$$

↦ **Solution :**

⋆ $\triangle_X$ est partagé en 3 classes ($h = 3$) et $\triangle_Y$ est partagé en 2 classes ($k = 2$) ainsi que l'indique le tableau ci-dessous. Les 100 couples de valeurs sont donc répartis entre "6" classes : c_{ij} présentées ci-dessous.

		$C_{1.} : 18 \leq X < 25$	$C_{2.} : 25 \leq X < 60$	$C_{3.} : X \geq 60$	**Total** $n_{n.j}$
$C_{.1}$: $Y = 0$	**effec obser** n_{ij} **effec théo** n^*_{ij}	$n_{11} = 12$ $n^*_{11} = 14,4$	$n_{21} = 40$ $n^*_{21} = 36$	$n_{31} = 20$ $n^*_{31} = 21,6$	$n_{.1} = 72$
$C_{.2}$: $Y \geq 1$		$n_{12} = 8$ $n^*_{12} = 5,6$	$n_{22} = 10$ $n^*_{22} = 14$	$n_{32} = 10$ $n^*_{32} = 8,4$	$n_{.2} = 28$
	Total : $n_{i.}$	$n_{1.} = 20$	$n_{2.} = 50$	$n_{3.} = 30$	$n = 100$

⋆ Les effectifs théoriques : $n^*_{ij} = \dfrac{n_{i.} \times n_{.j}}{n}$

⋆ $n^*_{11} = \dfrac{n_{1.} \times n_{.1}}{n} = \dfrac{72 \times 20}{100} = 14,4$

⋆ **Variable Statistique Utilisée :** Sous l'hypothèse H_0 on a :

$$Z = \sum_{i=1}^{3}\sum_{j=1}^{2} \frac{(N_{ij} - \dfrac{N_{i.} \times N_{.j}}{n})^2}{\dfrac{N_{i.} \times N_{.j}}{n}} \sim \chi^2_{(3-1)(2-1)} \equiv \chi^2_2$$

$$(n^*_{ij} \geq 5 \ \forall i,j)$$

$$Z \text{ Prend la valeur } z = \sum_{i=1}^{3}\sum_{j=1}^{2} \frac{(n_{ij} - n^*_{ij})^2}{n^*_{ij}}$$

⋆ **Règle de Décision :** Domaine de rejet de H_0 est : $\delta_\alpha = [c_\alpha;\ +\infty[$, avec

$$\mathrm{P}(\chi^2_2 \geq c_\alpha) = \alpha = 0,05 \text{ d'où } c_\alpha = 5,99 \text{ et } \delta_\alpha = [5,99;\ +\infty[$$

⋆ **Décision :**

$$z = \frac{(12-14,4)^2}{14,4} + \frac{(8-5,6)^2}{5,6} + ... + \frac{(10-8,4)^2}{8,4} = 3,439$$

Comme $z = 3,439 \notin \delta_\alpha$ donc on décide H_0 : *X et Y* indépendants.

↪ **Procédure sous SPSS :** Reprendre les données de l'exemple ci-dessous et appliquer la même procédure indiquée pour le test de Khi-deux pour deux échantillon indépendants (voir3.2.2), SPSS affichera les résultats figurants dans la figure 3.27.

Tableau croisé Déclaration * Age

Effectif

		Age			Total
		[18;25[	[25;60[	60 et +	
Déclaration	Y=0	12	40	20	72
	Y>0	8	10	10	28
Total		20	50	30	100

Tests du Khi-deux

	Valeur	ddl	Signification asymptotique (bilatérale)
Khi-deux de Pearson	3,439[a]	2	,179
Rapport de vraisemblance	3,439	2	,179
Association linéaire par linéaire	,064	1	,800
Nombre d'observations valides	100		

a. 0 cellules (,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 5,60.

FIGURE 3.27 –

Variable Ordinale

Test de Kruskal-Walis

↪ **Principe :** [11] Le test est une extension du test de Mann Whitney aux situations comprenant plus de deux échantillons indépendants. Il est aussi appelé test des rangs de Kruskal et Walis. Comme dans le cas de deux échantillons, la réalisation du test est basée sur le classement des observations par ordre croissant, la réalisation du test est basée sur le classement des observations par ordre croissant, la détermination du rang de chacune d'entre elles et le calcul de la somme des rangs relatives aux différents échantillons. Le coefficient H de Kruskal et Walis est alors calculé de la façon suivante, n_j désignant la taille de l'échantillon j :

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(N+1)$$

où
– k est le nombre d'échantillon.
– N : nombre de cas pour l'ensemble des échantillons.
– R_j : somme des rangs du j^{eme} échantillon.
Le coefficient H est distribué selon une loi de Khi-deux à $(k-1)$ degré de liberté lorsque la taille de chaque échantillon est supérieure à 5 . Pour des valeurs inférieures à 5, probabilités ont été tabulées (voir l'annexe)
Le domaine de rejet de H_0 est donné par :

$$\delta_\alpha = [c_\alpha;\ +\infty[\text{ tq } P(H \geq c_\alpha) = \alpha$$

Remarque 8 *On peut réécrire H de la manière suivante :*

$$KW = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i (\bar{R}_i - \frac{N+1}{2})^2$$

avec $\bar{R}_i = R_i/n_i$ (la moyenne des rangs). C'est cette deuxième expression qui sera utilisée dans le solution de l'exemple ci-dessous.

↪ **Exemple d'application :** [14] Un dirigeant de magasin à succursales multiples souhaite connaître l'impact de différents types de promotions envisagées sur le chiffre d'affaire. Concevant 3 types de campagnes de promotion P_1, P_2, P_3 ayant des coûts sensiblement égaux, il assigne à 10 magasins tests ces campagnes de promotion selon la répartition suivante : 3 pour P_1, 3 pour P_2 et 4 pour P_3. Le relevé du taux de croissance du chiffre d'affaires de chacun des 10 magasins pour la période des promotions est présenté ci-dessous, ce taux de croissance δ exprimé en % est calculé par référence au chiffre d'affaires de la période précédente de même durée :

	Taux de croissance δ			
Promotion P_1	2,1	4,0	3,5	
Promotion P_2	4,5	3,6	1,8	
Promotion P_3	2,5	2,2	3,1	3,8

Au vu de ces résultats, il convient de tester l'hypothèse

$$\begin{cases} H_0 : & \text{Les promotions ont la même influence sur le taux } \delta \textit{ d'accroissement}, \textbf{Contre} \\ H_1 : & \end{cases}$$

⋆ k=3, $n_1 = 3$, $n_2 = 3$, $n_3 = 4$, (on a 12 valeurs)
⋆ On range par ordre croissant ces 12 valeurs et on obtient le classement suivant :

P_1	2	6	9		$R_1 = 17$	$\bar{R_1} = 17/3$
P_2	1	7	10		$R_2 = 18$	$\bar{R_2} = 18/3$
P_3	3	4	5	8	$R_3 = 20$	$\bar{R_3} = 20/4$

⋆ La somme des rangs : $R_1 + R_2 + R_3 = \dfrac{n(n+1)}{2} = 55$
⋆ $KW \longrightarrow h^* = \dfrac{12}{10(10+1)}(3(5,66-5,5)^2 + 3(6-5,5))^2 + (5-5,5))^2) = 0,2$
⋆**Règle de Décision :** Domaine de rejet de H_0 est :

$$\delta_\alpha = [c_\alpha,\ +\infty[$$

avec

$$\mathrm{P}(H \geq \omega_\alpha) = \alpha = 5\% \text{ et } c_\alpha = 5,73$$

parsuite

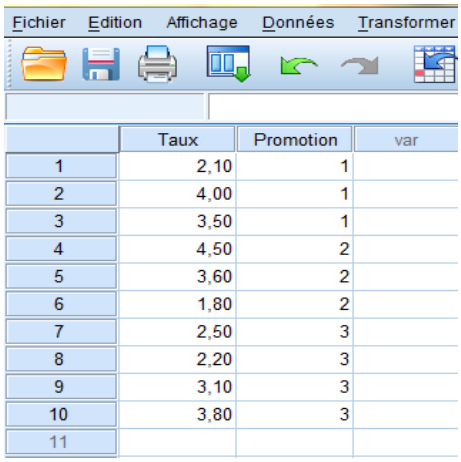
$$\delta_\alpha = [5,73,\ +\infty[$$

⋆ **La Décision :** La valeur h^* prise par H est égale à $0,2 \notin \delta_\alpha$ donc on décide H_0 : (Les différents type de promotions ont le même impact).
⋆ **Table de Kruska Wallis**

$(3,3,4) \longrightarrow$

n_i/α	0,05
433	5,73
	↑ c_α

↪ **Procédure sous SPSS :** Reprendre les données de l'exemple d'application ci-dessus essaye de créer un fichier SPSS pour ces données. Je pense que ce que vous avez obtenu se ressemble à ce que vous voyez dans la figure 3.28.



	Taux	Promotion	var
1	2,10	1	
2	4,00	1	
3	3,50	1	
4	4,50	2	
5	3,60	2	
6	1,80	2	
7	2,50	3	
8	2,20	3	
9	3,10	3	
10	3,80	3	
11			

FIGURE 3.28 –

Pour réaliser le test de Kruskal-Walis il suffit de suivre les étapes suivantes (voir figure3.29). Les résultats de ce test apparaissent dans la figure3.30

1. Cliquer sur **Analyse**, **Tests non paramétriques**'puis sur **K échantillons indépendants**.

2. Glisser la variable **Taux** dans la case **Liste des variables à tester**
3. Glissez la variable **Promotion** dans la case **Critère de regroupement** et cliquer sur **Définir Intervalle** pour préciser l'intervalle des codes des échantillons (1 et 3)
4. Enfin choisir le test **H de Kruskal-Walis** dans la liste **Type de test** puis cliquer sur **OK**

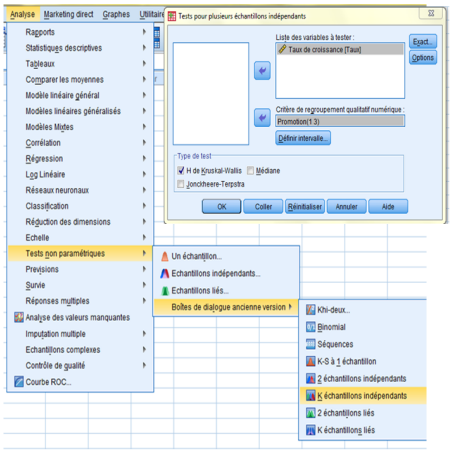


FIGURE 3.29 –

Rangs

	Promotion	N	Rang moyen
Taux de croissance	P1	3	5,67
	P2	3	6,00
	P3	4	5,00
	Total	10	

Test[a,b]

	Taux de croissance
Khi-deux	,200
ddl	2
Signification asymptotique	,905

a. Test de Kruskal Wallis

b. Critère de regroupement : Promotion

FIGURE 3.30 –

Variable de type :Intervalle ou Ratio

Analyse de la variance

Cette technique sera étudiée en détaille dans le chapitre quatre

3.2.5 K échantillons dépendants (appariés)

Variable Nominale

Test de Cochran Q

↪ **Principe :** [7] Ce test est une généralisation du test de Mc Nemar (pour deux échantillons appariés). Il consiste à tester si les échantillons appariés présentent des fréquences différentes.

Chacun des k traitements est appliqué indépendamment à chacun des n individus. Le résultat d'un traitement est "**Succès**" ou "**Echec**" c'est à dire des variables binaires (du type Oui (1) Non (0)). Les données peuvent donc se présenter sous la forme d'un tableau à n lignes et k colonnes rempli de 1 (succès) ou 0 (échec). On test H_0 : les traitements ont la même efficacité contre H_1 : il existe une différence d'efficacité entre les traitement. La statistique du test est :

$$Q = \frac{\sum_{j=1}^{k} k(k-1)(G_j - \frac{N}{k})^2}{\sum_{i=1}^{n} L_i(k - L_i)}$$

avec G_j nombre des 1 dans la colonnes j , L_i nombre des 1 dans la ligne i, et N nombre des 1 dans la matrice $(n \times k)$ en d'autre terme $N = \sum_{j=1}^{k} Gj = \sum_{i=1}^{n} L_i$. Si n est suffisamment grand, Q suit une loi de Khi-deux à $(k-1)$ degré de liberté.
La règle de décision est :

Si $Q > \chi^2_{1-\alpha}$	on rejette H_0
$Q \leq \chi^2_{1-\alpha}$	on n'a pas de raison de rejeter H_0

↪ **Exemple d'application :**[11] Une société commercialise son produit accompagné d'un cadeau. Elle considère que le recours au cadeau est un succès lorsque les ventes de la période avec cadeau sont de 25% supérieures à celle de la période sans cadeau. Elle dispose d'un relevé des succès et des échecs sur 15 zones géographiques différentes et pour 3 périodes temporelles : une période dite normale au cours de laquelle les prix du marché sont restés stables (période 1) ; une période de guerre des prix (période 2) et une période de retour à la normale (période 3). Le code 1 est affecté aux succès enregistrés et le code 0 en cas d'échecs. La société souhaite savoir si l'efficacité du cadeau est ou non affectée par la guerre des prix entre les marques concurrentes.

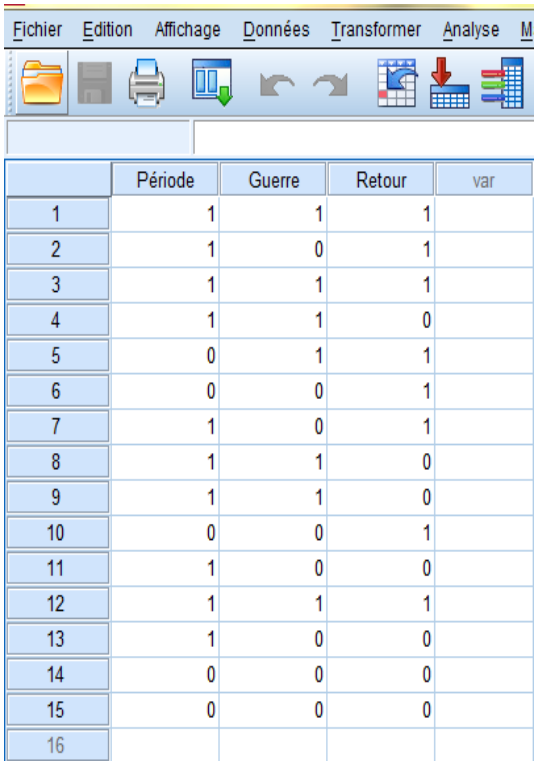
Zone géographique	**Période normale**	**Période guerre des prix**	**Période retour à la normale**	**Total**
1	1	1	1	3
2	1	0	1	2
3	1	1	1	3
4	1	1	0	2
5	0	1	1	2
6	0	0	1	1
7	1	0	1	2
8	1	1	0	2
9	1	1	0	2
10	0	0	1	1
11	1	0	0	1
12	1	1	1	3
13	1	0	0	1
14	0	0	0	0
15	0	0	0	0
	10 succès	**7 succès**	**8 succès**	**25 succès**

Le calcule de la valeur du Q de cochran nous donne donne la valeur suivante :

$$Q = \frac{3(3-1)(10 - \frac{25}{3})^2 + (7 - \frac{25}{3}) + (8 - \frac{25}{3})}{3(3-3) + 2(3-2) + ... + 0(3-0) + 0(3-0)} = 1,4$$

La valeur de Q étant inférieure à celle de $\chi^2_{0,95} = 5,99$, l'hypothèse nulle est acceptée. Il n'existe pas de différence entre les trois périodes. L'efficacité des cadeaux n'est pas affecté par la guerre des prix.

↪ **Procédure sous SPSS :** Reprenons les données de notre d'exemple d'application ci-dessus. La figure3.31 vous encourage à créer un fichier SPSS pour ces données.



	Période	Guerre	Retour	var
1	1	1	1	
2	1	0	1	
3	1	1	1	
4	1	1	0	
5	0	1	1	
6	0	0	1	
7	1	0	1	
8	1	1	0	
9	1	1	0	
10	0	0	1	
11	1	0	0	
12	1	1	1	
13	1	0	0	
14	0	0	0	
15	0	0	0	
16				

FIGURE 3.31 –

Pour la réalisation du test de Cochran il suffit de suivre les étapes suivantes(voir figure3.32) :

1. Cliquer sur **Analyse**, **Tests non paramétriques**'puis sur **K échantillons liés**.
2. Glisser les trois variables dans la case **Variables à tester**
3. Enfin choisir le test **Q de de Cochran** dans la liste **Type de test** puis cliquer sur **OK**

Les résultats du test apparaissent dans la figure 3.33. Le test montre une valeur de $\chi^2_{obs} = 1,4$ avec un $sig = 00,497 > 0,05$ d'ou le non rejet de H_0 : L'efficacité des cadeaux n'est pas affecté par la guerre des prix

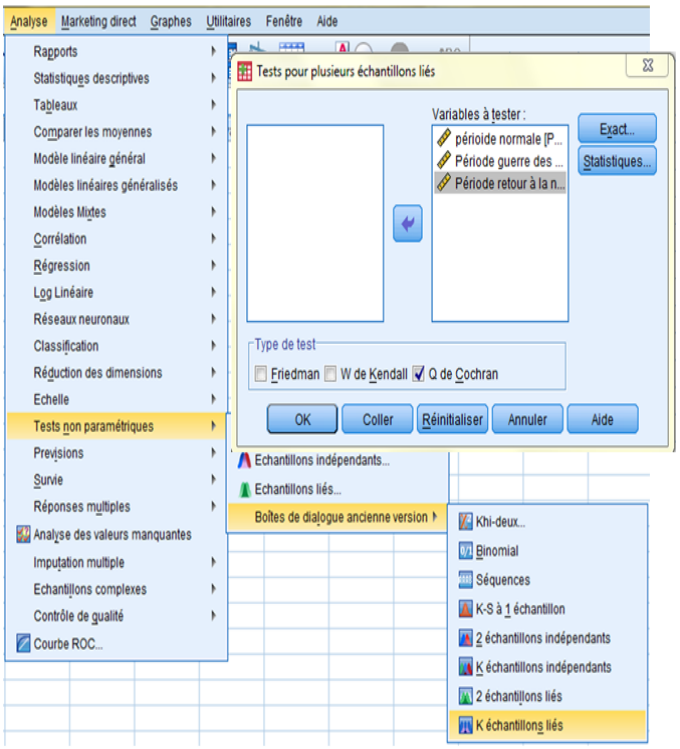


FIGURE 3.32 –

Fréquences

	Valeur	
	0	1
périoide normale	5	10
Période guerre des prix	8	7
Période retour à la normale	7	8

Test

N	15
Q de Cochran	1,400[a]
ddl	2
Signification asymptotique	,497

a. 1 est traité comme succès.

FIGURE 3.33 –

Variable Ordinale

Test de Friedman

↪ **Principe :** [7] Le test des des rangs de Friedman [8] pour k échantillons non indépendants permet de tester les mêmes hypothèses que le test du coefficient Q de Cochran, mais il s'agit de données ordinales. Il teste l'hypothèse nulle stipulant que les scores correspondant à chaque traitement proviennent de populations identiques et est particulièrement sensibles aux différences de tendance centrale au niveau des populations. Les données peuvent se présenter de façon suivante :

Traitements (attitudes vis-à-vis des produits 1 à k)

Blocs consommateurs	1	2	.....	.....	k
1	X_{11}	X_{12}	.....	.....	X_{1k}
2	X_{21}	X_{22}	.....	.....	X_{2k}
.			.....	.....	
.			.....	.....	
n	X_{n1}	X_{n2}			X_{nk}

Légende : n=nombre de lignes (c'est-à-dire de sujets).
k=nombre de colonnes (c'est-à-dire de traitements).
X_{ij}=résultats à l'intérieur d'un bloc (consommateur) : par exemple, évaluation d'un produit ou d'une publicité ; avec i allant de 1 à n, et j allant de 1 à k

TABLE 3.7 – Données d'attitude. Sourse :[7] page 387

Les données utilisées dans le test sont des rangs. Il va falloir transformer les données de départ en rangs pour chaque ligne séparément. Le coefficient de χ^2_r de Friedman est alors calculée de la façon suivante :

$$\chi^2_r = \frac{12}{nk(k+1)} \sum_{j=1}^{k} (R_j)^2 - 3n(k+1)$$

avec R_j somme des rangs de la j^{eme} colonne.
Le coefficient χ^2_r est distribué selon un χ^2 avec $k-1$ degré de liberté. Lorsque χ^2_r est inférieur à la valeur du $\chi^2_{1-\alpha}$ l'hypothèse nulle d'égalité des traitements est vérifiée.

↪ **Exemple d'application :** [11]

Supposons qu'un fabricant analyse sur trois années la satisfaction de six clients très importants. Ceux-ci ont jugé leur satisfaction sur un ensemble d'échelles de Likert. Plus les scores sont élevés plus le client est satisfait. Le fabricant utilise les scores globaux obtenus par client sous forme ordinale. Les scores obtenus sont les suivants :

Le calcule du χ^2_r de Friedman nous donne la valeur suivante :

$$\chi^2_r = (12/(6 \times 3 \times 4))[15^2 + 12^2 + 9^2] - (3 \times 6 \times 4) = 2,99$$

8. On le doit au célèbre économiste Milton Friedman, qui l'a conçu avant d'acquérir sa notoriété en économie

	Scores bruts			Rangs		
Client	**Année 1**	**Année 2**	**Année 3**	**Année 1**	**Année 2**	**Année 3**
1	82	76	83	2	3	1
2	32	39	42	3	2	1
3	44	48	46	3	1	2
4	49	50	55	3	2	1
5	73	64	60	1	2	3
6	20	25	31	3	2	1
				Total = 15	**Total= 12**	**Total =9**

TABLE 3.8 –

La table de Friedman (voir l'annexe) indique que pour cette valeur la probabilité d'observation est égale à 0,252. Celle-ci étant supérieure à 0,05, l'hypothèse d'égalité entre les conditions est acceptée. Il n'existe pas de différences de satisfaction entre les trois périodes analysées et pour chacun des 6 clients.

↪ **Procédure sous SPSS :**[4] Dans le tableau qui suit, vous avez les données fournies par des sujets qui ont estimé à quel point ils se sentaient en forme, à trois moments de la journée (sur une échelle allant de 0 à 5).

Participant	**MATIN**	**MIDI**	**SOIR**
1	1	2	3
2	2	4	5
3	1	2	2
4	2	1	3
5	1	3	3
6	2	5	5
7	1	2	3
8	2	2	2
9	1	3	3
10	2	1	3

Pour la réalisation du test de Friedman il suffit de suivre les étapes suivantes (voir figure3.34.

1. Sélectionner **Analyse**, **Tests non paramétriques** puis **K échantillons liés**
2. Déplacez les variables à tester dans la zone**Variables à tester**
3. Prenez soins de bien cocher l'option **Friedman**
4. Si vous avez envie de visualiser des statistiques descriptives vous pouvez passer par l'option **Statistiques**. Cliquez sur **OK**. et voilà les résultats dans la figure 3.35.

Le χ^2 vaut 12,25 et correspond à une signification de $p = 0,002$. La différence trouvée, pour une même personne, entre les différents moments de la journée, est significative.

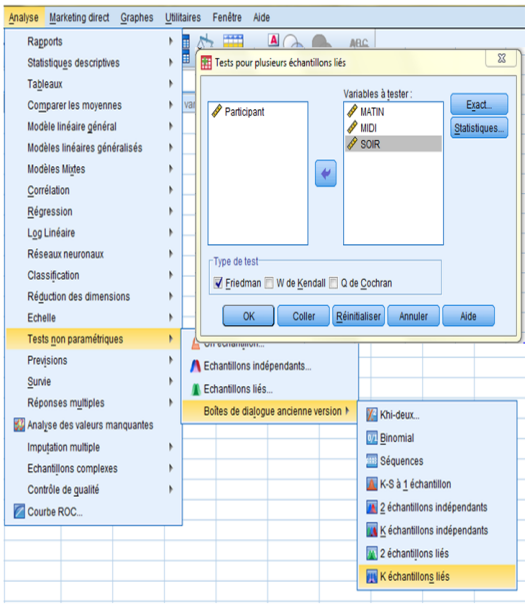


FIGURE 3.34 –

Rangs

	Rang moyen
MATIN	1,30
MIDI	2,00
SOIR	2,70

Test[a]

N	10
Khi-deux	12,250
ddl	2
Signification asymptotique	,002

a. Test de Friedman

FIGURE 3.35 –

Variable de type :Intervalle ou Ratio

Analyse de la variance à mesures répétées

Cette technique sera discutée dans le chapitre quatre intitulé **Analyse de la variance**

3.3 Travaux pratiques

TP3

Exercice 1 *Dans le cadre d'un pré-test, certaines données relatives à la marque Nike ont été recueillies auprès de 45 clients. Ces données figurent dans le tableau ci-dessous (Créez un fichier SPSS pour ces données et nommez le* **Nike.sav***). Elles concernent le niveau d'utilisation, le sexe, la notoriété, l'attitude, la préférence, l'intention et la fidélité vis-à-vis de la marque. Le niveau d'utilisation a été codé 1, 2 ou 3, selon qu'il était faible, moyen, ou important. Le sexe a été codé 1, pour les femmes, et 2 pour les hommes. La notoriété, l'attitude, la préférence, l'intention et la fidélité ont été mesurées sur une échelles de type Likert en sept points (1=très défavorable, 7= très favorable). On notera que cinq répondants présentent des valeurs manquantes, notées 9. Analyser les données et répondez aux questions suivantes. Dans chaque cas, formulez les hypothèses nulle et alternative, et réalisez le ou les tests statistiques appropriés.*

Numéro	Utilisation	Sexe	Notoriété	Attitude	Préférence	Intention	Fidélité
1	3	2	7	6	5	5	6
2	1	1	2	2	4	6	5
3	1	1	3	3	6	7	6
4	3	2	6	5	5	3	2
5	3	2	5	4	7	4	3
6	2	2	4	3	5	2	3
7	2	1	5	4	4	3	2
8	1	1	2	1	3	4	5
9	2	2	4	4	3	6	5
10	1	1	3	1	2	4	5
11	3	2	6	7	6	4	5
12	3	2	6	5	6	4	4
13	1	1	4	3	3	3	3
14	3	2	6	4	5	3	2
15	1	2	4	3	4	5	6
16	1	2	3	4	2	4	2
17	3	1	7	6	4	5	3
18	2	1	6	5	4	3	2
19	1	1	1	1	3	4	5
20	3	1	5	7	4	1	2
21	3	2	6	6	7	7	5
22	2	2	2	3	1	4	2
23	1	1	1	1	3	2	2
24	3	1	6	7	6	7	6
25	1	2	3	2	2	1	1
26	2	2	5	3	4	4	5
27	3	2	7	6	6	5	7
28	2	1	6	4	2	5	6
29	1	1	9	2	3	1	3
30	2	2	5	9	4	6	5

TABLE 3.9 – Données relatives à la marque Nike

Numéro	Utilisation	Sexe	Notoriété	Attitude	Préférence	Intention	Fidélité
31	1	2	1	2	9	3	2
32	1	2	4	6	5	9	3
33	2	1	3	4	3	2	9
34	2	1	4	6	5	7	6
35	3	1	5	7	7	3	3
36	3	1	6	5	7	3	4
37	3	2	6	7	5	3	4
38	3	2	5	6	4	3	2
39	3	2	7	7	6	3	4
40	1	1	4	3	4	6	5
41	1	1	2	3	4	5	6
42	1	1	1	3	2	3	4
43	1	1	2	4	3	6	7
44	1	1	3	3	4	6	5
45	1	1	1	1	4	5	3

TABLE 3.10 – Données relatives la marque Nike(suite)

Sourse :[12] page 410

- *Établissez la distribution de fréquences de chacune des variables suivantes et calculez les statistiques pertinentes : notoriété, attitude, préférence, intention, et fidélité de la marque.*
- *Effectuer un tri croisé de l'utilisation et du sexe. Interprétez les résultats.*
- *La notoriété de la marque Nike dépasse-elle 3,0 ?*
- *Le niveau de la notoriété de Nike varie-t-il en fonction du sexe des clients ? Les hommes et les femmes expriment-ils une attitude différente vis-à-vis de la marque ? Présentent-ils des différences en termes de fidélité ?*
- *Chez les répondants du pré-test, la notoriété de la marque est-elle supérieure à la fidélité ?*
- *La notoriété de Nike suit-elle une distribution normale ?*
- *La préférence pour Nike suit-elle une distribution normale ?*
- *Supposons que la notoriété de Nike soit mesurée sur une échelle ordinale plutôt que sur une échelle d'intervalles. Cette donnée varie-elle en fonction du sexe des clients ?*
- *Supposons que la fidélité à l'égard de Nike soit mesurée sur une échelle ordinale plutôt que sur une échelle d'intervalles. Cette donnée varie-elle en fonction du sexe des clients ?*
- *Supposons que l'attitude et la fidélité vis-à-vis de Nike soient mesurées sur une échelle ordinale plutôt que sur une échelle d'intervalles. Pour les personnes interrogées, la notoriété de la marque est-elle plus élevée que leur fidélité ?*

Exercice 2 *Lors d'un pré-test, on a demandé aux répondants d'exprimer leur attirance pour un mode de vie au grand air (V1), en utilisant une échelle en sept points (1= pas du tout attiré ; 7= très attiré). On leur a également demandé d'indiquer l'importance des variables suivantes sur une échelle en sept points (1= pas du tout important ; 7= très important) :*

- *V2= profiter de la nature*
- *V3 savoir le temps qu'il fait*
- *V4= vivre en harmonie avec l'environnement*
- *V5= faire régulièrement de l'exercice*
- *V6= rencontrer d'autres personnes*

Le sexe des répondants (V7) a été codé 1 pour les femmes et 2 pour les hommes, le lieu de résidence (V8) a été codé 1 pour la ville, 2 pour la banlieue et 3 pour la campagne. On a obtenu les données suivantes (voir tableau ci-dessous) Sourse :[12] page 413 Répondez aux questions

V1	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**V7**
7	3	6	4	5	2	1	1
1	1	1	2	1	2	1	1
6	2	5	4	4	5	1	1
4	3	4	6	3	2	1	1
1	2	2	3	1	2	1	1
6	3	5	4	6	2	1	1
5	3	4	3	4	5	1	1
6	4	5	4	5	1	1	1
3	3	2	2	2	2	1	1
2	4	2	6	2	2	1	1
6	4	5	3	5	5	1	2
2	3	1	4	2	1	1	2
7	2	6	4	5	6	1	2
4	6	4	5	3	3	1	2
1	3	1	2	1	4	1	2
6	6	6	3	4	5	2	2
5	5	6	4	4	6	2	2
7	7	4	4	7	7	2	2
2	6	3	7	4	3	2	2
3	7	3	6	4	4	2	2
1	5	2	6	3	3	2	3
5	6	4	7	5	6	2	3
2	4	1	5	4	4	2	3
4	7	4	7	4	6	2	3
6	7	4	2	1	7	2	3
3	6	4	6	4	4	2	3
4	7	7	4	2	5	2	3
3	7	2	6	4	3	2	3
4	6	3	7	2	7	2	3
5	6	2	6	7	2	2	3

TABLE 3.11 –

suivantes en vous aidant du logiciel SPSS. Dans chaque cas, formulez les hypothèses nulle et alternative et réalisez le ou les tests statistiques appropriés.

1. *L'attirance moyenne pour pour un mode de vie au grand air dépasse-t-elle 3,0 ?*

2. *L'importance moyenne attaché au fait de profiter de la nature dépasse-elle 3,0 ?*
3. *L'attirance moyenne pour pour un mode de vie au grand air varie-t-elle en fonction du sexe ?*
4. *L'importance des variables V2 à V6 est-elle fonction du sexe ?*
5. *Les répondants trouvent-ils plus important de profiter de la nature que de savoir le temps qu'il fait ?*
6. *Les répondants trouvent-ils plus important de savoir le temps qu'il fait que de rencontrer d'autre personnes ?*
7. *Les répondants trouvent-ils plus important de vivre en harmonie avec l'environnement que de faire régulièrement de l'exercice ?*
8. *L'importance des variables V2 à V6 est-elle fonction du sexe lorsqu'elles sont mesurées sur une échelle ordinale plutôt que sur une échelle d'intervalles ?*
9. *Les répondants trouvent-ils plus important de savoir le temps qu'il fait que de rencontrer d'autres personnes, lorsque ces deux variables sont mesurées sur une échelle ordinale plutôt que sur une échelle d'intervalles ?*

Chapitre 4

Analyse de la variance (ANOVA)

Introduction

Le pilotage d'unités de production, l'optimisation de la productivité, l'amélioration de la qualité des produits,..., sont autant de situations qui demandent de disposer de leviers d'action. Pour identifier ces leviers d'action, il est nécessaire de déterminer les facteurs qui influencent les processus en jeu. Ces processus sont rarement modélisables dans toute leur complexité. Néanmoins le problème d'identification de facteurs explicatifs peut souvent être simplifié : on recherche des variables qualitatives qui ont un impact sur les moyennes saisies sur plus de deux populations d'une variable statistique[17].

L'ANOVA à un facteur permet d'effectuer un test sur les moyennes de deux populations ou plus. L'ANOVA généralise ainsi les tests de comparaisons de moyenne examinés au chapitre précédent.

L'ANOVA à plusieus facteurs implique l'examen simultané de plusieurs variables indépendantes qualitatives. Elle permet l'évaluation de l'interaction de ces variables. Le test F sert à vérifier la signification de l'effet global, des effets principaux et des interactions. Il y a interaction lorsque l'effet d'une variable indépendante sur une variable dépendante diffère en fonction des modalités ou niveaux d'une autre variable indépendante.

L'ANCOVA (Analyse de Covariance) fait référence, en plus de variable(s) indépendante(s) qualitative(s), au test de variable(s) indépendante(s) quantitative(s). Cette dernière, appelée covariable, est souvent utilisée pour éliminer la variation externe de la variable dépendante.

Paradoxalement, ce sont des techniques d'analyse de la variance (ANOVA est l'acronyme de Analysis of variance), qui permettent de traiter ces nouveaux problèmes de comparaisons de moyennes (voir repère 10 pour plus de détailles).

Repère 10 (Comment comparer des moyennes à l'aide de l'analyse de la variance ?), *Source :[7]*
Il peut paraitre étonnant d'appeler analyse de variance une technique destinée à comparer des moyennes. Cependant, c'est réellement une comparaison de variances qui est effectuer. En effet, pour tester l'hypothèse d'égalité des moyennes de p populations, on doit prélever un échantillon aléatoire dans chaque population. Les moyennes des échantillons et la moyenne générale des observations permettent de définir deux types de variations : variation entre échantillons (ou variation factorielle) et variation résiduelle. L'importance de ces deux sources de variation est mesurée par deux quantités qui sont appelées carrés moyens ou **variances** *: ce sont les* **carrés moyens factoriels** *et les* **carrés moyens résiduels** *. Lorsqu'il existe des différences importantes entre les moyennes des populations, on doit s'attendre à ce qu'il en soit de même pour les échantillons. On doit avoir un carré moyen factoriel élevé par comparaison avec le carré moyen résiduel. Le rapport du carré moyen factoriel au carré moyen résiduel (rapport de variance) permet ainsi d'examiner l'égalité des moyennes.*

Notre démarche est basée sur la stratégie suivante :

· **Principe :** justifiant l'utilité et si c'est nécessaire les procédures décrivant les étapes de la technique.

· **Exemple d'application :** met en évidence l'application concrète de la technique dans le domaine du marketing.

· **Procédure sous SPSS :** nous facilite la réalisation de la technique sous SPSS.

Pour le dernier point de cette stratégie, on va utiliser les données du tableau 4.1,

Exemple 5 *[12] Une expérience visant à évaluer l'effet commercial des promotions sur le lieu de vente et des bons d'achats dans les divers établissements d'une grande chaine d'hypermarchés. Trois niveau de promotion ont été définis : élevé(1), moyen(2) et faible(3). Le couponnage a été noté sur deux niveaux, selon que les clients potentiels recevaient un bon d'achat de 20 euros(1) ou ne le recevaient pas(2). Ces deux variables ont été croisées pour obtenir un modèle* 3×2 *à six cellules. On a choisi 30 magasins au hasard, et cinq d'entre eux ont été aléatoirement assignés à chaque traitement, comme le montre le tableau 4.1. L'expérience a duré deux mois. Dans chaque magasin, les ventes ont été évaluées, normalisées en fonction des facteurs externes (taille du point de vente, fréquentation, ect.) et converties sur une échelle de 1 à 10. En complément, le niveau de vie relatif de la clientèle à fait l'objet d'une évaluation qualitative, mesurée elle aussi sur une échelle de 1 à 10. Les chiffres les plus élevés traduisent respectivement les ventes plus importantes et le meilleur niveau de vie.*

Essayer de créer un fichier SPSS pour ces données, nommer le **Promotion Magasin**

4.1 Analyse de variance à un facteur

↪ **Principe :** Les responsables d'études marketing s'intéressent souvent aux variations des valeurs moyennes de la variable dépendante en fonction des modalités (niveaux) d'une seule variable ou d'un seul facteur indépendant. Par exemple
- Les divers milieux socioprofessionnelles présentent-ils des différences en terme de consommation ?
- L'évaluation d'une marque varie-t-elle en fonction des publicités auxquelles les différents groupes de répondants se trouvent exposés ?
- Détaillants, grossistes, et agents témoignent-ils d'une attitude différente vis-à-vis de la politique de distribution de l'entreprise ?

N^o **du Magasin**	**Couponnage**	**Promotions**	**Ventes**	**Classement Clientèle**
1	1	1	10	9
2	1	1	9	10
3	1	1	10	8
4	1	1	8	4
5	1	1	9	6
6	1	2	8	8
7	1	2	8	4
8	1	2	7	10
9	1	2	9	6
10	1	2	6	9
11	1	3	5	8
12	1	3	7	9
13	1	3	6	6
14	1	3	4	10
15	1	3	5	4
16	2	1	8	10
17	2	1	9	6
18	2	1	7	8
19	2	1	7	4
20	2	1	6	9
21	2	2	4	6
22	2	2	5	8
23	2	2	5	10
24	2	2	6	4
25	2	2	4	9
26	2	3	2	4
27	2	3	3	6
28	2	3	2	10
29	2	3	1	9
30	2	3	2	8

TABLE 4.1 – Couponnage, promotions, ventes et classement de la clientèle

– Comment les intentions d'achats des consommateurs varient-elles en fonction du prix ?
– La connaissance qu'un consommateur a d'un magasin (élevée, moyenne, ou faible) exerce t-elle une influence sur ses préférences ?

Les réponses à ce genre de question peuvent être obtenues au moyen d'une ANOVA à un facteur[12].

L'analyse de la variance utilise un vocabulaire spécifique : les variables qualitatives susceptibles d'influencer sur la distribution de la variable quantitative étudiée sont appelées facteurs (ou facteurs de variabilité) et leurs modalités niveaux ou catégories. Dans ce chapitre, les facteurs sont toujours contrôlés, c'est-à-dire fixés par l'expérimentateur.

Hypothèses, notations et conditions d'application

L'analyse de la variance fait intervenir une variable quantitative mesurée sur plusieurs populations. Chaque population correspond à un niveau (une modalité) du facteur explicatif

envisagé.

Les notations des concepts nécessaires sont les suivantes :

- k variables aléatoires parentes $X_i (i = 1 \text{ à } k)$ d'espérance μ_i et d'écart-type σ_i. Les variables aléatoires X_i sont définies dans les mêmes termes mais sont mesurées sur k populations P_k. Par exemple, X_i est le temps d'assemblage dans l'usine A, X_2 est le temps d'assemblage dans l'usine B, etc.
- On tire un échantillon de taille n_i dans chaque population P_k. Ainsi, $(X_{i1}, X_{i2}, ..., X_{in_i})$ est un échantillon issu de X_i et X_{ij} est la $j^{\text{ème}}$ variable aléatoire issue de X_i $(j = 1 \text{ à } n_i)$.
- L'effectif total des échantillons : $n = \sum_{i=1}^{k} n_i$.

On teste : $\begin{cases} H_0 : & \mu_1 = ... = \mu_k \\ H_1 : & \text{au moins deux des espérances sont différentes} \end{cases}$

Le test repose sur trois hypothèses :

- Les variances sont toutes égales, autrement dit pour tout $i : \sigma_i^2 = \sigma^2$.
- Les n variables aléatoires X_{ij} sont indépendantes.
- Les variables aléatoires parentes X_i sont gaussiennes.

En résumé, les variable aléatoires sont indépendantes et pour tout i : $X_i \sim \mathcal{N}(\mu_i;\ \sigma)$.

Il convient de remarquer que si l' hypothèse nulle est retenue à l'issue du test, on considère que la variable qualitative (le facteur) définissant les populations n'a pas d'influence sur la variable quantitative, donc que l' on a une seule variable aléatoire parente et que toutes les observations sont issues de la même population. Les développements suivants vont permettre de réaliser ce test.

Principe de la méthode

La variance commune σ^2 joue un rôle fondamental car la méthode de L'ANOVA utilise deux estimations de cette variance. Les deux estimateurs correspondants sont construits ci-après. Le premier repose sur la variabilité des observations à l'intérieur de chaque échantillon, le second mesure la variabilité des moyennes entre les échantillon.

- **Estimation de la variance commune par la variance intra-échantillon**

  L'idée est d'utiliser l'information provenant des k échantillons pour construire un estimateur commun de σ^2. Cet estimateur commun n'est autre que la variance de pool (ou variance combinée) généralisée à plus de deux échantillons :

$$VAR_{intra} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + ... + (n_k - 1)S_k^2}{(n_1 - 1) + (n_2 - 1) + ... + (n_k - 1)} = \frac{1}{n - k} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

  Il est fondamental de noter que la variance intra-échantillon est un estimateur non biaisé de la variance commune σ^2 même si l'hypothèse nulle d'égalité des espérances est fausse.

- **Estimation de la variance commune par la variance inter-échantillon**

  Il s'agit à présent de construire un deuxième estimateur de la variance commune σ^2 en s'appuyant cette fois sur la variabilité observée entre les échantillons.

$$VAR_{inter} = \frac{1}{k - 1} \sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2$$

  $\bar{X} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} X_{ij}$ (n=$n_1 + n_2 + ... + n_k$) (La moyenne générale).

Sous H_0 la variance inter-échantillon est un estimateur non biaisé de la variance commune σ^2 des populations.

Nous avons montré que si $H_0 : \mu_1 = \mu_2 = \mu_3 = ... = \mu_k$ est vraie, la variance commune σ^2 peut être estimée sans biais à la fois par VAR_{intra} et par VAR_{inter}.Ainsi, si H_0 est vraie, le rapport VAR_{intra}/VAR_{inter} de ces deux estimateurs doit être proche de 1.

Au contraire, si H_0 n'est pas vraie, seule VAR_{intra} est un estimateur sans biais de σ^2 En outre, si l'hypothèse nulle n'est pas vérifiée, la variance inter-échantillon VAR_{inter} a tendance à surestimer la valeur de σ^2. Dans ce cas, le rapport entre les deux estimateurs VAR_{intra}/VAR_{inter} doit dépasser la valeur 1 . La variable de décision est ainsi VAR_{intra}/VAR_{inter}. Pour construire le test de L'ANOVA il est donc nécessaire de connaître la loi de cette variable.

- **Variable et règle de décision : le tableau ANOVA.**
  - **Variable de décision**

$$\frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2}{\frac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2} = \frac{VAR_{inter}}{VAR_{intra}} \sim \mathcal{F}_{(n-k)}^{(k-1)} \text{ (Fisher-Snédécor)}$$

  - **Région critique**

$$\mathcal{C}_r = [f_{(1-\alpha)};\ +\infty[$$

tel que $f_{(1-\alpha)}$ est le quantile d'ordre $(1-\alpha)$ de la loi de Fisher $(\mathcal{F}_{(n-k)}^{(k-1)})$

↪ **Exemple d'application : Répartition des compétences commerciales**.

La direction des ventes d'un important groupe spécialisé dans la distribution de matériel bureautique auprès des entreprises étudie la répartition stratégique de sa force commerciale sur le territoire national. La zone commerciale est divisée en sept unités régionales : A, B, C, D, E ,F,G. Ce découpage a été effectué il y a une dizaine d'années et, à présent, la direction souhaite vérifier que les potentialités commerciales de chaque territoire sont correctement exploitées. Pour effectuer cette étude, le groupe rassemble les informations portant sur les caractéristiques des zones géographiques et de la force de vente qui leur est affectée.

Le potentiel de chaque région était assez différent au moment du découpage, c'est pourquoi le nombre total de commerciaux diffère d'une région à l'autre (la région A qui est la plus "petite" compte 65 commerciaux). En outre, les commerciaux de groupe ont différents statuts qui correspondent à des objectifs mensuels de chiffre d'affaires. Les délégués commerciaux, qui sont les plus nombreux, ont un objectif mensuel moyen de l'ordre de 30 K euro (30 000 euros). Les ingénieurs commerciaux doivent, quand à eux, obtenir mensuellement des contrats pour 45 K euros, et les ingénieurs commerciaux "grand comptes", les moins nombreux, ont un objectif de 90 K euros.

La direction commerciale pense que le chiffre d'affaires annuel dépend de la taille de chaque région et qu'il est possible que la répartition des status des commerciaux dans chaque région soit également un facteur de réussite déterminant. Pour tester ces hypothèses, l'équipe d'administration des ventes a été chargée de rassembler le chiffre d'affaires généré par un échantillon de la force de vente de chacune des sept régions commerciales. Chaque échantillon tient compte de la taille relative de chaque région et de la répartition des status des commerciaux. Les données de chiffres d'affaires annuels des commerciaux représentatifs de chaque région sont présentées dans le tableau ci-dessous :

A	B	C	D	E	F	G
980,2	945,2	1400	915,3	967	654	1312,3
$\underbrace{1003,1}$	1200	767,6	1390	674,9	1080	1492,4
430	331,8	470,4	421,1	444	555	$\underbrace{1033,2}$
$\underbrace{395}$	45,6	356,5	234,2	557,2	271,2	572,7
550,5	452,7	493	632,2	487,2	650	$\underbrace{430,5}$
$\underbrace{600,4}$	326	576	526	700	565	1000,8
			800			526,1
			$\underbrace{733,3}$			611,6
						483,8
						$\underbrace{501,4}$
$\bar{X}_1 = 659,86$	$\bar{X}_2 = 550,21$	$\bar{X}_3 = 677,65$	$\bar{X}_4 = 706,51$	$\bar{X}_5 = 638,38$	$\bar{X}_6 = 629,2$	$\bar{X}_7 = 796,48$

Chiffres d'affaires annuels des sept échantillons de commerciaux (en K euro)

Dans ce tableau, les chiffre d'affaires réalisés ont été reportés du haut (pour les ingénieurs commerciaux "grands comptes") au bas (pour les délégués commerciaux).

Il s'agit d'analyser les chiffres d'affaires moyens dans ces échantillons et d'en tirer des conclusions sur la répartition des compétences commerciales dans les sept régions prédéfinies par le groupe (Comparer les chiffres d'affaires moyens entre régions).

La direction commerciale désire savoir si les chiffres d'affaires moyens des sept régions prédéfinies par le groupe sont différents. Le seuil du test à réaliser est fixé à 5%

1. **Les hypothèses :**

$$\begin{cases} H_0 : & \mu_1 = \mu_2 = ... = \mu_7 \text{ contre} \\ H_1 : & \exists (i,j)\ tq\ \mu_i \neq \mu_j \end{cases}$$

2. **Variable de décision :**

$$T = \frac{\frac{1}{7-1}\sum_{i=1}^{7} n_i(\bar{X}_i - \bar{X})^2}{\frac{1}{48-7}\sum_{i=1}^{7}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2} \sim \mathcal{F}_{(48-7)}^{(7-1)} \text{ (Fisher-Snédécor)}$$

3. **Région critique :**

$$\mathcal{C}_r = [f_{95\%};\ +\infty[$$

tel que $f_{95\%}$ est le quantile d'ordre 95% de la loi de Fisher ($\mathcal{F}_{41}^{6}$), d'après la table de la loi de Fisher $f_{95\%} = 2,32$ (voir l'annexe)

par suite :

$$\mathcal{C}_r = [2,32;\ +\infty[$$

4. **Décision :**

$$f_{obs} = \frac{\frac{270531,97}{6}}{\frac{4729233,11}{41}} = 0,39$$

La valeur $f_{obs} = 0,39$ n'appartient pas à la région critique $\mathcal{C} = [2,32;\ +\infty[$. Il n'est donc pas possible de rejeter l'hypothèse nulle d'égalité des moyennes des sept populations.

Le chiffre d'affaires moyen n'est pas différent d'une région à l'autre au seuil de 5% puisqu'il n'est pas possible de rejeter l'hypothèse nulle d'égalité des moyennes des sept populations. En outre, les régions ont un nombre différent de commerciaux en fonction de leur potentiel. En conclusion, il semble (la probabilité que cette conclusion soit mauvaise n'est pas de 5% mais égale au risque de deuxième espèce) y avoir une bonne affectation des compétences par le groupe au regard de ce premier critère.

↪ **Procédure sous SPSS :** Reprenons les données de l'exemple5. On souhaite étudier l'effet des promotions sur les ventes. Pour ce faire on doit suivre les étapes suivantes (Voir figure 4.1) :

1. On entre les données x_{ij} en colonne dans la variable **Ventes** et la modalité P_j du facteur **Promotions** correspondant à l'observation dans une colonne adjacente (1 pour P_1, 2 pour P_2, 3 pour P_3.)
2. On clique sur **Analyse**, **Comparer les moyennes** et on sélectionne **ANOVA à un facteur**.
3. Dans le menu **ANOVA** dans lequel on sélectionne la variable dépendante **Ventes** et la variable **Promotions** représentant le facteur appelée critère.
4. Si vous souhaitez effectuer des tests post hoc, cliquez sur le bouton correspondant (**les tests post hoc sont expliqués dans la section 4.1.1**). Une fois cochés les tests qui vous intéressent, cliquez sur **Poursuivre**. Si vous sohaitez obtenir des statistiques descriptives pour chaque groupe, cliquez sur **Options**.
5. Cliquer sur **Poursuivre**, puis sur **OK**.

Les résultats obtenus sont présentées dans la figure 4.2 ; La valeur de F est de 17,94 avec 2 et 27 degrés de liberté, d'où une probabilité de 0,000. La probabilité associée étant inférieure au seuil 0,05, l'hypothèse nulle supposant l'égalité des moyennes de la population se trouve rejetée. Les moyennes de l'échantillon $8,3$ $6,2$ et $3,7$ présentent de gros écarts. Les magasins offrant un niveau élevé de promotions bénéficient de la moyenne de ventes la plus importante $(8,3)$; inversement, la moyenne la plus basse $(3,7)$ est enregistrée par les magasins avec un faible niveau de promotions. Ce dernier point (Effets des modalités du facteur) sera étudier en détaille dans la section4.1.1

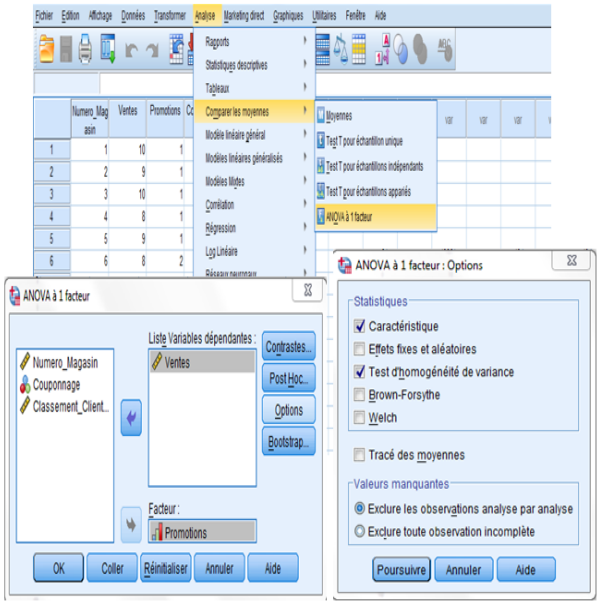


FIGURE 4.1 –

Descriptives

Ventes

	N	Moyenne	Ecart type	Erreur standard	Intervalle de confiance à 95 % pour la moyenne		Minimum	Maximum
					Borne inférieure	Borne supérieure		
Elevé	10	8,30	1,337	,423	7,34	9,26	6	10
Moyen	10	6,20	1,751	,554	4,95	7,45	4	9
Faible	10	3,70	2,003	,633	2,27	5,13	1	7
Total	30	6,07	2,532	,462	5,12	7,01	1	10

Test d'homogénéité des variances

Ventes

Statistique de Levene	ddl1	ddl2	Sig.
1,353	2	27	,275

ANOVA

Ventes

	Somme des carrés	ddl	Carré moyen	F	Sig.
Intergroupes	106,067	2	53,033	17,944	,000
Intragroupes	79,800	27	2,956		
Total	185,867	29			

FIGURE 4.2 –

4.1.1 Les calculs postérieurs au tableau ANOVA

[11] Lorsque l'hypothèse nulle d'égalité des moyennes est rejetée, c'est-à-dire lorsque le facteur manipulé exerce un effet non nul, différents calculs complémentaires sont nécessaires pour qualifier et localiser plus précisément cet effet :

1. **La signification pratique de l'effet d'un facteur :** Elle mesure l'intensité ou l'importance de cet effet. Deux indicateurs peuvent être utilisés η^2 et ω^2. Ces deux indicateur varient tous deux de 0 à 1 et indiquent quel pourcentage de la variance de la variable expliquée est dû au facteur manipulé.
   - $\eta^2 = \dfrac{SCE_F}{SCE_T}$
   - $\omega^2 = \dfrac{SCE_F - (dl_F \times CM_r)}{SCE_T + CM_r}$

   où

   SCE_F **:** somme des carrés des écarts factoriels,

   SCE_T **:** somme des carrés des écarts totaux,

   dl_F **:** degré de liberté du facteur,

   CM_r **:** carré moyen résiduel.

   La signification pratique pour le test de Student est calculée de la façon suivante :

   $$\omega^2 = \frac{t^2 + 1}{t^2 + n_1 + n_2 - 1}$$

   S'agissant de notre exemple de l'impact de Promotions sur les Ventes, l'importance de l'effet du facteur est respectivement égale à $57,1\%$ pour η^2 et $55,7\%$ pour ω^2, soit deux valeurs relativement proches : nous retiendrons toutefois la valeur du coefficient ω^2, en raison de son absence de biais. Nous pouvons donc conclure que $55,7\%$ de la variance des ventes est expliquée par la manipulation du facteur Promotions, soit une valeur relativement élevée.

2. **Les comparaisons multiples de moyennes**. Lorsqu'un facteur possède plus de deux modalité, ne permet pas d'identifier les modalités qui sont à l'origine de l'effet détecté, à partir de la seule lecture du tableau ANOVA. Les comparaisons multiples de moyennes le permettent.

   Deux situations peuvent se présenter. Si le chercheur n'a aucune idée sur les modalité qui sont à l'origine de l'effet, il doit utiliser les **tests post-hoc**. En revanche, s'il dispose de cette information a priori, il doit alors utiliser les **test a priori**.

   Les test post-hoc Les tests post-hoc consistent à effectuer des comparaisons par paires des moyennes entre elles. En présence d'un petit nombre de groupes expérimentaux, le test de Dunn-Bonferroni s'avère le plus intéressant. Inversement, mais pour les mêmes raisons, le test de Tukey est préconisé lorsque le nombre de groupes est plus élevé. Les autres tests sont moins précis et moins utilisés.

   Les tests a priori Les tests a priori sont dans l'ensemble plus sensibles que les précédents, dans la mesure où ils sont capables d'identifier plus finement les différences existantes. La méthode des **contrastes** est la plus connue.

   Un contraste est une combinaison linéaire de moyennes dont la somme des coefficients est égale à 0. Un contraste C peut être spécifié de la façon suivante :

   $$C = C_1\mu_1 + C_2\mu_2 + ... + C_k\mu_k$$

   avec C_i : coefficient du contraste avec somme des $C_i = 0$ et μ_i moyenne du groupe i. Ainsi pour l'exemple de Promotions ; l'hypothèse nulle $H_0 : \mu_1 = \mu_2$ peut être transformée selon le contraste suivant : : $1 * \mu_1 - 1 * \mu_2 + 0 * \mu_3$.

Deux contrastes sont dit orthogonaux lorsque la somme des produits des coefficients propres à chaque moyenne est égale à 0.

Pour un exposé plus approfondi sur les tests a priori et a posteriori, consulter par exemple [9] et [9] .

3. **La détection des "trends" et des effets de seuils** L'identification de la forme de la relation existant entre les modalité du facteur explicatif et la variable à expliquer s'avère intéressante lorsqu'il existe une hiérarchie dans les modalités. Elle permet notamment de mettre en évidence des effets de seuil. Supposons, par exemple, qu'on souhaite tester l'impact sur les ventes de différents niveaux de promotions : la visualisation de la forme que prend la courbe entre les niveaux de promotions classés par ordre croissant et les ventes permet de déterminer à partir de quel seuil de promotions les ventes croit. Pour plus de détailles consulter [11]

4.2 Analyse de variance à deux facteurs facteur

↪ **Principe :** [12] Dans le domaine des études marketing, on s'intéressent souvent aux effets simultanés de plusieurs facteurs. Par exemple :

- Comment les intentions d'achats des consommateurs varient-elles en fonction des différents niveaux de prix et des différents niveaux de distributions d'une marque ?
- Comment les intentions d'achats des consommateurs varient-elles en fonction des différents niveaux de prix et des différents niveaux de distributions d'une marque ?
- Comment l'interaction entre le niveau de publicité (élevé, moyen, faible) et le niveau des prix (élevé, moyen, faible) influence-t-elle les ventes d'une marque ?
- Le niveau d'études (inférieur au baccalauréat, baccalauréat, formation universitaire, diplôme universitaire) et l'age (moins de 35 ans, 35-55 ans, plus de 55 ans) affectent-ils la consommation d'une marque ?
- La connaissance d'un magasin donné (élevé, moyen, faible) et l'image de marque de ce magasin (positive, neutre, négative) influencent-elles les préférences du consommateurs ?

Il s'agit à présent d'étudier les situations où il est possible de relever plusieurs observations à la fois, à l'intérieur d'un même échantillon (selon le premier facteur) et d'une même catégorie (selon deuxième facteur). La méthode utilisée ici présente deux avantages :

- Le plus grand nombre d'observations disponibles dans les échantillons rend le test de différences des moyennes plus précis.
- Une troisième source de variation est identifiée : elle concerne les interactions entre les échantillons et les catégories qui sont susceptibles d'apparaître lorsque les effets spécifiques aux échantillons ne sont pas distribués uniformément parmi les catégories. Par exemple, l'utilisation d'un type de carburant particulier est censé être plus efficace (minimiser la consommation) associé à un modèle particulier de carburateur.

4.2.1 La détection de l'interaction

[11] Lorsqu'il existe un facteur statistiquement significatif, il se peut que son effet soit plus ou moins prononcé selon les modalité d'un autre facteur. On parle alors **d'interaction** entre les deux facteur. Plus généralement, une interaction a lieu lorsque la modalité d'un facteur se comporte de façon fort différente en fonction des modalités de l'un ou des autres facteurs.

Une interaction entre facteurs est représentée par un graphique dans lequel les droites tracées ne sont pas parallèles. L'interaction peut être **ordinal**(l'ordre des effets liés au premier facteur respecte celui des niveaux du second facteur) ou **non ordinal** (implique une modification dans l'ordre des effets). Si l'interaction est non ordinale, elle peut être croisée ou non croisée. La figure 4.3 illustre ces différents cas, en supposant l'existence de deux facteurs X_1 et X_2 dotés respectivement de trois niveaux (X_{11}, X_{12}, X_{13}) et de deux niveaux (X_{21}, X_{22})

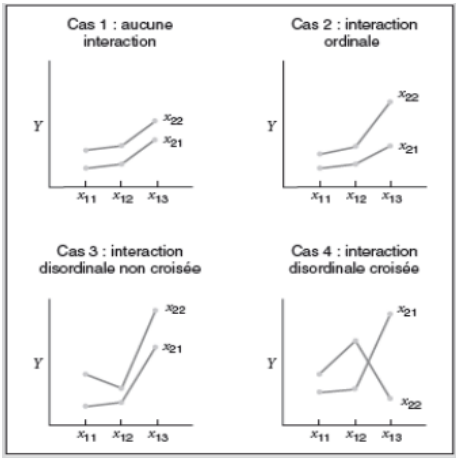


FIGURE 4.3 – Diagrammes d'interactions. Source :[12]

Le repère 11 présente les différents calculs associés à cette méthode.

Repère 11 (La boite à outils : ANOVA à deux facteurs), *Source :[17]*

Source de variation	**Somme des carrés**	**Degrés de liberté**	**Variances**	**Variable de décision**
inter-échantillon	SCI_e	$k-1$	$\frac{SCI_e}{k-1}$	$F_E = \frac{\frac{SCI_e}{k-1}}{\frac{SCE}{kh(g-1)}}$
inter-catégorie	SCI_c	$h-1$	$\frac{SCI_c}{h-1}$	$F_C = \frac{\frac{SCI_c}{h-1}}{\frac{SCE}{kh(g-1)}}$
interaction	SCI_{ec}	$(k-1)(h-1)$	$\frac{SCI_{ec}}{(k-1)(h-1)}$	$F_{EC} = \frac{\frac{SCI_{ec}}{(k-1)(h-1)}}{\frac{SCE}{kh(g-1)}}$
intra-échantillon (erreur)	SCE	$kh(g-1)$	$\frac{SCE}{kh(g-1)}$	
Total	SCT	$khg-1$		

- **Somme des carrés inter-échantillon :** $SCI_e = hg\sum_{i=1}^{k}(\bar{X}_{i\cdot\cdot} - \bar{X})^2$
- **Somme des carrés inter-catégorie :** $SCI_c = kg\sum_{j=1}^{h}(\bar{X}_{\cdot j\cdot} - \bar{X})^2$
- **Somme des carrés des interactions :** $SC_{ec} = g\sum_{i=1}^{k}\sum_{j=1}^{h}(\bar{X}_{ij\cdot} - \bar{X}_{\cdot j\cdot} - \bar{X}_{i\cdot\cdot} + \bar{X})^2$
- **Somme des carrés des erreurs :** $SCE = \sum_{i=1}^{k}\sum_{j=1}^{h}\sum_{l=1}^{g}(X_{ijl} - \bar{X}_{ij\cdot})^2$
- **Somme des carrés totaux :**$SCT = \sum_{i=1}^{k}\sum_{j=1}^{h}\sum_{l=1}^{g}(X_{ijl} - \bar{X})^2$

↪ **Exemple d'application :**[17] **Quelle est la meilleur place ?**

Les serveurs des débits de boissons ont pour coutume d'effectuer des rotations entre la salle, la terrasse et le bar. Ces rotations de services ont lieu, suivant les établissements, la journée ou de manière hebdomadaire. Cette tradition a pour but de ne pas désavantager les serveurs entre eux, car ceux-ci sont en général rémunérés à un taux fixe appliqué à leur chiffre d'affaires. Le patron d'un débit de boissons s'interroge sur l'opportunité de ces rotations. il pense par ailleurs qu'un second facteur explicatif des différences de salaires de ses employés est lié à leur expérience. Pour étudier la pertinence de son intuition, il relève les données présentées dans le tableau ci-dessous

Moyennes : d'échantillons, de catégories et de couples échantillons-catégories

	Salle	**Terrasse**	**Bar**	**Moyenne par catégorie**
sans	433	510	560	$\bar{X}_{.1.}$=506,2
sans	454	540	540	
	$\bar{X}_{11.}$=443,5	$\bar{X}_{21.}$=525,0	$\bar{X}_{31.}$=550,0	
faible	460	528	587	$\bar{X}_{.2.}$=527,3
faible	482	527	580	
	$\bar{X}_{12.}$=471,0	$\bar{X}_{22.}$=527,5	$\bar{X}_{32.}$=583,5	
moyenne	546	560	610	$\bar{X}_{.3.}$=576,0
moyenne	570	570	600	
	$\bar{X}_{13.}$=558,0	$\bar{X}_{23.}$=565,0	$\bar{X}_{33.}$=605,0	
assez élevée	615	625	610	$\bar{X}_{.4.}$=613,7
assez élevée	577	635	620	
	$\bar{X}_{14.}$=596,0	$\bar{X}_{24.}$=630,0	$\bar{X}_{34.}$=615,0	
élevée	630	568	580	$\bar{X}_{.5.}$=623,3
élevée	620	675	667	
	$\bar{X}_{15.}$=525,0	$\bar{X}_{25.}$=621,5	$\bar{X}_{35.}$=623,5	
Moyenne par échantillon	$\bar{X}_{1..}$=538,7	$\bar{X}_{2..}$=573,8	$\bar{X}_{3..}$=595,4	$\bar{X}$ =569,3

Les réalisations des variables de décision et les quantiles associés sont :

•Le test inter-échantillon est :

$$f_{Eobs} = \frac{\frac{16378,5}{3-1}}{\frac{11906,5}{3\times 5\times(2-1)}} = \frac{8189,1}{793,8} = 10,32 \text{ avec } f_{0,95} = 3,68 \text{ pour } F_E \sim \mathcal{F}(2;15)$$

•Le test inter-catégories est :

$$f_{Cobs} = \frac{\frac{64079,5}{5-1}}{\frac{11906,5}{3\times 5\times(2-1)}} = \frac{16019,9}{793,8} = 20,18 \text{ avec } f_{0,95} = 3,06 \text{ pour } F_C \sim \mathcal{F}(4;15)$$

•Le test interaction est :

$$f_{ECobs} = \frac{\frac{12430,1}{2\times 4}}{\frac{11906,5}{3\times 5\times(2-1)}} = \frac{1553,8}{793,8} = 1,96 \text{ avec } f_{0,95} = 2,64 \text{ pour } F_{EC} \sim \mathcal{F}(8;15)$$

Deux hypothèses nulles doivent être rejetées. Au seuil de 5% :

- L'impact du " lieu de travail " sur les chiffres d'affaires moyens est statistiquement significatif.

- L'impact de " l'expérience " sur les chiffres d'affaires moyens est également significatif.

Par contre, l'hypothèse d'absence d'interaction entre ces deux facteurs explicatifs ne peut pas être rejetée. Dans cette situation, la procédure d'ANOVA à deux facteurs permet de donner une réponse claire car elle isole les effets respectifs des deux facteurs sur la variable étudiée. Si une interaction entre le " lieu de travail " et le " niveau d'expérience " avait été mise en évidence, alors l' interprétation de l'analyse aurait été plus délicate. En effet, dans ce cas, l'effet du "lieu de travail " aurait pu être simplement dû à la présence conjointe des deux facteurs. De façon similaire, l'interaction aurait pu masquer l'effet individuel de l'un des deux facteurs. En d'autres termes, la présence d'un effet d'interaction doit susciter des questionnements supplémentaires, voire même une nouvelle spécification du test effectué.

↪ **Procédure sous SPSS :** (Voir figure 4.4) A partir des données de l'exemple 5, on peut examiner à présent l'effet du niveau des promotions et du couponnage sur les ventes. Pour la réalisation de cette ANOVA sous SPSS, il suffit de suivre les étapes suivantes :

1. Choisissez **Analyse**, **Modèle linéaire général** puis **Univarié**.
2. Déplacez la **VD** et les facteurs**VI** dans les zone ad hoc, comme dans la figure 4.4
3. Pour obtenir des informations sur les tailles d'effet (effets principaux et interactions) cliquez sur **Options** et sélectionnez **Estimations d'effets de taille**[1] ainsi que **Statistiques descriptives si vous le souhaitez**
4. Si vous souhaitez obtenir des diagrammes de variation (pour l'étude d'interaction) cliquez sur **Tracés** (voir figure 4.4 pour la suite)
5. Finissez par **OK** et l'analyse tourne.

Les résultats sont reproduit dans le tableau suivant (figure 4.5) : L'augmentation du niveau des promotions et la distribution de bon d'achats entrainent une augmentation des ventes, mais ces deux effets demeurent indépendants l'un de l'autre La figure 4.5 illustre bien cette absence d'interaction.

4.3 Analyse de covariance

↪ **Principe :** [12] Lorsqu'on examine les écarts des valeurs moyennes de la variable dépendante liés à l'effet des variables indépendantes contrôlées, il s'avère souvent nécessaire de prendre en compte l'influence de variables indépendantes non contrôlées (covariables). Ainsi :

- Pour déterminer de quelle manière les intentions d'achats des consommateurs varient en fonction des différents niveaux de prix, il faudra sans doute tenir compte de leur attitude par rapport à la marque.
- Pour déterminer de quelle manière différents groupes exposés à différentes publicités évaluent une marque, il faudra peut-être contrôler leur connaissance préalable de cette marque .
- Pour déterminer de quelle manière différents niveaux de prix affectent la consommation de céréales d'une famille, la prise en compte de la taille du foyer pourra s'avérer indispensable.

Pour légitimer l'ANCOVA, il faut en effet vérifier les mêmes conditions que pour l'ANOVA (voir 4.1). Mais il faut en outre que

1. Le logiciel a vraisemblablement été mal traduit, car c'est bien "estimations de tailles d'effet" qu'il faut comprendre.

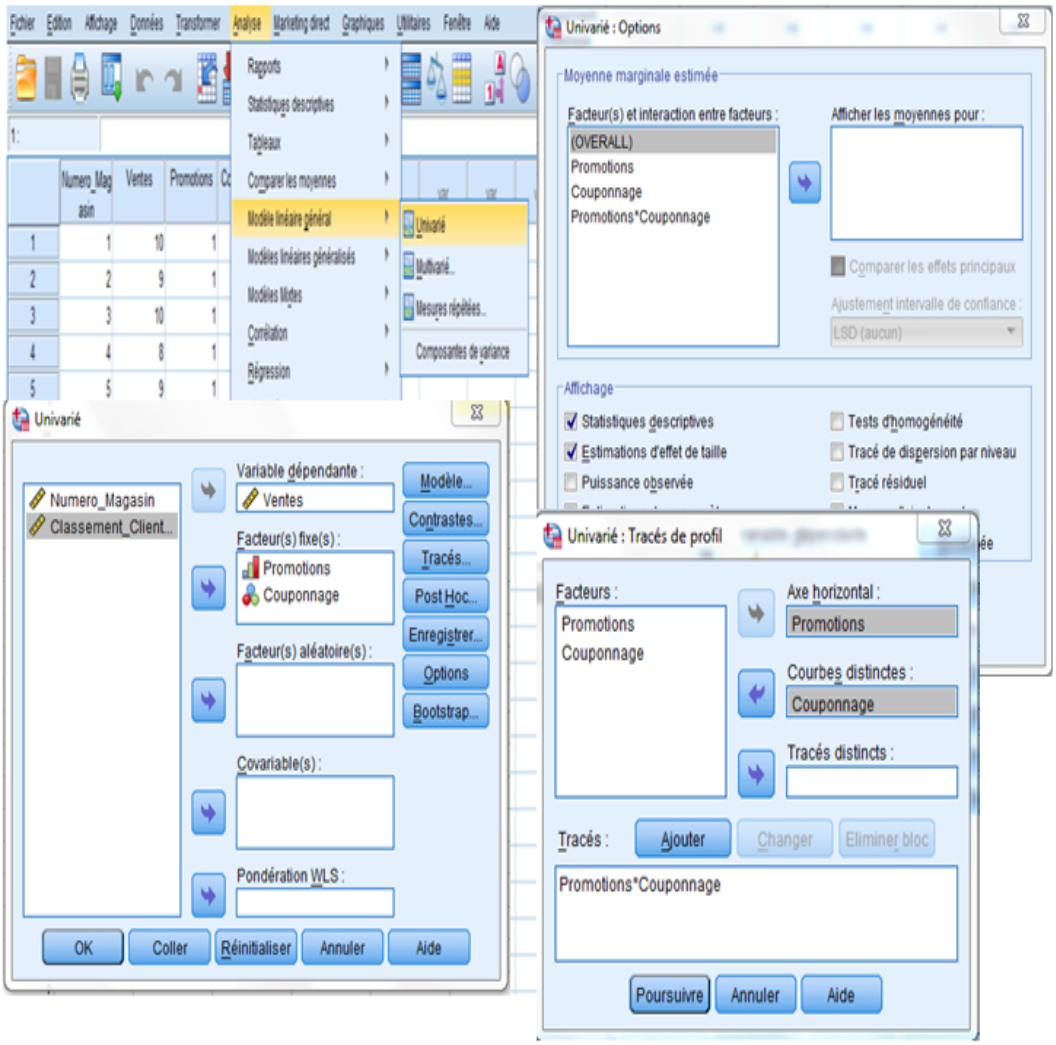


FIGURE 4.4 –

- **la covariable soit liée linéairement à la variable dépendante :** S'il n'y a pas de relations entre la variable dépendante et la covariable, et bien l'ANCOVA n'a aucun intérêt
- **la covariable soit fiable (pas d'erreur) :** Cela que si vous mesuriez la covariable en différentes occasions, vous devriez trouver des valeurs fortement corrélées.
- **les droites de régression soient parallèles :** C'est en fait assez logique. Si les droites ne sont pas parallèles, alors utiliser une procédure qui extrapole la moyenne de la covariable pour en déduire la valeur moyenne de la variable dépendante n'a aucun sens (pour plus de détailles sur ces trois conditions vous pouvez consulter [4]).

↪ **Exemple d'application :**[12] L'illustration de cette analyse s'appuie une nouvelle fois sur les données de l'exemple 5. On peut supposer que l'on souhaite déterminer l'effet des promotions en tenant compte de l'effet de la clientèle, dont on soupçonne que le niveau de vie puisse avoir un impact sur les les ventes du magasin. La variable dépendante représente les ventes. Comme précédemment, les promotions sont codées sur trois niveaux . La clientèle, mesurée sur une échelle d'intervalles, fait office de covariable.

↪ **Procédure sous SPSS :** (Voir figure 4.7) Voyons maintenant comment demander une ANCOVA à SPSS

1. Commencez par choisir **Analyse**, **Modèle linéaire général** puis **Univarié**. Cela vous amène au menu principal de l'analyse de covariance.

Tests des effets intersujets

Variable dépendante: Ventes

Source	Somme des carrés de type III	ddl	Carré moyen	F	Signification	Eta-carré partiel
Modèle corrigé	162,667[a]	5	32,533	33,655	,000	,875
Constante	1104,133	1	1104,133	1142,207	,000	,979
Promotions	106,067	2	53,033	54,862	,000	,821
Couponnage	53,333	1	53,333	55,172	,000	,697
Promotions * Couponnage	3,267	2	1,633	1,690	,206	,123
Erreur	23,200	24	,967			
Total	1290,000	30				
Total corrigé	185,867	29				

a. R-deux = ,875 (R-deux ajusté = ,849)

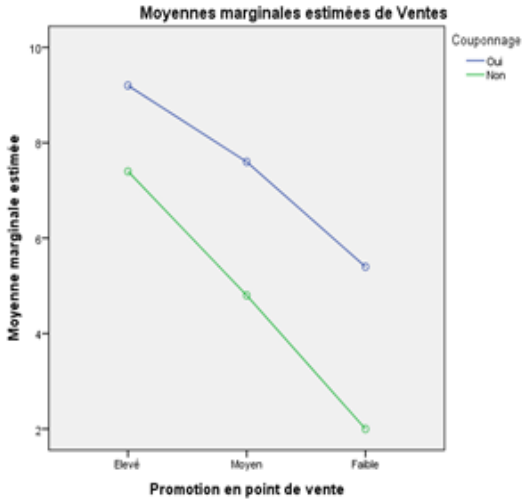


FIGURE 4.5 –

2. Définir correctement les variables : dépendante (**Ventes**) et indépendantes "Facteur(s) fixé(s)"(**Promotions**) ainsi que la covariable **Classement-clientèle**
3. Choisir **Options** et glisser la variable **Promotions** dans la zone de texte **Afficher moyennes** pour obtenir les moyennes brutes et ajustées (nommées "**moyennes marginales estimées**"). Cliquer sur poursuivre.
4. Il nous faut maintenant savoir si les droites de régression sont suffisamment parallèles. Pour cela , cliquer sur **Modèle**
5. Choisir **Personnalisé** Sélectionner les deux variables (**Promotions**et **Classement-clientèle**). Dans un encart situé au centre de la boite et nommé **Terme(s) construit(s)**, sélectionner d'abord **Effets principaux** dans le menu déroulant. Cliquer ensuite sur la flèche ↪ pour faire passer les deux variables à droite.
6. Retourner maintenant dans le menu déroulant pour remplacer **Effets principaux** par **Interaction**. Sélectionner à nouveau les deux variables dans la zone de gauche. Cliquer sur ↪. Une ligne supplémentaire doit alors apparaitre dans **Modèle**, indiquant que l'interaction est prise en compte (dans notre cas cela donne "Promotion*Classement-clientèle").
7. Cliquer sur poursuivre puis sur OK.

Dans les sorties qui s'affichent, vous remarquez un tableau 4.6(le seul qui soit véritable-

ment utile)

Tests des effets intersujets

Variable dépendante: Ventes

Source	Somme des carrés de type III	ddl	Carré moyen	F	Signification	Eta-carré partiel
Modèle corrigé	109,203[a]	5	21,841	6,837	,000	,588
Constante	103,346	1	103,346	32,353	,000	,574
Promotions	6,408	2	3,204	1,003	,382	,077
Classement_Clientele	,838	1	,838	,262	,613	,011
Promotions * Classement_Clientele	**2,298**	**2**	**1,149**	**,360**	**,702**	**,029**
Erreur	76,664	24	3,194			
Total	1290,000	30				
Total corrigé	185,867	29				

a. R-deux = ,588 (R-deux ajusté = ,502)

FIGURE 4.6 –

De ce tableau, nous ne retenons qu'une seule valeur, c'est la signification associée à l'interaction (en gras), soit dans notre cas $p = 0,702$. On peut conclure que l'interaction n'a pas d'effet significatif sur la VD, ce qui est une autre manière de dire que les droites de régression sont suffisamment parallèles.

8. Revenez donc à la fenêtre principale de SPSS, et sélectionnez à nouveau **Analyse**, **Modèle linéaire général** puis **Univarié**. Choisissez **Modèle...** et cliquez sur **Facteur complet.**
9. Cliquez ensuite sur **Poursuivre**, puis sur **OK** .

Les résultats sont reproduits dans la figure 4.8 : La somme des carrés imputable à la covariable se révèle très faible (0,838 avec 1 ddl). La valeur F correspondante est égale à 0,862 avec 1 et 23 ddl, ce qui n'est pas significatif pour un seuil de 5%¨. On en déduit que le niveau de vie de la clientèle n'exerce aucun effet sur les ventes du magasin.

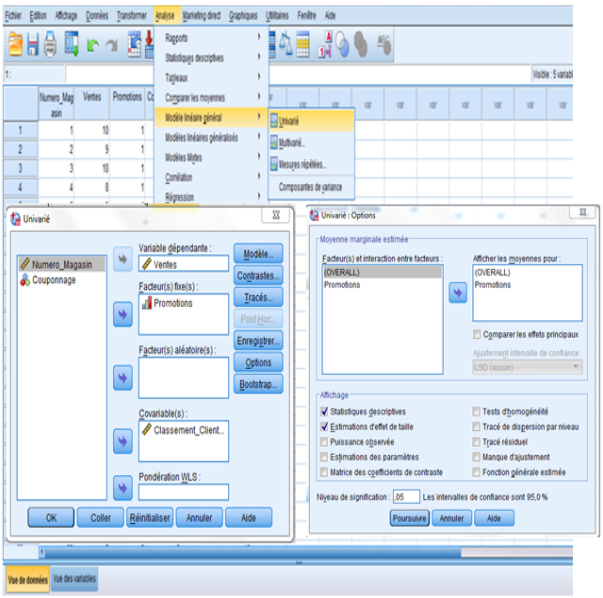


FIGURE 4.7 –

Statistiques descriptives

Variable dépendante: Ventes

Promotion en point de vente	Moyenne	Erreur type	N
Elevé	8,30	1,337	10
Moyen	6,20	1,751	10
Faible	3,70	2,003	10
Total	6,07	2,532	30

Tests des effets intersujets

Variable dépendante: Ventes

Source	Somme des carrés de type III	ddl	Carré moyen	F	Signification	Eta-carré partiel
Modèle corrigé	106,905[a]	3	35,635	11,734	,000	,575
Constante	103,346	1	103,346	34,029	,000	,567
Classement_Clientele	,838	1	,838	,276	,604	,011
Promotions	106,067	2	53,033	17,462	,000	,573
Erreur	78,962	26	3,037			
Total	1290,000	30				
Total corrigé	185,867	29				

a. R-deux = ,575 (R-deux ajusté = ,526)

FIGURE 4.8 –

4.4 Travaux pratiques

TP4

1. ANOVA One-Way

Supposons que trois types de promotions marketing aient été testés pour une nouvelle marque de soupe : un Poster (affichage) du produit dans le magasin(**R1**), une dégustation dans le magasin(**R2**) et la décoration autour du stand du produit(**R3**). Chacune de ces promotions spéciales a été testée dans cinq magasins différents. Ces magasins doivent être comparables de sorte qu'aucun «facteur caché» ne soit inclus dans l'analyse par erreur. Supposons que les cinq magasins dans lesquels un poster est utilisé sont également des boutiques qui ont été récemment rénovées et les dix autres magasins viennent à travers comme étant à l'ancienne. Si nous trouvons une différence significative entre les magasins avec et sans poster, la question est de savoir si cette différence a son origine dans la façon dont l'affiche est utilisée ou dans le look rénové dans les magasins. Le tableau 1 illustre les données de ventes recueillies après un mois.

Magasin/Promotion	**R1**	**R2**	**R3**
Magasin 1	7	15	9
Magasin 2	4	13	7
Magasin 3	4	17	5
Magasin 4	5	11	10
Magasin 5	6	13	8

TABLE 4.2 – Table 1 : les ventes

Cet exemple tente donc de déterminer l'effet d'un facteur ou variable indépendante(promotion), mesuré sur trois niveaux. Ces données sont saisies dans SPSS (**soup.sav**).

Problème

Effectuer une ANOVA One-Way pour déterminer l'effet des différentes promotions sur les ventes. En utilisant des tests post-hoc, découvrez également si l'effet diffère entre chaque paire de promotions.

2. Analyse de variance avec une covariable (ANCOVA)

Un manager d'une chaîne de magasins aimerait savoir si oui ou non la pulvérisation d'un parfum (huile de lavande) dans les magasins se traduira par des ventes plus élevées.

Pour ce faire, il sélectionne 30 magasins différents qui sont de nature similaire. Dans 10 des magasins, aucun parfum n'est pulvérisé, et ce groupe de magasins sert également de référence pour l'étude (groupe de contrôle). Dans 10 autres magasins, il ne vaporise qu'une très petite quantité d'huile de lavande. Dans les magasins restants, il permet de pulvériser une quantité importante. En ce qui concerne les magasins où l'huile de lavande est pulvérisée, des recherches antérieures avaient déjà montré que la pulvérisation limitée a habituellement conduit à une perception légère au parfum , tandis que pour les magasins où une quantité significative a été pulvérisée, la perception au parfum était forte ; En d'autres termes, les gens qui sont entrés dans le magasin ont immédiatement remarqué le parfum de lavande.

Comme dans le cas de l'ANOVA, cette étude doit également impliquer des magasins comparables, de sorte qu'aucun «facteur caché» n'est inclus dans l'analyse. Un facteur que le chercheur ne peut pas le contrôler est que les magasins ont des surfaces différentes et qu'une recherche antérieure avait déjà prouvé que les magasins de grandes surfaces

réalisent des ventes plus élevées. Afin de libérer l'expérience de tout effet potentiellement avoir une influence sur les résultats, le facteur surface est inclus comme covariable (variable de contrôle)dans la collecte et l'analyse de données

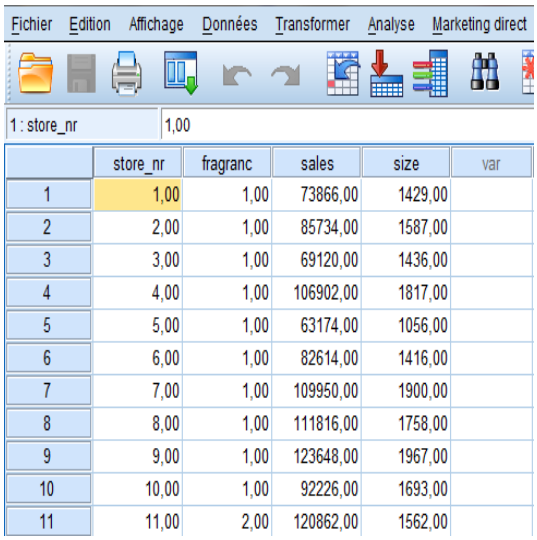
Contrairement au problème 1 (relative à l'ANOVA) dans lequel il n'y avait pas une idée apriori sur les différences entre les niveaux du facteur, le chercheur s'attend à ce que le groupe de référence connaisse les ventes les plus médiocres (en raison du manque de parfum de lavande agréable) par rapport aux magasins dans lesquels un parfum de lavande est en fait utilisé. Ce que le chercheur n'a aucune idée apriori est la question de savoir si oui ou non le groupe dans lequel une dose élevée de parfum de lavande est utilisée aura un meilleur rendement que le groupe dans lequel une dose plus faible est utilisée.

Après tout, la forte présence de parfum peut mettre les clients dans un meilleur état d'esprit, Comment cela peut aussi créer un effet négatif parce qu'un certain nombre de clients peuvent juger cela inapproprié et / ou provoquer un sentiment d'être manipulé.

Problème

Effectuer une ANCOVA et voir si l'utilisation d'une faible, élevée ou aucune dose de parfum de lavande a un effet sur les ventes. Ensuite, en utilisant les comparaisons de type contrastes, voir si les doses faibles et élevées conduisent à de meilleures ventes par rapport à une situation où aucun parfum de lavande n'est utilisé. Enfin, en utilisant des tests post-hoc, voir si il y a une différence dans les ventes entre les groupes à faible dose et à dose élevée. Alors que vous faites cela, gardez à l'esprit la possibilité d'une influence linéaire du facteur "surface des magasin" sur les ventes.

Les données sont dans le fichier "fragrance.sav", dont on peut trouver une illustration dans la figure 4.9.



	store_nr	fragranc	sales	size	var
1	1,00	1,00	73866,00	1429,00	
2	2,00	1,00	85734,00	1587,00	
3	3,00	1,00	69120,00	1436,00	
4	4,00	1,00	106902,00	1817,00	
5	5,00	1,00	63174,00	1056,00	
6	6,00	1,00	82614,00	1416,00	
7	7,00	1,00	109950,00	1900,00	
8	8,00	1,00	111816,00	1758,00	
9	9,00	1,00	123648,00	1967,00	
10	10,00	1,00	92226,00	1693,00	
11	11,00	2,00	120862,00	1562,00	

FIGURE 4.9 – les variables du fichier "fragrance.sav"

Les «ventes» sont les ventes hebdomadaires , «parfum» indique le groupe expérimental auquel le magasin appartient (0 = pas de Parfum de lavande, 1 = faible dose de parfum de lavande, 2 dose élevée de parfum de lavande). «Taille» est la superficie du magasin exprimée en m^2.

3. Analyse de variance pour un plan factoriel complet $2 \times 2 \times 2$

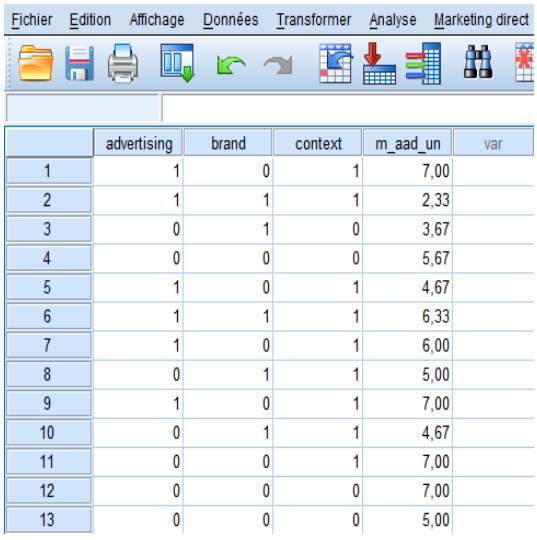
Lorsqu'un chercheur souhaite tester l'effet de différents niveaux de plusieurs facteurs sur une variable dépendante, une conception factorielle sera nécessaire. Dans l'exemple suivant, l'effet de trois variables dichotomiques sur une variable dépendante sera étudié,

dans ce cas l'attitude à l'égard d'une annonce publicitaire (Attitude envers l'annonce, ou Aad). Lorsqu'on étudie l'effet de plus d'un facteur, les effets d'interaction sont possibles, ce qui signifie que l'effet d'un facteur dépend des niveaux d'un autre facteur ; En d'autres termes, des combinaisons des facteurs peuvent avoir une influence significative sur la variable dépendante.

Une recherche antérieure a montré que pour les nouveaux produits, il est préférable d'utiliser une publicité rationnelle plutôt qu'une publicité émotionnelle (les consommateurs préfèrent l'information claire et rationnelle), tandis que pour les produits existants, une publicité émotionnelle est le meilleur choix. La question est maintenant de savoir si le contexte médiatique dans lequel la publicité sera montrée est également important. On pourrait aussi se demander si l'interaction entre les différents facteurs a une influence sur Aad. Étant donné qu'il existe trois variables indépendantes (type de marque, type d'annonce, type de contexte), chacune étant mesurée sur deux niveaux, il s'agit d'une conception factorielle $2 \times 2 \times 2$.

En d'autres termes, il existe huit combinaisons possibles différentes. Pour chacune de ces combinaisons, une publicité a été crée. Chaque publicité a été montrée à un groupe (différent) d'environ 25 personnes. Chacun des groupes a été sélectionné au hasard. Ensuite, on leur a demandé de remplir un questionnaire comprenait un certain nombre d'énoncés mesurant un aspect de la variable Aad : (la compréhension de la publicité [m-aad-be]) (échelle de type Likert à 7 points). Cette variable est calculée comme la moyenne des énoncés pertinents pour ce concept.

Dans la figure 4.10, une partie du fichier context.sav est affichée. Cet ensemble de données montre les différents facteurs et la variable «compréhension».



	advertising	brand	context	m_aad_un	var
1	1	0	1	7,00	
2	1	1	1	2,33	
3	0	1	0	3,67	
4	0	0	0	5,67	
5	1	0	1	4,67	
6	1	1	1	6,33	
7	1	0	1	6,00	
8	0	1	1	5,00	
9	1	0	1	7,00	
10	0	1	1	4,67	
11	0	0	1	7,00	
12	0	0	0	7,00	
13	0	0	0	5,00	

FIGURE 4.10 – les variables du fichier "context.sav"

Problème

Déterminer si les différentes variables dichotomiques ont une influence significative sur la compréhension du message publicitaire.

Vérifiez également s'il existe des effets d'interaction entre différentes variables indépendantes et présentez-les sous forme graphique.

Annexe :Les Tables Statistiques

- Table 1 : Coefficients binomiaux
- Table 2 : Loi normale centrée réduite
- Table 3 : Loi de Student
- Table 4 : Loi du Khi-deux
- Table 5a : Loi de Fisher-Snedecor
- Table 5b : Loi de Fisher-Snedecor
- Table 5c : Loi de Fisher-Snedecor
- Table 5d : Loi de Fisher-Snedecor
- Table 6 : Loi de Kolmogorov-Smirnov
- Table 7 : Loi de Wilcoxon
- Table 8 : Loi de Mann-Whitney
- Table 9 : Table de Kruskal-Wallis

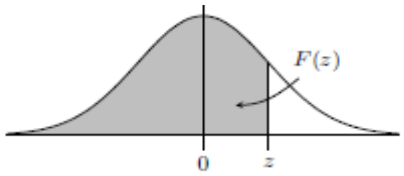
Table donnant les $\binom{n}{k}$

Rappel : $\binom{n}{k} = \binom{n}{n-k} = \begin{cases} \frac{n!}{k!(n-k)!} & \text{si } 0 \leq k \leq n \\ 0 & \text{sinon.} \end{cases}$

n \ k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	3	3	1	0	0	0	0	0	0	0	0	0	0	0	0
4	1	4	6	4	1	0	0	0	0	0	0	0	0	0	0	0
5	1	5	10	10	5	1	0	0	0	0	0	0	0	0	0	0
6	1	6	15	20	15	6	1	0	0	0	0	0	0	0	0	0
7	1	7	21	35	35	21	7	1	0	0	0	0	0	0	0	0
8	1	8	28	56	70	56	28	8	1	0	0	0	0	0	0	0
9	1	9	36	84	126	126	84	36	9	1	0	0	0	0	0	0
10	1	10	45	120	210	252	210	120	45	10	1	0	0	0	0	0
11	1	11	55	165	330	462	462	330	165	55	11	1	0	0	0	0
12	1	12	66	220	495	792	924	792	495	220	66	12	1	0	0	0
13	1	13	78	286	715	1287	1716	1716	1287	715	286	78	13	1	0	0
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001	364	91	14	1	0
15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003	1365	455	105	15	1
16	1	16	120	560	1820	4368	8008	11440	12870	11440	8008	4368	1820	560	120	16
17	1	17	136	680	2380	6188	12376	19448	24310	24310	19448	12376	6188	2380	680	136
18	1	18	153	816	3060	8568	18564	31824	43758	48620	43758	31824	18564	8568	3060	816
19	1	19	171	969	3876	11628	27132	50388	75582	92378	92378	75582	50388	27132	11628	3876
20	1	20	190	1140	4845	15504	38760	77520	125970	167960	184756	167960	125970	77520	38760	15504
21	1	21	210	1330	5985	20349	54264	116280	203490	293930	352716	352716	293930	203490	116280	54264
22	1	22	231	1540	7315	26334	74613	170544	319770	497420	646646	705432	646646	497420	319770	170544
23	1	23	253	1771	8855	33649	100947	245157	490314	817190	1144066	1352078	1352078	1144066	817190	490314
24	1	24	276	2024	10626	42504	134596	346104	735471	1307504	1961256	2496144	2704156	2496144	1961256	1307504
25	1	25	300	2300	12650	53130	177100	480700	1081575	2042975	3268760	4457400	5200300	5200300	4457400	3268760
26	1	26	325	2600	14950	65780	230230	657800	1562275	3124550	5311735	7726160	9657700	10400600	9657700	7726160
27	1	27	351	2925	17550	80730	296010	888030	2220075	4686825	8436285	13037895	17383860	20058300	20058300	17383860
28	1	28	378	3276	20475	98280	376740	1184040	3108105	6906900	13123110	21474180	30421755	37442160	40116600	37442160
29	1	29	406	3654	23751	118755	475020	1560780	4292145	10015005	20030010	34597290	51895935	67863915	77558760	77558760
30	1	30	435	4060	27405	142506	593775	2035800	5852925	14307150	30045015	54627300	86493225	119759850	145422675	155117520

FIGURE 11 – Table 1 : Coefficients binomiaux

$F(z) = \mathbb{P}[Z < z]$ en fonction de z pour $Z \rightsquigarrow \mathcal{N}(0;1)$.



z	0	1	2	3	4	5	6	7	8	9
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

Remarque :

Si $z < 0$, alors $F(z) = 1 - F(|z|)$.

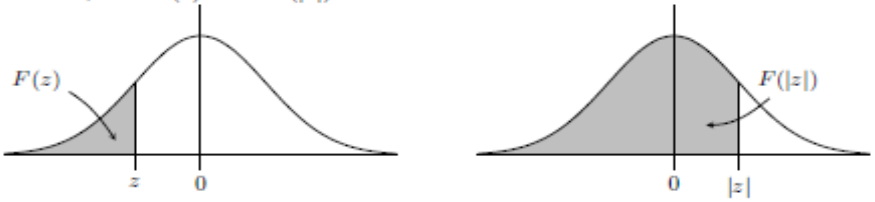
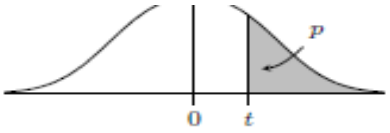


Figure 12 – Table 2 : Loi normale centrée réduite

t en fonction de p tel que $p = \mathbb{P}[T \geqslant t]$
pour T suivant une loi de Student.



ddl \ P	0,2	0,15	0,1	0,05	0,04	0,03	0,025	0,02	0,015	0,01	0,005
1	1,3764	1,9626	3,0777	6,3138	7,9158	10,5789	12,7062	15,8945	21,2049	31,8205	63,6567
2	1,0607	1,3862	1,8856	2,9200	3,3198	3,8964	4,3027	4,8487	5,6428	6,9646	9,9248
3	0,9785	1,2498	1,6377	2,3534	2,6054	2,9505	3,1824	3,4819	3,8960	4,5407	5,8409
4	0,9410	1,1896	1,5332	2,1318	2,3329	2,6008	2,7764	2,9985	3,2976	3,7469	4,6041
5	0,9195	1,1558	1,4759	2,0150	2,1910	2,4216	2,5706	2,7565	3,0029	3,3649	4,0321
6	0,9057	1,1342	1,4398	1,9432	2,1043	2,3133	2,4469	2,6122	2,8289	3,1427	3,7074
7	0,8960	1,1192	1,4149	1,8946	2,0460	2,2409	2,3646	2,5168	2,7146	2,9980	3,4995
8	0,8889	1,1081	1,3968	1,8595	2,0042	2,1892	2,3060	2,4490	2,6338	2,8965	3,3554
9	0,8834	1,0997	1,3830	1,8331	1,9727	2,1504	2,2622	2,3984	2,5738	2,8214	3,2498
10	0,8791	1,0931	1,3722	1,8125	1,9481	2,1202	2,2281	2,3593	2,5275	2,7638	3,1693
11	0,8755	1,0877	1,3634	1,7959	1,9284	2,0961	2,2010	2,3281	2,4907	2,7181	3,1058
12	0,8726	1,0832	1,3562	1,7823	1,9123	2,0764	2,1788	2,3027	2,4607	2,6810	3,0545
13	0,8702	1,0795	1,3502	1,7709	1,8989	2,0600	2,1604	2,2816	2,4358	2,6503	3,0123
14	0,8681	1,0763	1,3450	1,7613	1,8875	2,0462	2,1448	2,2638	2,4149	2,6245	2,9768
15	0,8662	1,0735	1,3406	1,7531	1,8777	2,0343	2,1314	2,2485	2,3970	2,6025	2,9467
16	0,8647	1,0711	1,3368	1,7459	1,8693	2,0240	2,1199	2,2354	2,3815	2,5835	2,9208
17	0,8633	1,0690	1,3334	1,7396	1,8619	2,0150	2,1098	2,2238	2,3681	2,5669	2,8982
18	0,8620	1,0672	1,3304	1,7341	1,8553	2,0071	2,1009	2,2137	2,3562	2,5524	2,8784
19	0,8610	1,0655	1,3277	1,7291	1,8495	2,0000	2,0930	2,2047	2,3456	2,5395	2,8609
20	0,8600	1,0640	1,3253	1,7247	1,8443	1,9937	2,0860	2,1967	2,3362	2,5280	2,8453
21	0,8591	1,0627	1,3232	1,7207	1,8397	1,9880	2,0796	2,1894	2,3278	2,5176	2,8314
22	0,8583	1,0614	1,3212	1,7171	1,8354	1,9829	2,0739	2,1829	2,3202	2,5083	2,8188
23	0,8575	1,0603	1,3195	1,7139	1,8316	1,9782	2,0687	2,1770	2,3132	2,4999	2,8073
24	0,8569	1,0593	1,3178	1,7109	1,8281	1,9740	2,0639	2,1715	2,3069	2,4922	2,7969
25	0,8562	1,0584	1,3163	1,7081	1,8248	1,9701	2,0595	2,1666	2,3011	2,4851	2,7874
26	0,8557	1,0575	1,3150	1,7056	1,8219	1,9665	2,0555	2,1620	2,2958	2,4786	2,7787
27	0,8551	1,0567	1,3137	1,7033	1,8191	1,9632	2,0518	2,1578	2,2909	2,4727	2,7707
28	0,8546	1,0560	1,3125	1,7011	1,8166	1,9601	2,0484	2,1539	2,2864	2,4671	2,7633
29	0,8542	1,0553	1,3114	1,6991	1,8142	1,9573	2,0452	2,1503	2,2822	2,4620	2,7564
30	0,8538	1,0547	1,3104	1,6973	1,8120	1,9546	2,0423	2,1470	2,2783	2,4573	2,7500
31	0,8534	1,0541	1,3095	1,6955	1,8100	1,9522	2,0395	2,1438	2,2746	2,4528	2,7440
32	0,8530	1,0535	1,3086	1,6939	1,8081	1,9499	2,0369	2,1409	2,2712	2,4487	2,7385
33	0,8526	1,0530	1,3077	1,6924	1,8063	1,9477	2,0345	2,1382	2,2680	2,4448	2,7333
34	0,8523	1,0525	1,3070	1,6909	1,8046	1,9457	2,0322	2,1356	2,2650	2,4411	2,7284
35	0,8520	1,0520	1,3062	1,6896	1,8030	1,9438	2,0301	2,1332	2,2622	2,4377	2,7238
36	0,8517	1,0516	1,3055	1,6883	1,8015	1,9419	2,0281	2,1309	2,2595	2,4345	2,7195
37	0,8514	1,0512	1,3049	1,6871	1,8001	1,9402	2,0262	2,1287	2,2570	2,4314	2,7154
38	0,8512	1,0508	1,3042	1,6860	1,7988	1,9386	2,0244	2,1267	2,2546	2,4286	2,7116
39	0,8509	1,0504	1,3036	1,6849	1,7975	1,9371	2,0227	2,1247	2,2524	2,4258	2,7079
40	0,8507	1,0500	1,3031	1,6839	1,7963	1,9357	2,0211	2,1229	2,2503	2,4233	2,7045
41	0,8505	1,0497	1,3025	1,6829	1,7952	1,9343	2,0195	2,1212	2,2482	2,4208	2,7012
42	0,8503	1,0494	1,3020	1,6820	1,7941	1,9330	2,0181	2,1195	2,2463	2,4185	2,6981
43	0,8501	1,0491	1,3016	1,6811	1,7931	1,9317	2,0167	2,1179	2,2445	2,4163	2,6951
44	0,8499	1,0488	1,3011	1,6802	1,7921	1,9305	2,0154	2,1164	2,2427	2,4141	2,6923
45	0,8497	1,0485	1,3006	1,6794	1,7911	1,9294	2,0141	2,1150	2,2411	2,4121	2,6896
46	0,8495	1,0483	1,3002	1,6787	1,7902	1,9283	2,0129	2,1136	2,2395	2,4102	2,6870
47	0,8493	1,0480	1,2998	1,6779	1,7894	1,9273	2,0117	2,1123	2,2380	2,4083	2,6846
48	0,8492	1,0478	1,2994	1,6772	1,7885	1,9263	2,0106	2,1111	2,2365	2,4066	2,6822
49	0,8490	1,0475	1,2991	1,6766	1,7878	1,9253	2,0096	2,1099	2,2351	2,4049	2,6800
50	0,8489	1,0473	1,2987	1,6759	1,7870	1,9244	2,0086	2,1087	2,2338	2,4033	2,6778
51	0,8487	1,0471	1,2984	1,6753	1,7863	1,9236	2,0076	2,1076	2,2325	2,4017	2,6757
52	0,8486	1,0469	1,2980	1,6747	1,7856	1,9227	2,0066	2,1066	2,2313	2,4002	2,6737
53	0,8485	1,0467	1,2977	1,6741	1,7849	1,9219	2,0057	2,1055	2,2301	2,3988	2,6718
54	0,8483	1,0465	1,2974	1,6736	1,7843	1,9211	2,0049	2,1046	2,2289	2,3974	2,6700
55	0,8482	1,0463	1,2971	1,6730	1,7836	1,9204	2,0040	2,1036	2,2278	2,3961	2,6682
56	0,8481	1,0461	1,2969	1,6725	1,7830	1,9197	2,0032	2,1027	2,2268	2,3948	2,6665
57	0,8480	1,0459	1,2966	1,6720	1,7825	1,9190	2,0025	2,1018	2,2258	2,3936	2,6649
58	0,8479	1,0458	1,2963	1,6716	1,7819	1,9183	2,0017	2,1010	2,2248	2,3924	2,6633
59	0,8478	1,0456	1,2961	1,6711	1,7814	1,9177	2,0010	2,1002	2,2238	2,3912	2,6618
60	0,8477	1,0455	1,2958	1,6706	1,7808	1,9170	2,0003	2,0994	2,2229	2,3901	2,6603
61	0,8476	1,0453	1,2956	1,6702	1,7803	1,9164	1,9996	2,0986	2,2220	2,3890	2,6589
62	0,8475	1,0452	1,2954	1,6698	1,7799	1,9158	1,9990	2,0979	2,2212	2,3880	2,6575
63	0,8474	1,0450	1,2951	1,6694	1,7794	1,9153	1,9983	2,0971	2,2204	2,3870	2,6561
64	0,8473	1,0449	1,2949	1,6690	1,7789	1,9147	1,9977	2,0965	2,2195	2,3860	2,6549
65	0,8472	1,0448	1,2947	1,6686	1,7785	1,9142	1,9971	2,0958	2,2188	2,3851	2,6536
66	0,8471	1,0446	1,2945	1,6683	1,7781	1,9137	1,9966	2,0951	2,2180	2,3842	2,6524
67	0,8470	1,0445	1,2943	1,6679	1,7776	1,9132	1,9960	2,0945	2,2173	2,3833	2,6512
68	0,8469	1,0444	1,2941	1,6676	1,7772	1,9127	1,9955	2,0939	2,2166	2,3824	2,6501
69	0,8469	1,0443	1,2939	1,6672	1,7769	1,9122	1,9949	2,0933	2,2159	2,3816	2,6490
70	0,8468	1,0442	1,2938	1,6669	1,7765	1,9118	1,9944	2,0927	2,2152	2,3808	2,6479

FIGURE 13 – Table 3 : Loi de Student

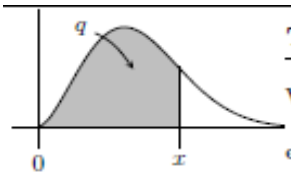
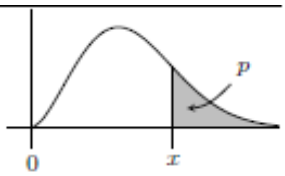


TABLE INVERSE DE LA LOI DU χ^2

Valeurs de x en fonction de q tel que $q = \mathbb{P}[\chi^2 \leqslant x]$
et de p tel que $p = \mathbb{P}[\chi^2 \geqslant x]$
en fonction du nombre de ddl du χ^2.

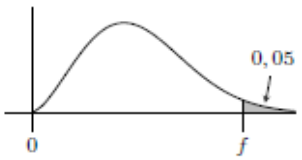


ddl \ q / p	0,005 / 0,995	0,01 / 0,99	0,02 / 0,98	0,025 / 0,975	0,05 / 0,95	0,1 / 0,9	0,9 / 0,1	0,95 / 0,05	0,975 / 0,025	0,98 / 0,02	0,99 / 0,01	0,995 / 0,005
1	0,00004	0,0002	0,001	0,001	0,004	0,016	2,706	3,841	5,024	5,412	6,635	7,879
2	0,010	0,020	0,040	0,051	0,103	0,211	4,605	5,991	7,378	7,824	9,210	10,60
3	0,072	0,115	0,185	0,216	0,352	0,584	6,251	7,815	9,348	9,837	11,34	12,84
4	0,207	0,297	0,429	0,484	0,711	1,064	7,779	9,488	11,14	11,67	13,28	14,86
5	0,412	0,554	0,752	0,831	1,145	1,610	9,236	11,07	12,83	13,39	15,09	16,75
6	0,676	0,872	1,134	1,237	1,635	2,204	10,64	12,59	14,45	15,03	16,81	18,55
7	0,989	1,239	1,564	1,690	2,167	2,833	12,02	14,07	16,01	16,62	18,48	20,28
8	1,344	1,646	2,032	2,180	2,733	3,490	13,36	15,51	17,53	18,17	20,09	21,95
9	1,735	2,088	2,532	2,700	3,325	4,168	14,68	16,92	19,02	19,68	21,67	23,59
10	2,156	2,558	3,059	3,247	3,940	4,865	15,99	18,31	20,48	21,16	23,21	25,19
11	2,603	3,053	3,609	3,816	4,575	5,578	17,28	19,68	21,92	22,62	24,72	26,76
12	3,074	3,571	4,178	4,404	5,226	6,304	18,55	21,03	23,34	24,05	26,22	28,30
13	3,565	4,107	4,765	5,009	5,892	7,042	19,81	22,36	24,74	25,47	27,69	29,82
14	4,075	4,660	5,368	5,629	6,571	7,790	21,06	23,68	26,12	26,87	29,14	31,32
15	4,601	5,229	5,985	6,262	7,261	8,547	22,31	25,00	27,49	28,26	30,58	32,80
16	5,142	5,812	6,614	6,908	7,962	9,312	23,54	26,30	28,85	29,63	32,00	34,27
17	5,697	6,408	7,255	7,564	8,672	10,09	24,77	27,59	30,19	31,00	33,41	35,72
18	6,265	7,015	7,906	8,231	9,390	10,86	25,99	28,87	31,53	32,35	34,81	37,16
19	6,844	7,633	8,567	8,907	10,12	11,65	27,20	30,14	32,85	33,69	36,19	38,58
20	7,434	8,260	9,237	9,591	10,85	12,44	28,41	31,41	34,17	35,02	37,57	40,00
21	8,034	8,897	9,915	10,28	11,59	13,24	29,62	32,67	35,48	36,34	38,93	41,40
22	8,643	9,542	10,60	10,98	12,34	14,04	30,81	33,92	36,78	37,66	40,29	42,80
23	9,260	10,20	11,29	11,69	13,09	14,85	32,01	35,17	38,08	38,97	41,64	44,18
24	9,886	10,86	11,99	12,40	13,85	15,66	33,20	36,42	39,36	40,27	42,98	45,56
25	10,52	11,52	12,70	13,12	14,61	16,47	34,38	37,65	40,65	41,57	44,31	46,93
26	11,16	12,20	13,41	13,84	15,38	17,29	35,56	38,89	41,92	42,86	45,64	48,29
27	11,81	12,88	14,13	14,57	16,15	18,11	36,74	40,11	43,19	44,14	46,96	49,64
28	12,46	13,56	14,85	15,31	16,93	18,94	37,92	41,34	44,46	45,42	48,28	50,99
29	13,12	14,26	15,57	16,05	17,71	19,77	39,09	42,56	45,72	46,69	49,59	52,34
30	13,79	14,95	16,31	16,79	18,49	20,60	40,26	43,77	46,98	47,96	50,89	53,67
31	14,46	15,66	17,04	17,54	19,28	21,43	41,42	44,99	48,23	49,23	52,19	55,00
32	15,13	16,36	17,78	18,29	20,07	22,27	42,58	46,19	49,48	50,49	53,49	56,33
33	15,82	17,07	18,53	19,05	20,87	23,11	43,75	47,40	50,73	51,74	54,78	57,65
34	16,50	17,79	19,28	19,81	21,66	23,95	44,90	48,60	51,97	53,00	56,06	58,96
35	17,19	18,51	20,03	20,57	22,47	24,80	46,06	49,80	53,20	54,24	57,34	60,27
36	17,89	19,23	20,78	21,34	23,27	25,64	47,21	51,00	54,44	55,49	58,62	61,58
37	18,59	19,96	21,54	22,11	24,07	26,49	48,36	52,19	55,67	56,73	59,89	62,88
38	19,29	20,69	22,30	22,88	24,88	27,34	49,51	53,38	56,90	57,97	61,16	64,18
39	20,00	21,43	23,07	23,65	25,70	28,20	50,66	54,57	58,12	59,20	62,43	65,48
40	20,71	22,16	23,84	24,43	26,51	29,05	51,81	55,76	59,34	60,44	63,69	66,77
45	24,31	25,90	27,72	28,37	30,61	33,35	57,51	61,66	65,41	66,56	69,96	73,17
50	27,99	29,71	31,66	32,36	34,76	37,69	63,17	67,50	71,42	72,61	76,15	79,49
60	35,53	37,48	39,70	40,48	43,19	46,46	74,40	79,08	83,30	84,58	88,38	91,95
70	43,28	45,44	47,89	48,76	51,74	55,33	85,53	90,53	95,02	96,39	100,4	104,2
80	51,17	53,54	56,21	57,15	60,39	64,28	96,58	101,9	106,6	108,1	112,3	116,3
90	59,20	61,75	64,63	65,65	69,13	73,29	107,6	113,1	118,1	119,6	124,1	128,3
100	67,33	70,06	73,14	74,22	77,93	82,36	118,5	124,3	129,6	131,1	135,8	140,2
110	75,55	78,46	81,72	82,87	86,79	91,47	129,4	135,5	140,9	142,6	147,4	151,9
120	83,85	86,92	90,37	91,57	95,70	100,6	140,2	146,6	152,2	153,9	159,0	163,6
130	92,22	95,45	99,07	100,3	104,7	109,8	151,0	157,6	163,5	165,2	170,4	175,3
140	100,7	104,0	107,8	109,1	113,7	119,0	161,8	168,6	174,6	176,5	181,8	186,8
150	109,1	112,7	116,6	118,0	122,7	128,3	172,6	179,6	185,8	187,7	193,2	198,4

FIGURE 14 – Table 4 : Loi du Khi-deux

Valeurs de f telles que $\mathbb{P}[F \geqslant f] = 0,05$

où F suit la loi de Fisher-Snedecor à ν_1, ν_2 degrés de liberté
ν_1 : nombre de ddl du numerateur
ν_2 : nombre de ddl du denominateur

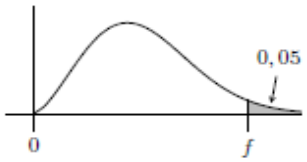


ν_2 \ ν_1	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
5	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66	4,64	4,62	4,60	4,59	4,58	4,57	4,56	4,55	4,54	4,53	4,53	4,52
6	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98	3,96	3,94	3,92	3,91	3,90	3,88	3,87	3,86	3,86	3,85	3,84	3,83
7	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55	3,53	3,51	3,49	3,48	3,47	3,46	3,44	3,43	3,43	3,42	3,41	3,40
8	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26	3,24	3,22	3,20	3,19	3,17	3,16	3,15	3,14	3,13	3,12	3,12	3,11
9	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05	3,03	3,01	2,99	2,97	2,96	2,95	2,94	2,93	2,92	2,91	2,90	2,89
10	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,89	2,86	2,85	2,83	2,81	2,80	2,79	2,77	2,76	2,75	2,75	2,74	2,73
11	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76	2,74	2,72	2,70	2,69	2,67	2,66	2,65	2,64	2,63	2,62	2,61	2,60
12	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66	2,64	2,62	2,60	2,58	2,57	2,56	2,54	2,53	2,52	2,51	2,51	2,50
13	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58	2,55	2,53	2,51	2,50	2,48	2,47	2,46	2,45	2,44	2,43	2,42	2,41
14	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51	2,48	2,46	2,44	2,43	2,41	2,40	2,39	2,38	2,37	2,36	2,35	2,34
15	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45	2,42	2,40	2,38	2,37	2,35	2,34	2,33	2,32	2,31	2,30	2,29	2,28
16	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40	2,37	2,35	2,33	2,32	2,30	2,29	2,28	2,26	2,25	2,24	2,24	2,23
17	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35	2,33	2,31	2,29	2,27	2,26	2,24	2,23	2,22	2,21	2,20	2,19	2,18
18	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31	2,29	2,27	2,25	2,23	2,22	2,20	2,19	2,18	2,17	2,16	2,15	2,14
19	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28	2,26	2,23	2,21	2,20	2,18	2,17	2,16	2,14	2,13	2,12	2,11	2,11
20	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25	2,22	2,20	2,18	2,17	2,15	2,14	2,12	2,11	2,10	2,09	2,08	2,07
21	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	2,22	2,20	2,18	2,16	2,14	2,12	2,11	2,10	2,08	2,07	2,06	2,05	2,05
22	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23	2,20	2,17	2,15	2,13	2,11	2,10	2,08	2,07	2,06	2,05	2,04	2,03	2,02
23	2,64	2,53	2,44	2,37	2,32	2,27	2,24	2,20	2,18	2,15	2,13	2,11	2,09	2,08	2,06	2,05	2,04	2,02	2,01	2,01	2,00
24	2,62	2,51	2,42	2,36	2,30	2,25	2,22	2,18	2,15	2,13	2,11	2,09	2,07	2,05	2,04	2,03	2,01	2,00	1,99	1,98	1,97
25	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16	2,14	2,11	2,09	2,07	2,05	2,04	2,02	2,01	2,00	1,98	1,97	1,96	1,96
26	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,12	2,09	2,07	2,05	2,03	2,02	2,00	1,99	1,98	1,97	1,96	1,95	1,94
27	2,57	2,46	2,37	2,31	2,25	2,20	2,17	2,13	2,10	2,08	2,06	2,04	2,02	2,00	1,99	1,97	1,96	1,95	1,94	1,93	1,92
28	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12	2,09	2,06	2,04	2,02	2,00	1,99	1,97	1,96	1,95	1,93	1,92	1,91	1,91
29	2,55	2,43	2,35	2,28	2,22	2,18	2,14	2,10	2,08	2,05	2,03	2,01	1,99	1,97	1,96	1,94	1,93	1,92	1,91	1,90	1,89
30	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,06	2,04	2,01	1,99	1,98	1,96	1,95	1,93	1,92	1,91	1,90	1,89	1,88
31	2,52	2,41	2,32	2,25	2,20	2,15	2,11	2,08	2,05	2,03	2,00	1,98	1,96	1,95	1,93	1,92	1,91	1,90	1,88	1,88	1,87
32	2,51	2,40	2,31	2,24	2,19	2,14	2,10	2,07	2,04	2,01	1,99	1,97	1,95	1,94	1,92	1,91	1,90	1,88	1,87	1,86	1,85
33	2,50	2,39	2,30	2,23	2,18	2,13	2,09	2,06	2,03	2,00	1,98	1,96	1,94	1,93	1,91	1,90	1,89	1,87	1,86	1,85	1,84
34	2,49	2,38	2,29	2,23	2,17	2,12	2,08	2,05	2,02	1,99	1,97	1,95	1,93	1,92	1,90	1,89	1,88	1,86	1,85	1,84	1,83
35	2,49	2,37	2,29	2,22	2,16	2,11	2,07	2,04	2,01	1,99	1,96	1,94	1,92	1,91	1,89	1,88	1,87	1,85	1,84	1,83	1,82
36	2,48	2,36	2,28	2,21	2,15	2,11	2,07	2,03	2,00	1,98	1,95	1,93	1,92	1,90	1,88	1,87	1,86	1,85	1,83	1,82	1,81
37	2,47	2,36	2,27	2,20	2,14	2,10	2,06	2,02	2,00	1,97	1,95	1,93	1,91	1,89	1,88	1,86	1,85	1,84	1,83	1,82	1,81
38	2,46	2,35	2,26	2,19	2,14	2,09	2,05	2,02	1,99	1,96	1,94	1,92	1,90	1,88	1,87	1,85	1,84	1,83	1,82	1,81	1,80
39	2,46	2,34	2,26	2,19	2,13	2,08	2,04	2,01	1,98	1,95	1,93	1,91	1,89	1,88	1,86	1,85	1,83	1,82	1,81	1,80	1,79
40	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,97	1,95	1,92	1,90	1,89	1,87	1,85	1,84	1,83	1,81	1,80	1,79	1,78
41	2,44	2,33	2,24	2,17	2,12	2,07	2,03	2,00	1,97	1,94	1,92	1,90	1,88	1,86	1,85	1,83	1,82	1,81	1,80	1,79	1,78
42	2,44	2,32	2,24	2,17	2,11	2,06	2,03	1,99	1,96	1,94	1,91	1,89	1,87	1,86	1,84	1,83	1,81	1,80	1,79	1,78	1,77
43	2,43	2,32	2,23	2,16	2,11	2,06	2,02	1,99	1,96	1,93	1,91	1,89	1,87	1,85	1,83	1,82	1,81	1,79	1,78	1,77	1,76
44	2,43	2,31	2,23	2,16	2,10	2,05	2,01	1,98	1,95	1,92	1,90	1,88	1,86	1,84	1,83	1,81	1,80	1,79	1,78	1,77	1,76
45	2,42	2,31	2,22	2,15	2,10	2,05	2,01	1,97	1,94	1,92	1,89	1,87	1,86	1,84	1,82	1,81	1,80	1,78	1,77	1,76	1,75
46	2,42	2,30	2,22	2,15	2,09	2,04	2,00	1,97	1,94	1,91	1,89	1,87	1,85	1,83	1,82	1,80	1,79	1,78	1,77	1,76	1,75
47	2,41	2,30	2,21	2,14	2,09	2,04	2,00	1,96	1,93	1,91	1,88	1,86	1,84	1,83	1,81	1,80	1,78	1,77	1,76	1,75	1,74
48	2,41	2,29	2,21	2,14	2,08	2,03	1,99	1,96	1,93	1,90	1,88	1,86	1,84	1,82	1,81	1,79	1,78	1,77	1,76	1,75	1,74
49	2,40	2,29	2,20	2,13	2,08	2,03	1,99	1,96	1,93	1,90	1,88	1,85	1,84	1,82	1,80	1,79	1,78	1,76	1,75	1,74	1,73
50	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95	1,92	1,89	1,87	1,85	1,83	1,81	1,80	1,78	1,77	1,76	1,75	1,74	1,73
55	2,38	2,27	2,18	2,11	2,06	2,01	1,97	1,93	1,90	1,88	1,85	1,83	1,81	1,79	1,78	1,76	1,75	1,74	1,73	1,72	1,71
60	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,89	1,86	1,84	1,82	1,80	1,78	1,76	1,75	1,73	1,72	1,71	1,70	1,69
65	2,36	2,24	2,15	2,08	2,03	1,98	1,94	1,90	1,87	1,85	1,82	1,80	1,78	1,76	1,75	1,73	1,72	1,71	1,70	1,69	1,68
70	2,35	2,23	2,14	2,07	2,02	1,97	1,93	1,89	1,86	1,84	1,81	1,79	1,77	1,75	1,74	1,72	1,71	1,70	1,68	1,67	1,66
75	2,34	2,22	2,13	2,06	2,01	1,96	1,92	1,88	1,85	1,83	1,80	1,78	1,76	1,74	1,73	1,71	1,70	1,69	1,67	1,66	1,65
80	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88	1,84	1,82	1,79	1,77	1,75	1,73	1,72	1,70	1,69	1,68	1,67	1,65	1,64

Figure 15 – Table 5a : Loi de Fisher-Snedecor

VALEURS DE f TELLES QUE $\mathbb{P}[F \geqslant f] = 0,05$

où F suit la loi de Fisher-Snedecor à ν_1, ν_2 degrés de liberté
ν_1 : nombre de ddl du numérateur
ν_2 : nombre de ddl du dénominateur

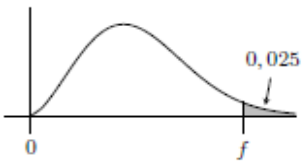


ν_2 \ ν_1	26	27	28	29	30	31	32	33	34	35	40	42	43	45	50	55	60	65	70	75	80
5	4,52	4,51	4,50	4,50	4,50	4,49	4,49	4,48	4,48	4,48	4,46	4,46	4,46	4,45	4,44	4,44	4,43	4,43	4,42	4,42	4,41
6	3,83	3,82	3,82	3,81	3,81	3,80	3,80	3,80	3,79	3,79	3,77	3,77	3,77	3,76	3,75	3,75	3,74	3,73	3,73	3,73	3,72
7	3,40	3,39	3,39	3,38	3,38	3,37	3,37	3,36	3,36	3,36	3,34	3,34	3,33	3,33	3,32	3,31	3,30	3,30	3,29	3,29	3,29
8	3,10	3,10	3,09	3,08	3,08	3,07	3,07	3,07	3,06	3,06	3,04	3,04	3,04	3,03	3,02	3,01	3,01	3,00	2,99	2,99	2,99
9	2,89	2,88	2,87	2,87	2,86	2,86	2,85	2,85	2,85	2,84	2,83	2,82	2,82	2,81	2,80	2,79	2,79	2,78	2,78	2,77	2,77
10	2,72	2,72	2,71	2,70	2,70	2,69	2,69	2,69	2,68	2,68	2,66	2,66	2,65	2,65	2,64	2,63	2,62	2,61	2,61	2,60	2,60
11	2,59	2,59	2,58	2,58	2,57	2,57	2,56	2,56	2,55	2,55	2,53	2,53	2,52	2,52	2,51	2,50	2,49	2,48	2,48	2,47	2,47
12	2,49	2,48	2,48	2,47	2,47	2,46	2,46	2,45	2,45	2,44	2,43	2,42	2,42	2,41	2,40	2,39	2,38	2,38	2,37	2,37	2,36
13	2,41	2,40	2,39	2,39	2,38	2,38	2,37	2,37	2,36	2,36	2,34	2,33	2,33	2,33	2,31	2,30	2,30	2,29	2,28	2,28	2,27
14	2,33	2,33	2,32	2,31	2,31	2,30	2,30	2,29	2,29	2,28	2,27	2,26	2,26	2,25	2,24	2,23	2,22	2,22	2,21	2,21	2,20
15	2,27	2,27	2,26	2,25	2,25	2,24	2,24	2,23	2,23	2,22	2,20	2,20	2,20	2,19	2,18	2,17	2,16	2,15	2,15	2,14	2,14
16	2,22	2,21	2,21	2,20	2,19	2,19	2,18	2,18	2,17	2,17	2,15	2,14	2,14	2,14	2,12	2,11	2,11	2,10	2,09	2,09	2,08
17	2,17	2,17	2,16	2,15	2,15	2,14	2,14	2,13	2,13	2,12	2,10	2,10	2,09	2,09	2,08	2,07	2,06	2,05	2,05	2,04	2,03
18	2,13	2,13	2,12	2,11	2,11	2,10	2,10	2,09	2,09	2,08	2,06	2,06	2,05	2,05	2,04	2,03	2,02	2,01	2,00	2,00	1,99
19	2,10	2,09	2,08	2,08	2,07	2,07	2,06	2,06	2,05	2,05	2,03	2,02	2,02	2,01	2,00	1,99	1,98	1,97	1,97	1,96	1,96
20	2,07	2,06	2,05	2,05	2,04	2,03	2,03	2,02	2,02	2,01	1,99	1,99	1,98	1,98	1,97	1,96	1,95	1,94	1,93	1,93	1,92
21	2,04	2,03	2,02	2,02	2,01	2,00	2,00	1,99	1,99	1,98	1,96	1,96	1,95	1,95	1,94	1,93	1,92	1,91	1,90	1,90	1,89
22	2,01	2,00	2,00	1,99	1,98	1,98	1,97	1,97	1,96	1,96	1,94	1,93	1,93	1,92	1,91	1,90	1,89	1,88	1,88	1,87	1,86
23	1,99	1,98	1,97	1,97	1,96	1,95	1,95	1,94	1,94	1,93	1,91	1,91	1,90	1,90	1,88	1,87	1,86	1,86	1,85	1,84	1,84
24	1,97	1,96	1,95	1,95	1,94	1,93	1,93	1,92	1,92	1,91	1,89	1,89	1,88	1,88	1,86	1,85	1,84	1,83	1,83	1,82	1,82
25	1,95	1,94	1,93	1,93	1,92	1,91	1,91	1,90	1,90	1,89	1,87	1,86	1,86	1,86	1,84	1,83	1,82	1,81	1,81	1,80	1,80
26	1,93	1,92	1,91	1,91	1,90	1,89	1,89	1,88	1,88	1,87	1,85	1,85	1,84	1,84	1,82	1,81	1,80	1,79	1,79	1,78	1,78
27	1,91	1,90	1,90	1,89	1,88	1,88	1,87	1,87	1,86	1,86	1,84	1,83	1,83	1,82	1,81	1,79	1,79	1,78	1,77	1,76	1,76
28	1,90	1,89	1,88	1,88	1,87	1,86	1,86	1,85	1,85	1,84	1,82	1,81	1,81	1,80	1,79	1,78	1,77	1,76	1,75	1,75	1,74
29	1,88	1,88	1,87	1,86	1,85	1,85	1,84	1,84	1,83	1,83	1,81	1,80	1,79	1,79	1,77	1,76	1,75	1,75	1,74	1,73	1,73
30	1,87	1,86	1,85	1,85	1,84	1,83	1,83	1,82	1,82	1,81	1,79	1,78	1,78	1,77	1,76	1,75	1,74	1,73	1,72	1,72	1,71
31	1,86	1,85	1,84	1,83	1,83	1,82	1,82	1,81	1,81	1,80	1,78	1,77	1,77	1,76	1,75	1,74	1,73	1,72	1,71	1,70	1,70
32	1,85	1,84	1,83	1,82	1,82	1,81	1,80	1,80	1,79	1,79	1,77	1,76	1,76	1,75	1,74	1,72	1,71	1,71	1,70	1,69	1,69
33	1,83	1,83	1,82	1,81	1,81	1,80	1,79	1,79	1,78	1,78	1,76	1,75	1,74	1,74	1,72	1,71	1,70	1,69	1,69	1,68	1,67
34	1,82	1,82	1,81	1,80	1,80	1,79	1,78	1,78	1,77	1,77	1,75	1,74	1,73	1,73	1,71	1,70	1,69	1,68	1,68	1,67	1,66
35	1,82	1,81	1,80	1,79	1,79	1,78	1,77	1,77	1,76	1,76	1,74	1,73	1,72	1,72	1,70	1,69	1,68	1,67	1,66	1,66	1,65
36	1,81	1,80	1,79	1,78	1,78	1,77	1,76	1,76	1,75	1,75	1,73	1,72	1,71	1,71	1,69	1,68	1,67	1,66	1,66	1,65	1,64
37	1,80	1,79	1,78	1,77	1,77	1,76	1,76	1,75	1,74	1,74	1,72	1,71	1,71	1,70	1,68	1,67	1,66	1,65	1,65	1,64	1,63
38	1,79	1,78	1,77	1,77	1,76	1,75	1,75	1,74	1,74	1,73	1,71	1,70	1,70	1,69	1,68	1,66	1,65	1,64	1,64	1,63	1,62
39	1,78	1,77	1,77	1,76	1,75	1,75	1,74	1,73	1,73	1,72	1,70	1,69	1,69	1,68	1,67	1,66	1,65	1,64	1,63	1,62	1,62
40	1,77	1,77	1,76	1,75	1,74	1,74	1,73	1,73	1,72	1,72	1,69	1,69	1,68	1,67	1,66	1,65	1,64	1,63	1,62	1,61	1,61
41	1,77	1,76	1,75	1,74	1,74	1,73	1,72	1,72	1,71	1,71	1,69	1,68	1,67	1,67	1,65	1,64	1,63	1,62	1,61	1,61	1,60
42	1,76	1,75	1,75	1,74	1,73	1,72	1,72	1,71	1,71	1,70	1,68	1,67	1,67	1,66	1,65	1,63	1,62	1,61	1,61	1,60	1,59
43	1,75	1,75	1,74	1,73	1,72	1,72	1,71	1,71	1,70	1,70	1,67	1,66	1,66	1,65	1,64	1,63	1,62	1,61	1,60	1,59	1,59
44	1,75	1,74	1,73	1,73	1,72	1,71	1,71	1,70	1,69	1,69	1,67	1,66	1,65	1,65	1,63	1,62	1,61	1,60	1,59	1,59	1,58
45	1,74	1,73	1,73	1,72	1,71	1,71	1,70	1,69	1,69	1,68	1,66	1,65	1,65	1,64	1,63	1,61	1,60	1,59	1,59	1,58	1,57
46	1,74	1,73	1,72	1,71	1,71	1,70	1,69	1,69	1,68	1,68	1,65	1,65	1,64	1,64	1,62	1,61	1,60	1,59	1,58	1,57	1,57
47	1,73	1,72	1,72	1,71	1,70	1,70	1,69	1,68	1,68	1,67	1,65	1,64	1,64	1,63	1,61	1,60	1,59	1,58	1,57	1,57	1,56
48	1,73	1,72	1,71	1,70	1,70	1,69	1,68	1,68	1,67	1,67	1,64	1,64	1,63	1,62	1,61	1,60	1,59	1,58	1,57	1,56	1,56
49	1,72	1,71	1,71	1,70	1,69	1,69	1,68	1,67	1,67	1,66	1,64	1,63	1,63	1,62	1,60	1,59	1,58	1,57	1,56	1,56	1,55
50	1,72	1,71	1,70	1,69	1,69	1,68	1,67	1,67	1,66	1,66	1,63	1,63	1,62	1,61	1,60	1,59	1,58	1,57	1,56	1,55	1,54
55	1,70	1,69	1,68	1,67	1,67	1,66	1,65	1,65	1,64	1,64	1,61	1,60	1,60	1,59	1,58	1,56	1,55	1,54	1,54	1,53	1,52
60	1,68	1,67	1,66	1,66	1,65	1,64	1,64	1,63	1,62	1,62	1,59	1,59	1,58	1,57	1,56	1,55	1,53	1,52	1,52	1,51	1,50
65	1,67	1,66	1,65	1,64	1,63	1,63	1,62	1,61	1,61	1,60	1,58	1,57	1,57	1,56	1,54	1,53	1,52	1,51	1,50	1,49	1,49
70	1,65	1,65	1,64	1,63	1,62	1,62	1,61	1,60	1,60	1,59	1,57	1,56	1,55	1,55	1,53	1,52	1,50	1,49	1,49	1,48	1,47
75	1,64	1,63	1,63	1,62	1,61	1,60	1,60	1,59	1,59	1,58	1,55	1,55	1,54	1,53	1,52	1,50	1,49	1,48	1,47	1,47	1,46
80	1,63	1,63	1,62	1,61	1,60	1,59	1,59	1,58	1,58	1,57	1,54	1,54	1,53	1,52	1,51	1,49	1,48	1,47	1,46	1,45	1,45

FIGURE 16 – Table 5b : Loi de Fisher-Snedecor

VALEURS DE f TELLES QUE $\mathbb{P}[F \geqslant f] = 0,025$

où F suit la loi de Fisher-Snedecor à ν_1, ν_2 degrés de liberté
ν_1 : nombre de ddl du numérateur
ν_2 : nombre de ddl du dénominateur

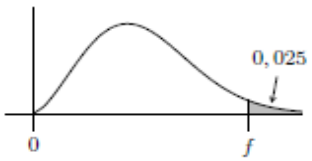


ν_2 \ ν_1	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
5	7,15	6,98	6,85	6,76	6,68	6,62	6,57	6,52	6,49	6,46	6,43	6,40	6,38	6,36	6,34	6,33	6,31	6,30	6,29	6,28	6,27
6	5,99	5,82	5,70	5,60	5,52	5,46	5,41	5,37	5,33	5,30	5,27	5,24	5,22	5,20	5,18	5,17	5,15	5,14	5,13	5,12	5,11
7	5,29	5,12	4,99	4,90	4,82	4,76	4,71	4,67	4,63	4,60	4,57	4,54	4,52	4,50	4,48	4,47	4,45	4,44	4,43	4,41	4,40
8	4,82	4,65	4,53	4,43	4,36	4,30	4,24	4,20	4,16	4,13	4,10	4,08	4,05	4,03	4,02	4,00	3,98	3,97	3,96	3,95	3,94
9	4,48	4,32	4,20	4,10	4,03	3,96	3,91	3,87	3,83	3,80	3,77	3,74	3,72	3,70	3,68	3,67	3,65	3,64	3,63	3,61	3,60
10	4,24	4,07	3,95	3,85	3,78	3,72	3,66	3,62	3,58	3,55	3,52	3,50	3,47	3,45	3,44	3,42	3,40	3,39	3,38	3,37	3,35
11	4,04	3,88	3,76	3,66	3,59	3,53	3,47	3,43	3,39	3,36	3,33	3,30	3,28	3,26	3,24	3,23	3,21	3,20	3,18	3,17	3,16
12	3,89	3,73	3,61	3,51	3,44	3,37	3,32	3,28	3,24	3,21	3,18	3,15	3,13	3,11	3,09	3,07	3,06	3,04	3,03	3,02	3,01
13	3,77	3,60	3,48	3,39	3,31	3,25	3,20	3,15	3,12	3,08	3,05	3,03	3,00	2,98	2,96	2,95	2,93	2,92	2,91	2,89	2,88
14	3,66	3,50	3,38	3,29	3,21	3,15	3,09	3,05	3,01	2,98	2,95	2,92	2,90	2,88	2,86	2,84	2,83	2,81	2,80	2,79	2,78
15	3,58	3,41	3,29	3,20	3,12	3,06	3,01	2,96	2,92	2,89	2,86	2,84	2,81	2,79	2,77	2,76	2,74	2,73	2,71	2,70	2,69
16	3,50	3,34	3,22	3,12	3,05	2,99	2,93	2,89	2,85	2,82	2,79	2,76	2,74	2,72	2,70	2,68	2,67	2,65	2,64	2,63	2,61
17	3,44	3,28	3,16	3,06	2,98	2,92	2,87	2,82	2,79	2,75	2,72	2,70	2,67	2,65	2,63	2,62	2,60	2,59	2,57	2,56	2,55
18	3,38	3,22	3,10	3,01	2,93	2,87	2,81	2,77	2,73	2,70	2,67	2,64	2,62	2,60	2,58	2,56	2,54	2,53	2,52	2,50	2,49
19	3,33	3,17	3,05	2,96	2,88	2,82	2,76	2,72	2,68	2,65	2,62	2,59	2,57	2,55	2,53	2,51	2,49	2,48	2,46	2,45	2,44
20	3,29	3,13	3,01	2,91	2,84	2,77	2,72	2,68	2,64	2,60	2,57	2,55	2,52	2,50	2,48	2,46	2,45	2,43	2,42	2,41	2,40
21	3,25	3,09	2,97	2,87	2,80	2,73	2,68	2,64	2,60	2,56	2,53	2,51	2,48	2,46	2,44	2,42	2,41	2,39	2,38	2,37	2,36
22	3,22	3,05	2,93	2,84	2,76	2,70	2,65	2,60	2,56	2,53	2,50	2,47	2,45	2,43	2,41	2,39	2,37	2,36	2,34	2,33	2,32
23	3,18	3,02	2,90	2,81	2,73	2,67	2,62	2,57	2,53	2,50	2,47	2,44	2,42	2,39	2,37	2,36	2,34	2,33	2,31	2,30	2,29
24	3,15	2,99	2,87	2,78	2,70	2,64	2,59	2,54	2,50	2,47	2,44	2,41	2,39	2,36	2,35	2,33	2,31	2,30	2,28	2,27	2,26
25	3,13	2,97	2,85	2,75	2,68	2,61	2,56	2,51	2,48	2,44	2,41	2,38	2,36	2,34	2,32	2,30	2,28	2,27	2,26	2,24	2,23
26	3,10	2,94	2,82	2,73	2,65	2,59	2,54	2,49	2,45	2,42	2,39	2,36	2,34	2,31	2,29	2,28	2,26	2,24	2,23	2,22	2,21
27	3,08	2,92	2,80	2,71	2,63	2,57	2,51	2,47	2,43	2,39	2,36	2,34	2,31	2,29	2,27	2,25	2,24	2,22	2,21	2,19	2,18
28	3,06	2,90	2,78	2,69	2,61	2,55	2,49	2,45	2,41	2,37	2,34	2,32	2,29	2,27	2,25	2,23	2,22	2,20	2,19	2,17	2,16
29	3,04	2,88	2,76	2,67	2,59	2,53	2,48	2,43	2,39	2,36	2,32	2,30	2,27	2,25	2,23	2,21	2,20	2,18	2,17	2,15	2,14
30	3,03	2,87	2,75	2,65	2,57	2,51	2,46	2,41	2,37	2,34	2,31	2,28	2,26	2,23	2,21	2,20	2,18	2,16	2,15	2,14	2,12
31	3,01	2,85	2,73	2,64	2,56	2,50	2,44	2,40	2,36	2,32	2,29	2,26	2,24	2,22	2,20	2,18	2,16	2,15	2,13	2,12	2,11
32	3,00	2,84	2,71	2,62	2,54	2,48	2,43	2,38	2,34	2,31	2,28	2,25	2,22	2,20	2,18	2,16	2,15	2,13	2,12	2,10	2,09
33	2,98	2,82	2,70	2,61	2,53	2,47	2,41	2,37	2,33	2,29	2,26	2,23	2,21	2,19	2,17	2,15	2,13	2,12	2,10	2,09	2,08
34	2,97	2,81	2,69	2,59	2,52	2,45	2,40	2,35	2,31	2,28	2,25	2,22	2,20	2,17	2,15	2,13	2,12	2,10	2,09	2,07	2,06
35	2,96	2,80	2,68	2,58	2,50	2,44	2,39	2,34	2,30	2,27	2,23	2,21	2,18	2,16	2,14	2,12	2,10	2,09	2,07	2,06	2,05
36	2,94	2,78	2,66	2,57	2,49	2,43	2,37	2,33	2,29	2,25	2,22	2,20	2,17	2,15	2,13	2,11	2,09	2,08	2,06	2,05	2,04
37	2,93	2,77	2,65	2,56	2,48	2,42	2,36	2,32	2,28	2,24	2,21	2,18	2,16	2,14	2,12	2,10	2,08	2,07	2,05	2,04	2,03
38	2,92	2,76	2,64	2,55	2,47	2,41	2,35	2,31	2,27	2,23	2,20	2,17	2,15	2,13	2,11	2,09	2,07	2,05	2,04	2,03	2,01
39	2,91	2,75	2,63	2,54	2,46	2,40	2,34	2,30	2,26	2,22	2,19	2,16	2,14	2,12	2,10	2,08	2,06	2,04	2,03	2,02	2,00
40	2,90	2,74	2,62	2,53	2,45	2,39	2,33	2,29	2,25	2,21	2,18	2,15	2,13	2,11	2,09	2,07	2,05	2,03	2,02	2,01	1,99
41	2,89	2,74	2,62	2,52	2,44	2,38	2,33	2,28	2,24	2,20	2,17	2,15	2,12	2,10	2,08	2,06	2,04	2,03	2,01	2,00	1,99
42	2,89	2,73	2,61	2,51	2,43	2,37	2,32	2,27	2,23	2,20	2,16	2,14	2,11	2,09	2,07	2,05	2,03	2,02	2,00	1,99	1,98
43	2,88	2,72	2,60	2,50	2,43	2,36	2,31	2,26	2,22	2,19	2,16	2,13	2,10	2,08	2,06	2,04	2,02	2,01	1,99	1,98	1,97
44	2,87	2,71	2,59	2,50	2,42	2,36	2,30	2,26	2,22	2,18	2,15	2,12	2,10	2,07	2,05	2,03	2,02	2,00	1,99	1,97	1,96
45	2,86	2,70	2,58	2,49	2,41	2,35	2,29	2,25	2,21	2,17	2,14	2,11	2,09	2,07	2,04	2,03	2,01	1,99	1,98	1,96	1,95
46	2,86	2,70	2,58	2,48	2,41	2,34	2,29	2,24	2,20	2,17	2,13	2,11	2,08	2,06	2,04	2,02	2,00	1,99	1,97	1,96	1,94
47	2,85	2,69	2,57	2,48	2,40	2,33	2,28	2,23	2,19	2,16	2,13	2,10	2,07	2,05	2,03	2,01	1,99	1,98	1,96	1,95	1,94
48	2,84	2,69	2,56	2,47	2,39	2,33	2,27	2,23	2,19	2,15	2,12	2,09	2,07	2,05	2,02	2,01	1,99	1,97	1,96	1,94	1,93
49	2,84	2,68	2,56	2,46	2,39	2,32	2,27	2,22	2,18	2,15	2,11	2,09	2,06	2,04	2,02	2,00	1,98	1,97	1,95	1,94	1,92
50	2,83	2,67	2,55	2,46	2,38	2,32	2,26	2,22	2,18	2,14	2,11	2,08	2,06	2,03	2,01	1,99	1,98	1,96	1,95	1,93	1,92
55	2,81	2,65	2,53	2,43	2,36	2,29	2,24	2,19	2,15	2,11	2,08	2,05	2,03	2,01	1,99	1,97	1,95	1,93	1,92	1,90	1,89
60	2,79	2,63	2,51	2,41	2,33	2,27	2,22	2,17	2,13	2,09	2,06	2,03	2,01	1,98	1,96	1,94	1,93	1,91	1,90	1,88	1,87
65	2,77	2,61	2,49	2,39	2,32	2,25	2,20	2,15	2,11	2,07	2,04	2,01	1,99	1,97	1,95	1,93	1,91	1,89	1,88	1,86	1,85
70	2,75	2,59	2,47	2,38	2,30	2,24	2,18	2,14	2,10	2,06	2,03	2,00	1,97	1,95	1,93	1,91	1,89	1,88	1,86	1,85	1,83
75	2,74	2,58	2,46	2,37	2,29	2,22	2,17	2,12	2,08	2,05	2,01	1,99	1,96	1,94	1,92	1,90	1,88	1,86	1,85	1,83	1,82
80	2,73	2,57	2,45	2,35	2,28	2,21	2,16	2,11	2,07	2,03	2,00	1,97	1,95	1,92	1,90	1,88	1,87	1,85	1,83	1,82	1,81

FIGURE 17 – Table 5c : Loi de Fisher-Snedecor

VALEURS DE f TELLES QUE $\mathbb{P}[F \geqslant f] = 0,025$

où F suit la loi de Fisher-Snedecor à ν_1, ν_2 degrés de liberté
ν_1 : nombre de ddl du numerateur
ν_2 : nombre de ddl du denominateur



ν_2 \ ν_1	26	27	28	29	30	31	32	33	34	35	40	42	43	45	50	55	60	65	70	75	80
5	6,26	6,25	6,24	6,23	6,23	6,22	6,21	6,21	6,20	6,20	6,18	6,17	6,16	6,16	6,14	6,13	6,12	6,11	6,11	6,10	6,10
6	5,10	5,09	5,08	5,07	5,07	5,06	5,05	5,05	5,04	5,04	5,01	5,00	5,00	4,99	4,98	4,97	4,96	4,95	4,94	4,94	4,93
7	4,39	4,39	4,38	4,37	4,36	4,36	4,35	4,34	4,34	4,33	4,31	4,30	4,30	4,29	4,28	4,26	4,25	4,25	4,24	4,23	4,23
8	3,93	3,92	3,91	3,90	3,89	3,89	3,88	3,87	3,87	3,86	3,84	3,83	3,83	3,82	3,81	3,79	3,78	3,78	3,77	3,76	3,76
9	3,59	3,58	3,58	3,57	3,56	3,55	3,55	3,54	3,53	3,53	3,51	3,50	3,49	3,49	3,47	3,46	3,45	3,44	3,43	3,43	3,42
10	3,34	3,34	3,33	3,32	3,31	3,30	3,30	3,29	3,29	3,28	3,26	3,25	3,24	3,24	3,22	3,21	3,20	3,19	3,18	3,18	3,17
11	3,15	3,14	3,13	3,13	3,12	3,11	3,10	3,10	3,09	3,09	3,06	3,05	3,05	3,04	3,03	3,01	3,00	2,99	2,99	2,98	2,97
12	3,00	2,99	2,98	2,97	2,96	2,96	2,95	2,94	2,94	2,93	2,91	2,90	2,89	2,89	2,87	2,86	2,85	2,84	2,83	2,82	2,82
13	2,87	2,86	2,85	2,85	2,84	2,83	2,82	2,82	2,81	2,80	2,78	2,77	2,77	2,76	2,74	2,73	2,72	2,71	2,70	2,70	2,69
14	2,77	2,76	2,75	2,74	2,73	2,72	2,72	2,71	2,71	2,70	2,67	2,67	2,66	2,65	2,64	2,63	2,61	2,60	2,60	2,59	2,58
15	2,68	2,67	2,66	2,65	2,64	2,64	2,63	2,62	2,62	2,61	2,59	2,58	2,57	2,56	2,55	2,54	2,52	2,51	2,51	2,50	2,49
16	2,60	2,59	2,58	2,58	2,57	2,56	2,55	2,55	2,54	2,53	2,51	2,50	2,50	2,49	2,47	2,46	2,45	2,44	2,43	2,42	2,42
17	2,54	2,53	2,52	2,51	2,50	2,49	2,49	2,48	2,47	2,47	2,44	2,43	2,43	2,42	2,41	2,39	2,38	2,37	2,36	2,35	2,35
18	2,48	2,47	2,46	2,45	2,44	2,44	2,43	2,42	2,42	2,41	2,38	2,38	2,37	2,36	2,35	2,33	2,32	2,31	2,30	2,30	2,29
19	2,43	2,42	2,41	2,40	2,39	2,39	2,38	2,37	2,37	2,36	2,33	2,32	2,32	2,31	2,30	2,28	2,27	2,26	2,25	2,24	2,24
20	2,39	2,38	2,37	2,36	2,35	2,34	2,33	2,33	2,32	2,31	2,29	2,28	2,27	2,27	2,25	2,24	2,22	2,21	2,20	2,20	2,19
21	2,34	2,33	2,33	2,32	2,31	2,30	2,29	2,29	2,28	2,27	2,25	2,24	2,23	2,23	2,21	2,19	2,18	2,17	2,16	2,16	2,15
22	2,31	2,30	2,29	2,28	2,27	2,26	2,26	2,25	2,24	2,24	2,21	2,20	2,20	2,19	2,17	2,16	2,14	2,13	2,13	2,12	2,11
23	2,28	2,27	2,26	2,25	2,24	2,23	2,22	2,22	2,21	2,20	2,18	2,17	2,16	2,15	2,14	2,12	2,11	2,10	2,09	2,08	2,08
24	2,25	2,24	2,23	2,22	2,21	2,20	2,19	2,19	2,18	2,17	2,15	2,14	2,13	2,12	2,11	2,09	2,08	2,07	2,06	2,05	2,05
25	2,22	2,21	2,20	2,19	2,18	2,17	2,17	2,16	2,15	2,15	2,12	2,11	2,10	2,10	2,08	2,06	2,05	2,04	2,03	2,02	2,02
26	2,19	2,18	2,17	2,17	2,16	2,15	2,14	2,13	2,13	2,12	2,09	2,08	2,08	2,07	2,05	2,04	2,03	2,02	2,01	2,00	1,99
27	2,17	2,16	2,15	2,14	2,13	2,13	2,12	2,11	2,10	2,10	2,07	2,06	2,06	2,05	2,03	2,01	2,00	1,99	1,98	1,97	1,97
28	2,15	2,14	2,13	2,12	2,11	2,10	2,10	2,09	2,08	2,08	2,05	2,04	2,03	2,03	2,01	1,99	1,98	1,97	1,96	1,95	1,94
29	2,13	2,12	2,11	2,10	2,09	2,08	2,08	2,07	2,06	2,06	2,03	2,02	2,01	2,01	1,99	1,97	1,96	1,95	1,94	1,93	1,92
30	2,11	2,10	2,09	2,08	2,07	2,07	2,06	2,05	2,04	2,04	2,01	2,00	1,99	1,99	1,97	1,95	1,94	1,93	1,92	1,91	1,90
31	2,10	2,08	2,07	2,07	2,06	2,05	2,04	2,03	2,03	2,02	1,99	1,98	1,98	1,97	1,95	1,94	1,92	1,91	1,90	1,89	1,89
32	2,08	2,07	2,06	2,05	2,04	2,03	2,02	2,02	2,01	2,00	1,98	1,97	1,96	1,95	1,93	1,92	1,91	1,89	1,88	1,88	1,87
33	2,06	2,05	2,04	2,03	2,03	2,02	2,01	2,00	2,00	1,99	1,96	1,95	1,95	1,94	1,92	1,90	1,89	1,88	1,87	1,86	1,85
34	2,05	2,04	2,03	2,02	2,01	2,00	2,00	1,99	1,98	1,97	1,95	1,94	1,93	1,92	1,90	1,89	1,88	1,86	1,85	1,85	1,84
35	2,04	2,03	2,02	2,01	2,00	1,99	1,98	1,97	1,97	1,96	1,93	1,92	1,92	1,91	1,89	1,87	1,86	1,85	1,84	1,83	1,82
36	2,03	2,01	2,00	2,00	1,99	1,98	1,97	1,96	1,96	1,95	1,92	1,91	1,90	1,90	1,88	1,86	1,85	1,84	1,83	1,82	1,81
37	2,01	2,00	1,99	1,98	1,97	1,97	1,96	1,95	1,94	1,94	1,91	1,90	1,89	1,88	1,87	1,85	1,84	1,82	1,81	1,81	1,80
38	2,00	1,99	1,98	1,97	1,96	1,95	1,95	1,94	1,93	1,93	1,90	1,89	1,88	1,87	1,85	1,84	1,82	1,81	1,80	1,79	1,79
39	1,99	1,98	1,97	1,96	1,95	1,94	1,94	1,93	1,92	1,91	1,89	1,88	1,87	1,86	1,84	1,83	1,81	1,80	1,79	1,78	1,78
40	1,98	1,97	1,96	1,95	1,94	1,93	1,93	1,92	1,91	1,90	1,88	1,87	1,86	1,85	1,83	1,82	1,80	1,79	1,78	1,77	1,76
41	1,97	1,96	1,95	1,94	1,93	1,92	1,92	1,91	1,90	1,90	1,87	1,86	1,85	1,84	1,82	1,81	1,79	1,78	1,77	1,76	1,75
42	1,96	1,95	1,94	1,93	1,92	1,92	1,91	1,90	1,89	1,89	1,86	1,85	1,84	1,83	1,81	1,80	1,78	1,77	1,76	1,75	1,74
43	1,96	1,95	1,93	1,93	1,92	1,91	1,90	1,89	1,88	1,88	1,85	1,84	1,83	1,82	1,80	1,79	1,77	1,76	1,75	1,74	1,74
44	1,95	1,94	1,93	1,92	1,91	1,90	1,89	1,88	1,88	1,87	1,84	1,83	1,82	1,82	1,80	1,78	1,77	1,75	1,74	1,73	1,73
45	1,94	1,93	1,92	1,91	1,90	1,89	1,88	1,88	1,87	1,86	1,83	1,82	1,82	1,81	1,79	1,77	1,76	1,75	1,74	1,73	1,72
46	1,93	1,92	1,91	1,90	1,89	1,88	1,88	1,87	1,86	1,85	1,82	1,81	1,81	1,80	1,78	1,76	1,75	1,74	1,73	1,72	1,71
47	1,93	1,91	1,90	1,89	1,89	1,88	1,87	1,86	1,85	1,85	1,82	1,81	1,80	1,79	1,77	1,76	1,74	1,73	1,72	1,71	1,70
48	1,92	1,91	1,90	1,89	1,88	1,87	1,86	1,85	1,85	1,84	1,81	1,80	1,79	1,79	1,77	1,75	1,73	1,72	1,71	1,70	1,69
49	1,91	1,90	1,89	1,88	1,87	1,86	1,86	1,85	1,84	1,83	1,80	1,79	1,79	1,78	1,76	1,74	1,73	1,72	1,71	1,70	1,69
50	1,91	1,90	1,89	1,88	1,87	1,86	1,85	1,84	1,83	1,83	1,80	1,79	1,78	1,77	1,75	1,74	1,72	1,71	1,70	1,69	1,68
55	1,88	1,87	1,86	1,85	1,84	1,83	1,82	1,81	1,81	1,80	1,77	1,76	1,75	1,74	1,72	1,71	1,69	1,68	1,67	1,66	1,65
60	1,86	1,85	1,83	1,82	1,82	1,81	1,80	1,79	1,78	1,78	1,74	1,73	1,73	1,72	1,70	1,68	1,67	1,65	1,64	1,63	1,63
65	1,84	1,83	1,82	1,81	1,80	1,79	1,78	1,77	1,76	1,76	1,72	1,71	1,71	1,70	1,68	1,66	1,65	1,63	1,62	1,61	1,60
70	1,82	1,81	1,80	1,79	1,78	1,77	1,76	1,75	1,75	1,74	1,71	1,70	1,69	1,68	1,66	1,64	1,63	1,62	1,60	1,59	1,59
75	1,81	1,80	1,78	1,77	1,76	1,76	1,75	1,74	1,73	1,72	1,69	1,68	1,68	1,67	1,65	1,63	1,61	1,60	1,59	1,58	1,57
80	1,79	1,78	1,77	1,76	1,75	1,74	1,73	1,73	1,72	1,71	1,68	1,67	1,66	1,65	1,63	1,61	1,60	1,59	1,57	1,56	1,55

FIGURE 18 – Table 5d : Loi de Fisher-Snedecor

VALEURS DE d_α TELLES QUE $\mathbb{P}[\max(|F^{\text{th}} - F^{\text{exp}}|) \geqslant d_\alpha] = \alpha$

n \ α	0,2	0,1	0,05	0,025	0,02	0,01	0,005
1	0,9000	0,9500	0,9750	0,9875	0,9900	0,9950	0,9975
2	0,6838	0,7764	0,8419	0,8882	0,9000	0,9293	0,9500
3	0,5648	0,6360	0,7076	0,7679	0,7846	0,8290	0,8643
4	0,4927	0,5652	0,6239	0,6739	0,6889	0,7342	0,7764
5	0,4470	0,5094	0,5633	0,6126	0,6272	0,6685	0,7054
6	0,4104	0,4680	0,5193	0,5640	0,5774	0,6166	0,6529
7	0,3815	0,4361	0,4834	0,5256	0,5384	0,5758	0,6098
8	0,3583	0,4096	0,4543	0,4945	0,5065	0,5418	0,5743
9	0,3391	0,3875	0,4300	0,4681	0,4796	0,5133	0,5444
10	0,3226	0,3687	0,4092	0,4456	0,4566	0,4889	0,5187
11	0,3083	0,3524	0,3912	0,4261	0,4367	0,4677	0,4964
12	0,2958	0,3382	0,3754	0,4090	0,4192	0,4490	0,4767
13	0,2847	0,3255	0,3614	0,3938	0,4036	0,4325	0,4592
14	0,2748	0,3142	0,3489	0,3802	0,3897	0,4176	0,4435
15	0,2659	0,3040	0,3376	0,3679	0,3771	0,4042	0,4293
16	0,2578	0,2947	0,3273	0,3568	0,3657	0,3920	0,4164
17	0,2504	0,2863	0,3180	0,3466	0,3553	0,3809	0,4046
18	0,2436	0,2785	0,3094	0,3372	0,3457	0,3706	0,3938
19	0,2373	0,2714	0,3014	0,3286	0,3369	0,3612	0,3838
20	0,2316	0,2647	0,2941	0,3206	0,3287	0,3524	0,3745
21	0,2262	0,2586	0,2872	0,3132	0,3210	0,3443	0,3659
22	0,2212	0,2528	0,2809	0,3062	0,3139	0,3367	0,3578
23	0,2165	0,2475	0,2749	0,2997	0,3073	0,3295	0,3503
24	0,2120	0,2424	0,2693	0,2936	0,3010	0,3229	0,3432
25	0,2079	0,2377	0,2640	0,2879	0,2952	0,3166	0,3365
26	0,2040	0,2332	0,2591	0,2825	0,2896	0,3106	0,3302
27	0,2003	0,2290	0,2544	0,2774	0,2844	0,3050	0,3243
28	0,1968	0,2250	0,2499	0,2725	0,2794	0,2997	0,3186
29	0,1935	0,2212	0,2457	0,2679	0,2747	0,2947	0,3133
30	0,1903	0,2176	0,2417	0,2636	0,2702	0,2899	0,3082
31	0,1873	0,2141	0,2379	0,2594	0,2660	0,2853	0,3033
32	0,1844	0,2108	0,2342	0,2554	0,2619	0,2809	0,2987
33	0,1817	0,2077	0,2308	0,2517	0,2580	0,2768	0,2943
34	0,1791	0,2047	0,2274	0,2480	0,2543	0,2728	0,2901
35	0,1766	0,2018	0,2242	0,2446	0,2507	0,2690	0,2860
36	0,1742	0,1991	0,2212	0,2412	0,2473	0,2653	0,2821
37	0,1719	0,1965	0,2183	0,2380	0,2440	0,2618	0,2784
38	0,1697	0,1939	0,2154	0,2350	0,2409	0,2584	0,2748
39	0,1675	0,1915	0,2127	0,2320	0,2379	0,2552	0,2713
40	0,1655	0,1891	0,2101	0,2291	0,2349	0,2521	0,2680
41	0,1635	0,1869	0,2076	0,2264	0,2321	0,2490	0,2648
42	0,1616	0,1847	0,2052	0,2238	0,2294	0,2461	0,2617
43	0,1597	0,1826	0,2028	0,2212	0,2268	0,2433	0,2587
44	0,1580	0,1805	0,2006	0,2187	0,2243	0,2406	0,2559
45	0,1562	0,1786	0,1984	0,2163	0,2218	0,2380	0,2531
46	0,1546	0,1767	0,1963	0,2140	0,2194	0,2354	0,2504
47	0,1530	0,1748	0,1942	0,2118	0,2171	0,2330	0,2478
48	0,1514	0,1730	0,1922	0,2096	0,2149	0,2306	0,2452
49	0,1499	0,1713	0,1903	0,2075	0,2128	0,2283	0,2428
50	0,1484	0,1696	0,1884	0,2055	0,2107	0,2260	0,2404
51	0,1470	0,1680	0,1866	0,2035	0,2086	0,2239	0,2381
52	0,1456	0,1664	0,1848	0,2016	0,2067	0,2217	0,2358
53	0,1442	0,1648	0,1831	0,1997	0,2048	0,2197	0,2336
54	0,1429	0,1633	0,1814	0,1979	0,2029	0,2177	0,2315
55	0,1416	0,1619	0,1798	0,1961	0,2011	0,2157	0,2294
56	0,1404	0,1604	0,1782	0,1944	0,1993	0,2138	0,2274
57	0,1392	0,1591	0,1767	0,1927	0,1976	0,2120	0,2255
58	0,1380	0,1577	0,1752	0,1911	0,1959	0,2102	0,2236
59	0,1369	0,1564	0,1737	0,1895	0,1943	0,2084	0,2217
60	0,1357	0,1551	0,1723	0,1879	0,1927	0,2067	0,2199
61	0,1346	0,1539	0,1709	0,1864	0,1911	0,2051	0,2181
62	0,1336	0,1526	0,1696	0,1849	0,1896	0,2034	0,2164
63	0,1325	0,1514	0,1682	0,1835	0,1881	0,2018	0,2147
64	0,1315	0,1503	0,1669	0,1821	0,1867	0,2003	0,2130
65	0,1305	0,1491	0,1657	0,1807	0,1853	0,1988	0,2114
66	0,1295	0,1480	0,1644	0,1793	0,1839	0,1973	0,2098
67	0,1286	0,1469	0,1632	0,1780	0,1825	0,1958	0,2083
68	0,1277	0,1459	0,1620	0,1767	0,1812	0,1944	0,2068
69	0,1267	0,1448	0,1609	0,1755	0,1799	0,1930	0,2053
70	0,1259	0,1438	0,1597	0,1742	0,1786	0,1917	0,2039

Pour n>70 : $d_\alpha \simeq \frac{C_\alpha}{\sqrt{n}}$
où C_α est donné par :

α	C_α
0,2	1,073
0,1	1,224
0,05	1,358
0,03	1,480
0,02	1,517
0,01	1,628
0,005	1,731

FIGURE 19 – Table 6 : Loi de Kolmogorov-Smirnov

TABLE DES VALEURS MAXIMALES w_α TELLES QUE $\mathbb{P}[W \leqslant w_\alpha] < \alpha$

Où n' est le nombre de différences non nulles.

α \ n'	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
0,05	0	2	3	5	8	10	13	17	21	25	29	34	40	46	52	58	65	73	81	89
0,02	-	0	1	3	5	7	9	12	15	19	23	27	32	37	43	49	55	62	69	76
0,01	-	-	0	1	3	5	7	9	12	15	19	23	27	32	37	42	48	54	61	68

Pour les grands échantillons, $W \simeq \mathcal{N}\left(\frac{n'(n'+1)}{4} ; \sqrt{\frac{n'(n'+1)(2n'+1)}{24}}\right)$

FIGURE 20 – Table 7 : Loi de Wilcoxon

TABLE DES VALEURS MAXIMALES u_α TELLES QUE $\mathbb{P}[U \leqslant u_\alpha] < \alpha$

Où n_1 et n_2 sont les tailles des échantillons.

$\lvert n_1 - n_2 \rvert$	α \ $\min(n_1;n_2)$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0,05	-	-	-	0	2	5	8	13	17	23	30	37	45	55	64	75	87	99	113	127
	0,01	-	-	-	-	0	2	4	7	11	16	21	27	34	42	51	60	70	81	93	105
1	0,05	-	-	-	1	3	6	10	15	20	26	33	41	50	59	70	81	93	106	119	134
	0,01	-	-	-	-	1	3	6	9	13	18	24	31	38	46	55	65	75	87	99	112
2	0,05	-	-	0	2	5	8	12	17	23	29	37	45	54	64	75	86	99	112	126	141
	0,01	-	-	-	0	1	4	7	11	16	21	27	34	42	50	60	70	81	92	105	118
3	0,05	-	-	1	3	6	10	14	19	26	33	40	49	59	69	80	92	105	119	133	149
	0,01	-	-	-	0	2	5	9	13	18	24	30	37	45	54	64	74	86	98	111	125
4	0,05	-	-	1	4	7	11	16	22	28	36	44	53	63	74	85	98	111	125	140	156
	0,01	-	-	-	1	3	6	10	15	20	26	33	41	49	58	69	79	91	104	117	131
5	0,05	-	-	2	4	8	13	18	24	31	39	47	57	67	78	90	103	117	132	147	163
	0,01	-	-	-	1	4	7	12	17	22	29	36	44	53	63	73	84	96	109	123	138
6	0,05	-	0	2	5	9	14	20	26	34	42	51	61	72	83	96	109	123	138	154	171
	0,01	-	-	0	2	5	9	13	18	24	31	39	47	57	67	78	89	102	115	129	144
7	0,05	-	0	3	6	11	16	22	29	37	45	55	65	76	88	101	115	129	145	161	178
	0,01	-	-	0	2	6	10	15	20	27	34	42	51	60	71	82	94	107	121	135	151
8	0,05	-	0	3	7	12	17	24	31	39	48	58	69	80	93	106	120	135	151	168	186
	0,01	-	-	0	3	7	11	16	22	29	37	45	54	64	75	87	99	112	127	142	157
9	0,05	-	0	4	8	13	19	26	34	42	52	62	73	85	98	111	126	141	158	175	193
	0,01	-	-	1	3	7	12	18	24	31	39	48	58	68	79	91	104	118	132	148	164
10	0,05	-	1	4	9	14	21	28	36	45	55	65	77	89	102	117	132	147	164	182	200
	0,01	-	-	1	4	8	13	19	26	33	42	51	61	72	83	96	109	123	138	154	170
11	0,05	-	1	5	10	15	22	30	38	48	58	69	81	94	107	122	137	154	171	189	208
	0,01	-	-	1	5	9	15	21	28	36	44	54	64	75	87	100	114	128	144	160	177
12	0,05	-	1	5	11	17	24	32	41	50	61	73	85	98	112	127	143	160	177	196	215
	0,01	-	-	2	5	10	16	22	30	38	47	57	68	79	92	105	119	134	150	166	184
13	0,05	-	1	6	11	18	25	34	43	53	64	76	89	102	117	132	149	166	184	203	222
	0,01	-	-	2	6	11	17	24	32	40	50	60	71	83	96	109	124	139	155	172	190
14	0,05	-	1	6	12	19	27	36	45	56	67	80	93	107	122	138	154	172	190	210	230
	0,01	-	-	2	6	12	18	25	34	43	52	63	74	87	100	114	129	145	161	179	197
15	0,05	-	2	7	13	20	29	38	48	59	71	83	97	111	127	143	160	178	197	217	237
	0,01	-	-	2	7	13	19	27	35	45	55	66	78	91	104	119	134	150	167	185	203
16	0,05	-	2	7	14	22	30	40	50	62	74	87	101	116	131	148	166	184	203	224	245
	0,01	-	-	3	8	14	21	29	37	47	58	69	81	94	108	123	139	155	173	191	210
17	0,05	-	2	8	15	23	32	42	53	64	77	90	105	120	136	153	171	190	210	231	252
	0,01	-	0	3	8	14	22	30	39	49	60	72	85	98	113	128	144	161	179	197	217
18	0,05	-	2	8	16	24	33	44	55	67	80	94	109	125	141	159	177	196	216	238	259
	0,01	-	0	3	9	15	23	32	41	52	63	75	88	102	117	132	149	166	184	203	223
19	0,05	-	3	9	17	25	35	46	57	70	83	98	113	129	146	164	183	202	223	245	267
	0,01	-	0	4	9	16	24	33	43	54	66	78	92	106	121	137	154	172	190	210	230
20	0,05	-	3	9	17	27	37	48	60	73	87	101	117	133	151	169	188	209	230	252	274
	0,01	-	0	4	10	17	25	35	45	56	68	81	95	110	125	142	159	177	196	216	237

Pour les grand échantillons, $U \simeq \mathcal{N}\left(\frac{n_1 n_2}{2} \; ; \; \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}\right)$

FIGURE 21 – Table 8 : Loi de Mann-Whitney

Valeurs critiques de la loi $Y = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1)$ à 5% et à 1%

où n_i est taille du ième groupe. n est l'effectif total
k = nombre de groupes

k = 3					k=4						k=5						
Tailles des échantillons			5%	1%	Tailles des échantillons				5%	1%	Tailles des échantillons					5%	1%
4	2	2	5,33	-	3	3	1	1	6,33	-	3	2	2	2	1	7,309	8,127
4	3	1	5,208	-	3	3	2	1	6,244	7,2	3	2	2	2	2	7,682	8,682
4	3	2	5,444	6,444	3	3	2	2	6,527	7,636	3	3	1	1	1	7,111	-
4	3	3	5,791	6,745	3	3	3	1	6,6	7,4	3	3	2	1	1	7,2	8,073
4	4	1	4,967	6,667	3	3	3	2	6,727	8,015	3	3	2	2	1	7,591	8,576
4	4	2	5,455	7,036	4	3	3	3	7	8,538	3	3	2	2	2	7,91	9,115
4	4	3	5,598	7,144	4	1	1	1	-	-	3	3	3	1	1	7,576	8,424
4	4	4	5,692	7,654	4	2	1	1	5,833	-	3	3	3	2	1	7,769	9,051
5	2	1	5	-	4	2	2	1	6,133	7	3	3	3	2	2	8,044	9,505
5	2	2	5,16	6,533	4	2	2	2	6,545	7,391	3	3	3	3	1	8	9,451
5	3	1	4,96	-	4	3	1	1	6,178	7,067	3	3	3	3	2	8,2	9,876
5	3	2	5,251	6,909	4	3	2	1	6,309	7,455	3	3	3	3	3	8,333	10,2
5	3	3	5,648	7,079	4	3	2	2	6,621	7,871							
5	4	1	4,985	6,955	4	3	3	1	6,545	7,758							
5	4	2	5,273	7,205	4	3	3	2	6,795	8,333							
5	4	3	5,656	7,445	4	3	3	3	6,984	8,659							
5	4	4	5,657	7,76	4	4	1	1	5,945	7,909							
5	5	1	5,127	7,309	4	4	2	1	6,386	7,909							
5	5	2	5,338	7,338	4	4	2	2	6,731	8,346							
5	5	3	5,705	7,578	4	4	3	1	6,635	8,231							
5	5	4	5,666	7,823	4	4	3	2	6,874	8,621							
5	5	5	5,78	8	4	4	3	3	7,038	8,876							
6	1	1	-	-	4	4	4	1	6,725	8,588							
6	2	1	4,822	-	4	4	4	2	6,957	8,871							
6	2	2	5,345	6,655	4	4	4	3	7,142	9,075							
6	3	1	4,855	6,873	4	4	4	4	7,235	9,287							
6	3	2	5,348	6,97													
6	3	3	5,615	7,41													
6	4	1	4,947	7,106													
6	4	2	5,34	7,34													
6	4	3	5,61	7,5													
6	4	4	5,681	7,795													
6	5	1	4,99	7,182													
6	5	2	5,338	7,376													
6	5	3	5,602	7,59													
6	5	4	5,661	7,936													
6	5	5	5,729	8,028													
6	6	1	4,945	7,121													
6	6	2	5,41	7,467													
6	6	3	5,625	7,725													
6	6	4	5,724	8													
6	6	5	5,765	8,124													
6	6	6	5,801	8,222													
7	7	7	5,819	8,378													
8	8	8	5,805	8,465													

Figure 22 – Table 9 : Table de Kruskal-Wallis

Bibliographie

[1] A Ancelle. *Statistique Épidémiologie.* Maloine, 2002.

[2] D Bouget and A Viéno. *Traitement de l'information : Statistiques et Probabilités.* Vuibert, 1995.

[3] M Carricano and F Poujol. *Analyse de données avec SPSS.* Pearson Education, 2008.

[4] C.-P Dancey and J Reidy. *Statistiques sans maths pour psychlogues.* de boeck, 2007.

[5] M.-b Dosse. *Statistique bivariée avec R.* Pur, 2011.

[6] J.-J Droesbeke, C Dehon, and C Vermandele. *Eléments de statistique.* SMA, 2008.

[7] Y Evrad, B Pras, and E Roux. *Market,* 4^e *édition.* Dunod, 2009.

[8] S Ganassali. *enquêtes et analyse de données avec Sphinx.* Pearson Education, 2014.

[9] D.-C. Howell. *Méthodes statistiques en sciences humaines.* de boeck, 2008.

[10] W Janssens, K Wijnen, P De Pelsmacker, and P.-V Kenhove. *Marketing research with SPSS.* Prentice Hall, 2008.

[11] A Joliber and P Jourdan. *Marketing research.* Dunod, 2011.

[12] N Malhotra. *Études marketing,* 6^e *édition.* Pearson Education, 2011.

[13] Alain Méot. *Introduction aux statistiques inférentielles.* de booeck, 2003.

[14] P.-C Pupion. *Statistique pour la gestion,* 3^e *édition.* Dunod, 2012.

[15] B Py. *La statistique sans formule mathématique,* 2^e *édition.* Person Education, 2010.

[16] J Stafford and P Bodson. *L'analyse multivariée avec SPSS.* Presses de l'Université du Québec, 2006.

[17] B Tribout. *Statistique pour économistes et gestionnaires.* Pearson Education, 2007.

Table des matières