République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA
FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE
DÉPARTEMENT DE MATHÉMATIQUE

Polycopié du TP

$\underbrace{BIOST\widehat{ATIST}IQUES}$

Statistiques Appliquées à l'Expérimentation En Sciences Biologique sous SPSS (Troisième Année Licence)

Préparé par :
Dr. CHERFAOUI Mouloud

Université de Biskra, 2020/2021

Table des matières

Chapitre 1

Estimation sous SPSS

Introduction

L'objectif du présent chapitre est d'illustrer, à travers d'un exemple numérique, les principales étapes à suivre dans le logiciel SPSS pour obtenir une estimation par intervalle de confiance d'une moyenne à un seuil de confiance donné. Avant la présentation de ces étapes, un bref rappel sur la notion d'estimation ponctuelle et par intervalle de confiance d'un paramètre sera présentée.

1.1 Notion d'estimation d'un paramètre

L'estimation consiste à évaluer certaines caractéristiques d'une variable aléatoire à base des observations réalisées sur un échantillon. C'est-à-dire, elle consiste à évaluer un paramètre inconnu θ (dont θ peut être une moyenne, une variance, une fréquence,...) de la population à partir d'un échantillon représentatif tiré de cette population.

La valeur estimée du paramètre θ est généralement notée $\hat{\theta}$, alors

- θ : est la vraie valeur (inconnue) du paramètre.
- $\hat{\theta}$: est la valeur estimée du paramètre θ.

Un survol de la littérature nous permet de distinguer deux principales classes de méthodes d'estimation, à savoir : l'estimation ponctuelle et l'estimation par intervalle de confiance. Les principes de ces deux types d'estimation sont résumés dans les deux prochaines sections.

1.1.1 Estimation ponctuelle

Lorsqu'un paramètre θ est estimé par un seul nombre déduit des résultats de l'échantillon, ce nombre est appelé une estimation ponctuelle de ce paramètre. Entre autres on peut citer l'exemple de :

- L'estimateur de la moyenne μ qui est donné par la moyenne de l'échantillon

$$\hat{\mu} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- L'estimateur de la variance σ^2 qui est donné par la variance corrigée de l'échantillon

$$\hat{\sigma}_c^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

- L'estimateur de la variance σ^2, lorsque la moyenne μ est connue, qui est donné par

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

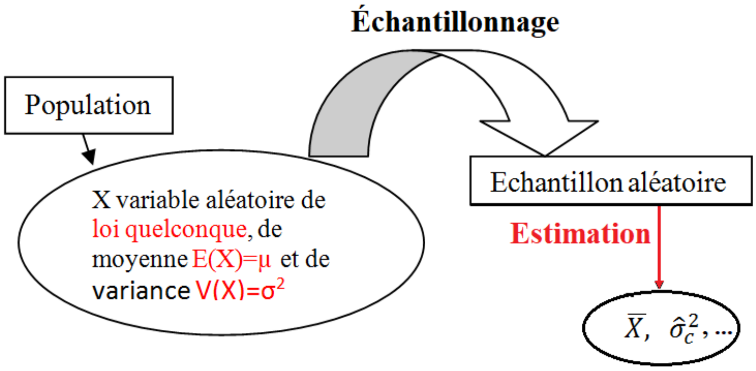


FIGURE 1.1: *Illustration graphique de la notion d'estimation.*

1.1.2 Estimation par intervalle de confiance

Les estimations ponctuelles, bien qu'utiles en pratique, ne fournissent aucune information concernant la précision des estimations. Ceci le fait qu'elles ne tiennent pas compte de l'erreur possible dans l'estimation due aux fluctuations (erreurs) d'échantillonnage. La théorie d'estimation par des intervalles de confiance (IC) consiste à construire, autour d'une estimation ponctuelle $\hat{\theta}$, un intervalle qui aura une forte probabilité $(1-\alpha)$ de contenir la vraie valeur du paramètre θ recherché.

Typiquement, on cherche deux nombres réels a et b tel que :

$$P(a \leq \theta \leq b) = 1 - \alpha,$$

où

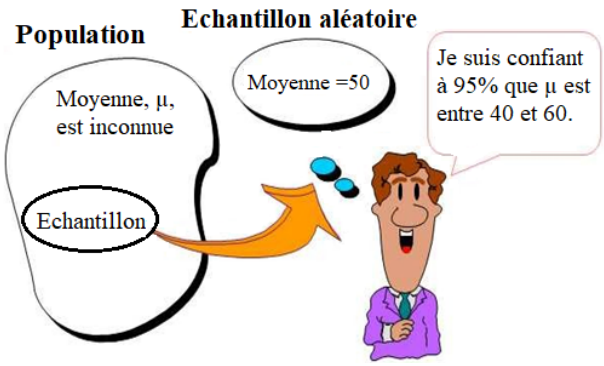
- L'intervalle $[a, b]$ est appelé intervalle de confiance pour θ et il est noté par

$$IC_{1-\alpha}(\theta) = [a, b].$$

- La probabilité $(1-\alpha)$ est appelé le niveau de confiance et α est appelé le risque de se tromper que l'on est prêt à prendre c'est-à-dire la probabilité que la vraie valeur du paramètre θ n'appartienne pas à l'intervalle $[a, b]$ $(P(\theta \notin [a, b]) = \alpha)$.

Dans la pratique, généralement, on prend $\alpha \in \{10\%, 5\%, 2\%, 1\%, 0.1\%\}$.

Par exemple si on prend $\alpha = 5\%$, ce qui correspond à un IC à 95%, se traduit par :" il y a 95% de chance que la vraie valeur inconnue de θ soit comprise entre a et b".



1.2 Intevalle de confiance d'une moyenne sous SPSS

Dans les problèmes statistiques on peut s'intéresser à l'estimation de divers caractéristiques d'une population mais dans ce document nous allons se limiter uniquement à la détermination de la moyenne.

L'objectif de la présente section est de lister les étapes à suivre dans le logiciel SPSS pour construire un intervalle de confiance d'une moyenne. Pour ce faire, considérant l'exemple suivant :

Exemple 1 *La mesure de la taille (en cm) de 10 enfants d'une ville donnée à fourni ce qui suit :*

70.5	*85.2*	*93.8*	*99.1*	*101*	*105.8*	*110.3*	*121.2*	*138.6*	*166.1*

Questions :

1. *Déterminer une estimation ponctuelle de la moyenne et de l'écart-type des tailles de ces enfants.*
2. *Déterminer une estimation par intervalle de confiance de la taille moyenne de ces enfants pour un risque $\alpha = 5\%$.*
3. *Déterminer une estimation par intervalle de confiance de la taille moyenne de ces enfants pour un risque $\alpha = 1\%$.*

Pour répondre aux questions de cet exemple sous SPSS, il suffit de suivre les étapes suivantes :

Etape 1. Saisie des données :

Entrez les données dans SPSS, dont nous avons une seule variable quantitative (**Tailles des enfants**) à définir dans SPSS : cliquez sur "**Affichages des variables"** puis saisissez tous ses caractéristiques : **Nom**, **Type**, **Largeur**,...), ensuite cliquez sur "**Affichage des données"** pour entrer les valeurs des variables où chaque colonne représente une variable (Voir Figure 1.2).

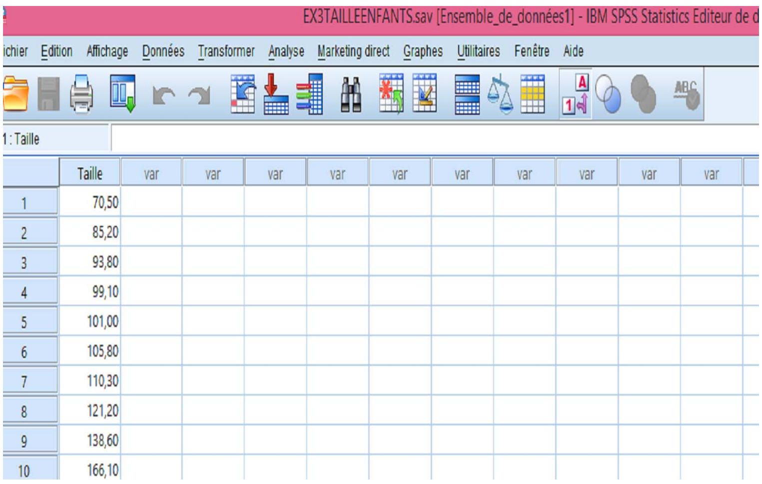


FIGURE 1.2: *Saisie des données sous SPSS.*

Remarque 1.1

1. *Il faut sauvegardez votre fichier.*
2. *Lors de l'introduction des nombres décimales dans SPSS si la configuration du PC est par défaut sont présenté par une virgule et non par un point. Exemple : on doit écrire "70, 50" et non "70.50".*

Etape 2. Réalisation de l'estimation sous SPSS (partie 1). Sélectionnez sur la barre de menu

Analyse → Comparer les moyennes → Test T pour échantillon unique (Voir Figure 1.3).

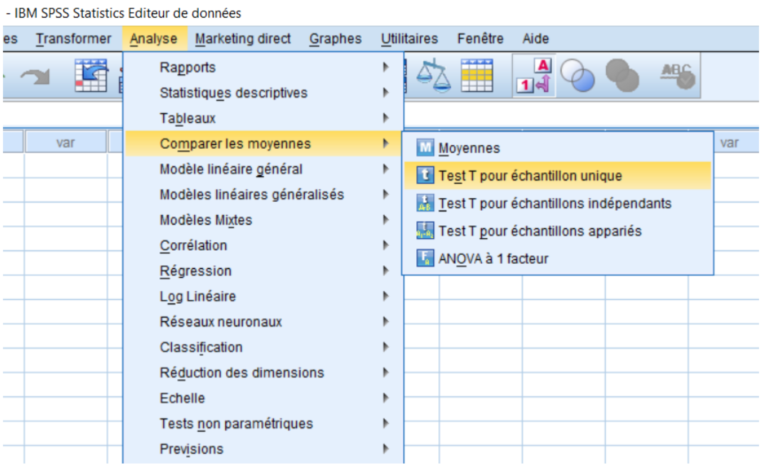


FIGURE 1.3: *Sélection de la méthode d'estimation par IC sous SPSS.*

Etape 3. En cliquant sur " **Test T pour échantillon unique**" (voir Etape 2) une fenêtre sera affichée dont la forme est présentée dans la figure 1.4 à gauche. Alors il faut suivre ce qui suit :

1. Mettez la variable **Taille enfant** dans la case "**Variables à tester"**.
2. Vérifiez dans la case "**Valeur de test"** qu'elle est égale à 0.
3. Cliquez sur "**option"** une fenêtre va s'ouvrir (voir Figure 1.4 à droite).
4. Choisissez le niveau de confiance, par exemple dans la question 2, on a le risque $\alpha = 5\%$, donc le niveau de confiance égale à $1 - \alpha = 1 - 0.05 = 0.95$ (95%).
5. Cliquez sur "**poursuivre"**.
6. Cliquez sur "**OK"** pour afficher les résultats.

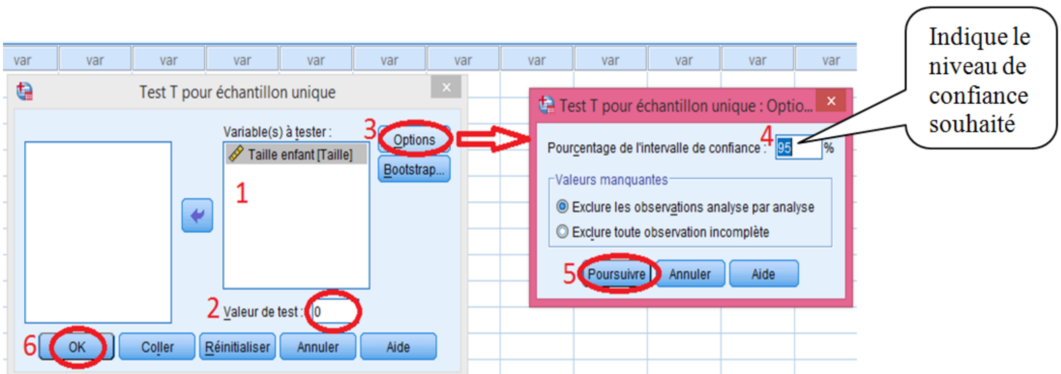
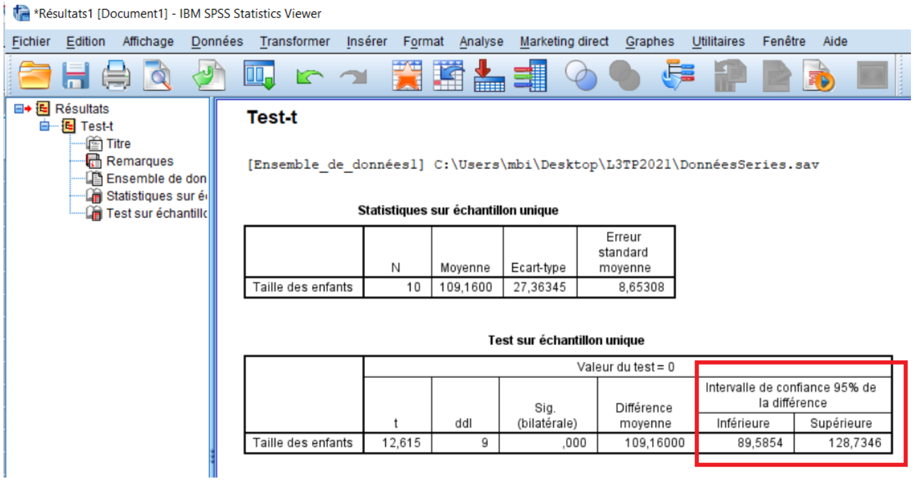


FIGURE 1.4: *Sélection de la variable à traiter et le seuil de confiance.*

Etape 4. Résultats : L'application des trois étapes précédentes sans erreurs sur les données de l'exemple permet d'obtenir les résultats (2 tableaux) qui sont présentés dans la Figure 1.5. Ces résultats contiennent des réponses sur la questions 1 (tableau 1) et la question 2 (tableau 2).

Etape 5. Interprétation des résultats :

Le premier tableau fourni :



	N	Moyenne	Ecart-type	Erreur standard moyenne
Taille des enfants	10	109,1600	27,36345	8,65308

	Valeur du test = 0					
	t	ddl	Sig. (bilatérale)	Différence moyenne	Intervalle de confiance 95% de la différence	
					Inférieure	Supérieure
Taille des enfants	12,615	9	,000	109,16000	89,5854	128,7346

FIGURE 1.5: *Fenêtre d'affichage des résultats de l'estimation sous SPSS.*

- La taille de l'échantillon : $N = 10$.
- Une estimation ponctuelle de la taille moyenne des enfants : $\hat{\mu} = \overline{X} = 109.16\ cm$.
- Une estimation ponctuelle de l'écart-type : $\hat{\sigma}_c = \sqrt{\hat{\sigma}_c^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2} = 27.36345$ cm.
- Une estimation ponctuelle pour l'erreur standard moyenne : $\frac{\hat{\sigma}_c}{\sqrt{n}} = 8.65308\ cm$.

Pour le deuxième tableau, l'information cruciale dans ce tableau, pour l'objectif de notre exemple, est la colonne intitulée "**Intervalle de confiance 95% de la différence"** qui représente l'intervalle de confiance de niveau 95% de la taille moyenne des enfants (les bornes inférieure et supérieure de cet intervalle), d'où :

$$IC_{95\%}(\mu) = [89.5854, 128.7346].$$

Remarque 1.2 *Pour répondre à la question 3 de l'exemple, c'est-à-dire pour obtenir un intervalle de confiance de la taille moyenne des enfants au risque $\alpha = 1\%$, il suffit de suivre les mêmes étapes précédentes avec le changement du seuil de confiance au niveau de l'étape 3, dans ce cas, il est égale à 99% (voir figure 1.4). L'intervalle de confiance de la taille moyenne des enfants est donné cette fois-ci comme suite :*

$$IC_{99\%}(\mu) = [81.0389, 137.2811].$$

Test sur échantillon unique

	Valeur du test = 0					
	t	ddl	Sig. (bilatérale)	Différence moyenne	Intervalle de confiance 99% de la différence	
					Inférieure	Supérieure
Taille enfant	12,615	9	,000	109,16000	81,0389	137,2811

FIGURE 1.6: *Intervalle de confiance de la taille moyenne des enfants au risque $\alpha = 1\%$.*

Chapitre 2

Tests d'hypothèses

Introduction

Les tests de conformité statistiques constituent une approche décisionnelle de la statistique inférentielle. Un tel test a pour objet de décider sur la base d'un échantillon si une caractéristique de la loi mère (ou de la population) répond ou non à une certaine spécification que l'on appelle hypothèse, par exemple : on désir savoir si la moyenne d'une loi est significativement supérieure à 10.

Dans ce chapitre, dans un premier lieu nous allons présenter les notions de base nécessaires pour la compréhension d'un test statistique et ces différents types. Par la suite nous nous intéressons à la mise en œuvre du test de conformité et d'homogénéité de moyennes de Student sous SPSS.

2.1 Principe d'un test statistique

Les tests d'hypothèses sont des outils statistiques d'aide à la décision. Ils vont permettre de comparer un ou plusieurs échantillons, et de valider ou d'invalider une hypothèse donnée avec un certain risque de se tromper. En effet, un test statistique est une mise à l'épreuve d'une hypothèse concernant une population sur la base de données fournies à partir d'un échantillon (ou plusieurs) représentatif de la population, qui permet de prendre la décision de rejeter ou de ne pas rejeter les hypothèses.

2.2 Eléments d'un test

On a vu que le principe d'un test statistique est une règle de décision qui permet, sur la base des données observées et avec des risques d'erreur déterminés, d'accepter ou de refuser une hypothèse statistique. Ça sous-entend, qu'un test statistique est constitué principalement de deux éléments à savoir : hypothèses et risque.

2.2.1 Hypothèses d'un test

Une hypothèse est l'information qu'on veut confirmer ou infirmer concernant les caractéristiques (valeurs des paramètres, forme de la distribution des observations,...) d'une population.

Le principe est de comparer la probabilité d'une hypothèse versus le contraire de cette même hypothèse. Ainsi, on distingue deux types d'hypothèses, dont une et une seule est vraie.

L'hypothèse nulle : Cette hypothèse est notée H_0, elle est l'hypothèse principale du test et qu'on considère vraie a priori.

L'hypothèse alternative : Cette hypothèse est notée H_1. C'est l'hypothèse qu'on choisisse en cas de rejet de l'hypothèse H_0. Quoi qu'on a une liberté pour le choix de cette hypothèse ça reste qu'elle doit être choisis d'une manière qu'elle soit compatible avec le problème étudie et elle soit différente de l'hypothèse H_0.

2.2.2 Risque d'un test

Dans un test statistique y a toujours des erreurs de décision c'est-à-dire on ne pourra jamais conclure avec certitude sur le rejet ou le non rejet d'une hypothèse. A cet effet, pour effectuer un test statistique, il faudra choisir un certain risque d'erreur qui décrit la probabilité de se tromper en prenant la décision retenue. Dans la pratique, on choisit un risque α qui qui reflète la probabilité de rejeter l'hypothèse H_0 alors qu'elle est vraie :

$$\alpha = P(\text{rejeter } H_0/H_0 \text{ est vraie}).$$

Remarque 2.1

- *On appelle α le seuil de risque ou niveau de signification (ou encore seuil de signification).*
- *Les valeurs usuelles de α sont* 10%, 5%, 2%, 1%, 0.1%.

2.3 Tests bilatéral et unilatéral

Avant d'appliquer tout test statistique, il s'agit de bien définir le problème posé. En effet, selon les hypothèses formulées, on appliquera soit un test bilatéral, soit un test unilatéral.

2.3.1 Test bilatéral

Un test bilatéral s'applique quand on cherchera une différence entre deux paramètres, ou entre un paramètre et une valeur donnée sans se préoccuper des détails sur le signe ou le sens de la différence.

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases} ; \quad \begin{cases} H_0 : \theta_1 = \theta_2 \\ H_1 : \theta_1 \neq \theta_2 \end{cases} .$$

2.3.2 Tests unilatéral

Un test unilatéral s'applique quand on cherchera à savoir si un paramètre est supérieur (respectivement inférieur) à un autre ou à une valeur donnée et on parlera d'un test unilatéral à droite (respectivement à gauche).

- Test unilatéral à droite : $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases} ; \quad \begin{cases} H_0 : \theta_1 = \theta_2 \\ H_1 : \theta_1 > \theta_2 \end{cases} .$
- Test unilatéral à gauche : $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases} ; \quad \begin{cases} H_0 : \theta_1 = \theta_2 \\ H_1 : \theta_1 < \theta_2 \end{cases} .$

2.4 La démarche de réalisation d'un test sous SPSS

On peut résumer la démarche d'un test sous SPSS de la manière suivante :

1. Choix des hypothèses H_0 et H_1.
2. Fixation du seuil de risque α.
3. Choix du type du test qu'on doit appliquer (selon le problème posé).
4. Exécution du test sous SPSS.
5. Interprétation des résultats : Prise de décision à l'aide de la valeur "signification (Sig)" fournie par le logiciel SPSS : rejet ou non rejet de H_0 au risque α.

2.5 Test de conformité de la moyenne sous SPSS

Il existe plusieurs types de tests paramétriques, dont l'objet est de tester certaine hypothèse relative à un ou plusieurs paramètres d'une variable aléatoire de loi spécifique, qui sont regroupés principalement en deux classes, à savoir : la classe des tests de conformité et la classe des tests d'homogénéité.

Dans cette section, nous allons s'intéresser sur le test de Student pour la conformité d'une moyenne.

2.5.1 Formulation d'un test de conformité de la moyenne

Dans ce test d'hypothèses, on veut valider l'hypothèse que la moyenne μ de la population entière diffère d'une certaine valeur μ_0, qui est déterminée par une expérience précédente ou par analogie avec une situation semblable. Par exemple, dans une certaine expérience, on veut voir si le rendement moyen d'un engrais pour la culture du blé (tonnes par hectare) est significativement supérieur à 8 tonnes/hectare. Alors, on pose les hypothèses suivantes :

$$\begin{cases} H_0 : \mu = 8, \\ H_1 : \mu > 8. \end{cases}$$

Dans le cas général, les hypothèses à tester sont de la forme suivante

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \quad \text{ou } (H_1 : \mu > \mu_0)\,;\,(H_1 : \mu < \mu_0)\ . \end{cases}$$

2.5.2 Test T pour la conformité de la moyenne sous SPSS

Pour expliquer les étapes à suivre dans le logiciel SPSS pour la mise en œuvre du test de conformité d'une moyenne reprenant l'exemple 1 présenté dans le chapitre précédent et on se pose dans ce qui suit :

"Au vu de l'échantillon de l'exemple 1, peut-on considérer, au seuil de risque 5%*, que la taille moyenne des enfants est significativement égale à* 110 *cm ?"*

La formulation des hypothèses du test souhaité dans cette exemple est exprimé comme suit :

$$\begin{cases} H_0 : \mu = 110 \\ H_1 : \mu \neq 110 \end{cases}.$$

La procédure de validation d'une hypothèse sur la valeur de la moyenne d'une population lorsqu'on dispose d'un seul échantillon, plus précisément le test de Student pour la conformité de la moyenne, dans le logiciel SPSS, est la même procédure que celle de l'estimation par intervalle de confiance d'une moyenne c'est-à-dire "**Test T pour échantillon unique**". En effet, pour répondre aux questions de l'exemple à l'aide de logiciel SPSS, il suffit de suivre les étapes suivantes :

Etape 1. Saisie des données :

Entrez les données dans SPSS, dont nous avons une seule variable quantitative (**Tailles des enfants**) à définir dans SPSS : cliquez sur "**Affichages des variables**" puis saisissez tous ses caractéristiques : **Nom**, **Type**, **Largeur**,...), ensuite cliquez sur "**Affichage des données**" pour entrer les valeurs des variables où chaque colonne représente une variable (Voir Figure 1.2).

Etape 2. Test T pour échantillon unique sous SPSS : Sélectionnez sur la barre de menu

Analyse → Comparer les moyennes → Test T pour échantillon unique (Voir Figure 1.3).

Etape 3. En cliquant sur " **Test T pour échantillon unique**" une fenêtre sera affichée dont la forme est présentée dans la figure 2.1. Alors il faut faire ce qui suit :

1. Mettez la variable **Taille enfant** dans la case "**Variables à tester**".
2. Mettez la valeur de μ_0 dans la case "**Valeur de test**". Dans le cas de notre exemple, il faut mettre la valeur **110**.

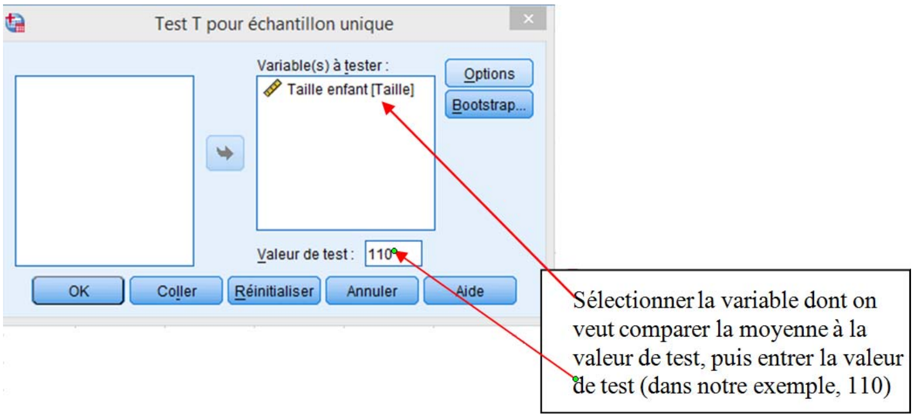


FIGURE 2.1: *Etapes de la réalisation d'un test de conformité de la moyenne sous SPSS.*

3. Une fois que les étapes 1-3 sont chevées sans erreurs, il reste à cliquez sur le boutton "**OK**" pour obtenir les résultats. Les résultas obtenus dans le cas de l'exemple considéré sont présentés dans la figure 2.6.

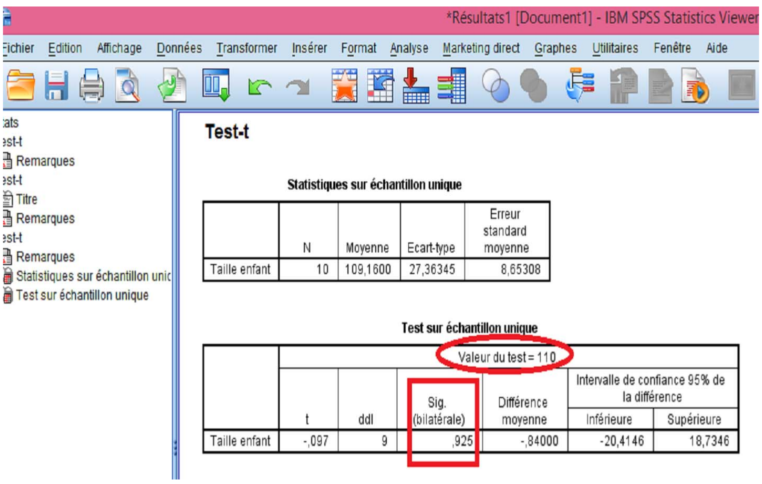


FIGURE 2.2: *Fenêtre des résultats du test de conformité de la moyenne sous SPSS.*

Etape 4. Interprétation des Résultats : Dans le cas d'un test bilatérale ($H_1 : \mu \neq \mu_0$), afin de prendre une décision sur le rejet ou le non rejet de l'hypothèse H_0, on procède comme suit :

$$\begin{cases} \text{non rejet de } H_0 \text{ au risque } \alpha\,; & \text{si } \alpha < signification\,; \\ \text{rejet de } H_0 \text{ au risque } \alpha\,; & \text{si } \alpha \geqslant signification\,; \end{cases}.$$

A partir des résultats obtenus sur notre exemple, on constate que pour un seuil de risque $\alpha = 5\%$, qu'on ne peut pas rejeter H_0, c'est-à-dire, la taille moyenne des enfants est significativement égale à 110 cm car $\alpha < signification$ (0.05 < 0.925).

Remarque 2.2 *Dans le cas d'un test unilatérale (i.e. cas $H_1 : \mu < \mu_0$ ou $H_1 : \mu > \mu_0$), afin de prendre une décision si on rejette ou non l'hypothèse H_0 on procède comme suit :*

- $\unrhd$ *on rejette pas H_0 si $2\alpha < signification$ (i.e. la moyenne μ est significativement égale à μ_0) ;*
- $\unrhd$ *on rejette H_0 si $2\alpha \geq signification$ (i.e. la moyenne μ significativement supérieur à μ_0 si $H_1 : \mu > \mu_0$ et la moyenne μ significativement inférieur à μ_0 si $H_1 : \mu < \mu_0$).*

2.6 Test d'homogénéité de deux moyennes sous SPSS

Contrairement aux tests de conformité qui sont utilisés lorsque nous désirons savoir si une caractéristique d'une population donnée correspond à une valeur fixe, les tests d'homogénéité consiste à vérifier si une caractéristique de deux ou de plusieurs populations sont semblables. Ça sous-entend qu'on ne peut parler sur les tests d'homogénéité que lorsque nous disposons de deux échantillons indépendents ou plus (≥ 2).

Dans ce passage nous allons exposer les étapes de la mise en œuvre du test d'homogénéité de Student, pour la comparaison des moyennes de deux échantillons, sous SPSS.

2.6.1 Formulation d'un test d'homogénéité de deux moyennes

Dans ce type de tests, on veut valider l'hypothèse que la moyenne μ_1 de la première population diffère ou non de la moyenne μ_2 d'une deuxième population. Par exemple, dans une certaine expérience, on veut voir si le rendement moyen d'un engrais de `Type 1` pour la culture du blé est significativement supérieur au rendement moyen d'un engrais de `Type 2`. Dans ce cas, on devrait poser les hypothèses suivantes :

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

Dans le cas général, les hypothèses à tester sont de la forme suivante :

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \quad \text{ou } (H_1 : \mu_1 > \mu_2)\,; \,(H_1 : \mu_1 < \mu_2) \end{cases}.$$

2.6.2 Test T pour l'hmogénéité de moyennes sous SPSS

Pour expliquer les étapes à suivre dans le logiciel SPSS pour la mise en œuvre du test d'homogénéité de moyennes considérons l'exemple numérique suivant :

Exemple 2 *:* Afin de comparer deux types d'arbre vis-à-vis leurs hauteurs, nous avons réalisés un recueil de hauteur de quelques arbres, dont les mesures sont rangées dans le tableau suivant.

Arbre 1	23.3	24.0	24.3	24.5	25.0	25.9
Arbre 2	21.1	21.1	22.1	22.4	23.3	

Question : Supposons qu'on désir savoir si les deux types d'arbres ont la même hauteur en moyenne. Alors, pour un seuil de risque $\alpha = 2\%$, que peut-on conclure sur la hauteur moyenne des deux types d'arbres.

D'après l'énoncé de l'exemple, le test qui nous permettera de répendre à la question posé est le test d'homogéniété de moyennes de deux populations dont la formulation est :

$$H_0 : \mu_1 = \mu_2 \quad \text{contre} \quad H_1 : \mu_1 \neq \mu_2.$$

Pour réaliser ce dernier test à l'aide du logiciel SPSS, il suffit de suivre les étapes suivantes :

Etape 1. *Introduction des données :* Les données des deux échantillons doivent être introduites comme étant une seule variable (i.e. comme étant un seul échantillon), et d'ajouter une deuxième variable qui indique le numéro de l'échantillon dont l'observation X_i provienne c'est-à-dire de transformer nos données sous forme de couple (X_i, G_i), où X_i est une observation et G_i est numéro de l'échantillon qui contient cette observation. Ainsi pour les données de notre exemple on aura ce qui suit :

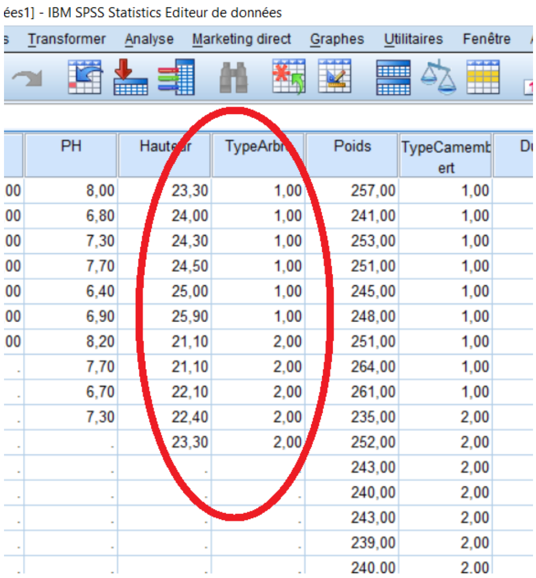


Figure 2.3: *Introduction des données pour le test d'homogénéité T sous SPSS.*

Etape 2. Réalisation de l'estimation sous SPSS : Sélectionnez sur la barre de menu

Analyse → Comparer les moyennes → Test T pour échantillons indépendants.

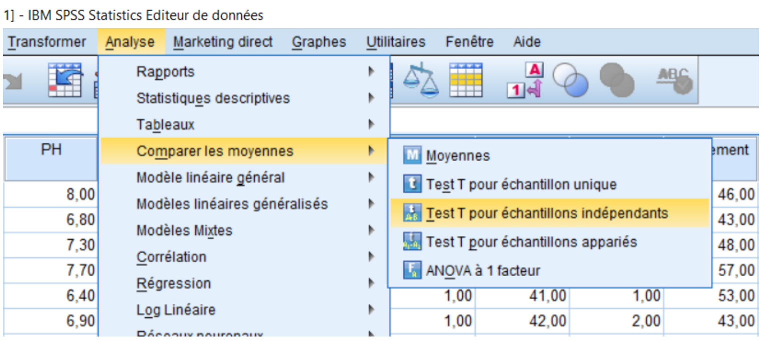


Figure 2.4: *Sélection du test d'homogénéité dans SPSS.*

Etape 3. En cliquant sur " **Test T pour échantillons indépendants**" une fenêtre sera affichée dont la forme est présentée dans la figure 2.5. Les taches qu'il faut effectuer sur cette fenêtre sont :

(A) Mettez la variable **Hauteur** dans la case "**Variable(s) à tester**".

(B) Mettez la variable **TypeArbre** dans la case "**Critère de regroupement qualitatif numérique**".

(C) Sélectionnez la variable **TypeArbre** est cliquez sur le bouton **Définir des groupes...** et une petite nouvelle fenêtre vas apparaitre.

(D) Au niveau des cases Groupe introduire les numéros des deux échantillons souhaité à comparer. Dans notre cas "***1***" pour l'échantillon de la hauteur de premier type d'arbre et "***2***" pour l'échantillon de la hauteur de deuxième type d'arbre.

(E) Une fois les groupes sont définis il faut cliquer sur le bouton **Poursuivre**.

(F) Il reste qu'à cliquer sur le bouton **OK** pour afficher les résultats du test.

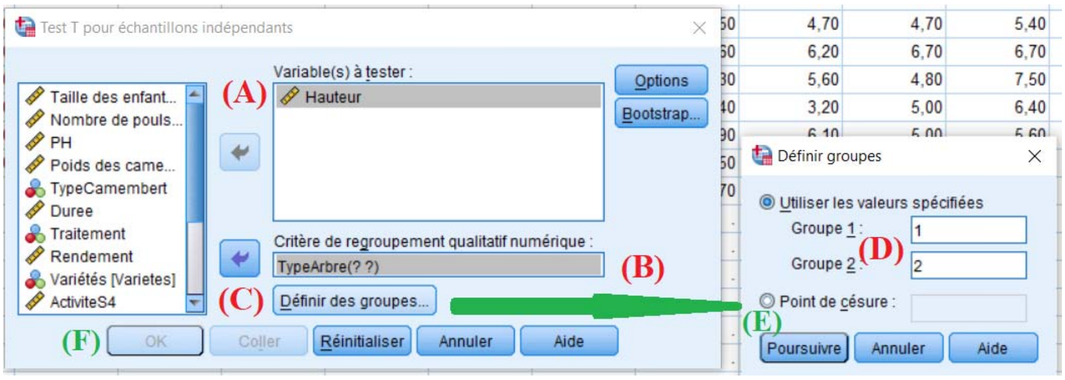


FIGURE 2.5: *Etapes de la réalisation d'un test d'homogénéité T de deux moyennes sous SPSS.*

Etape 4. Prise de décisions : Une fois que les étapes 1–3 sont chevées sans erreurs deux tableaux de résultats seront affichés dans la fenêtre d'affichage des résultats où le premier tableau résume quelques caractéristiques descriptives des deux échantillons (la taille d'échantillon, la moyenne, l'écart-type et l'erreur standard moyenne) tandis que le deuxième tableau résume les résultats associés purement au test de Student pour l'homogénéité de moyennes. Les résultats obtenus, dans le cas de l'exemple considéré, sont présentés dans la figure 2.6.

Test-t

[Ensemble_de_données1] C:\Users\mbi\Desktop\DonnéesSeries.sav

Statistiques de groupe

	TypeArbre	N	Moyenne	Ecart-type	Erreur standard moyenne
Hauteur	Arbre 1	6	24,5000	,88769	,36240
	Arbre 2	5	22,0000	,93274	,41713

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
Hauteur	Hypothèse de variances égales	,076	sig1= ,789	4,547	9	sig2= ,001	2,50000	,54981	1,25623	3,74377
	Hypothèse de variances inégales			4,524	8,461	sig3= ,002	2,50000	,55257	1,23775	3,76225

FIGURE 2.6: *Résultats du test d'homogénéité T de deux moyennes sous SPSS.*

Dans le cadre d'un test bilatérale (i.e. $H_1 : \mu_1 \neq \mu_2$) et pour un seuil de risque α, la décision sur le rejet ou non de l'hypothèse H_0 se fait selon le schéma suivant :

- ▷ Si $\alpha < sig_1$ alors les deux variances sont significativement égales et dans ce cas pour la décision sur les deux moyennes il faut utiliser la valeur de signification sig_2 où :
  - ⊵ on ne rejette pas H_0 si $\alpha < sig_2$ (i.e. les deux moyenne sont significativement égales).
  - ⊵ on rejette H_0 si $\alpha \geq sig_2$ (i.e. les deux moyenne sont significativement inégales).
- ▷ Si $\alpha \geq sig_1$ alors les deux variances sont significativement inégales (sont différentes) et dans ce cas pour la décision sur les deux moyennes il faut utiliser la valeur de signification sig_3 où :
  - ⊵ on ne rejette pas H_0 si $\alpha < sig_3$ (i.e. les deux moyenne sont significativement égales).
  - ⊵ on rejette H_0 si $\alpha \geq sig_3$ (i.e. les deux moyenne sont significativement inégales).

Ainsi, à partir des résultats obtenus sur les données de notre exemple, on déduit que les deux types d'arbres ont une variation de hauteur significativement égale par contre leurs hauteurs moyennes sont significativement différentes et ceci le fait que :

$$\alpha = 0.05 < sig_1 = 0.789 \ (\sigma_1^2 = \sigma_2^2) \text{ et } \alpha = 0.05 > sig_2 = 0.001 \ (\mu_1 = \mu_2).$$

Remarque 2.3 *Dans le cadre d'un test* ***unilatérale*** *(i.e. $H_1 : \mu_1 < \mu_2$ ou $H_1 : \mu_1 > \mu_2$), pour un seuil de risque α la décision sur le rejet ou non de l'hypothèse H_0 se fait comme suit :*

- ▷ *Si $\alpha < sig_1$ alors les deux variances sont significativement égales et dans ce cas pour la décision sur les deux moyennes, il faut utiliser la valeur de signification sig_2 où :*
  - ⊵ *on ne rejette pas H_0 si $2\alpha < sig_2$ (i.e. les deux moyenne sont significativement égales).*
  - ⊵ *on rejette H_0 si $2\alpha \geq sig_2$ (i.e. $\mu_1 < \mu_2$ si $H_1 : \mu_1 < \mu_2$ ou $\mu_1 > \mu_2$ si $H_1 : \mu_1 > \mu_2$).*
- ▷ *Si $\alpha \geq sig_1$ alors les deux variances sont significativement inégales (sont différentes) et dans ce cas pour la décision sur les deux moyennes, il faut utiliser la valeur de signification sig_3 où :*
  - ⊵ *on ne rejette pas H_0 si $2\alpha < sig_3$ (i.e. les deux moyenne sont significativement égales).*
  - ⊵ *on rejette H_0 si $2\alpha \geq sig_3$ (i.e. $\mu_1 < \mu_2$ si $H_1 : \mu_1 < \mu_2$ ou $\mu_1 > \mu_2$ si $H_1 : \mu_1 > \mu_2$).*

Chapitre 3

L'analyse de la variance à 1 facteur sous SPSS

Introduction

Dans ce chapitre nous allons présenter dans un premier lieu les principales étapes à suivre pour réaliser une analyse de la variance à un seul facteur à l'aide du logiciel SPSS. Par la suite, nous allons intéresser aux étapes de vérification de homocédasticité (égalité de variances) et de détermination des sous-ensembles homogènes lorsque le facteur à un effet significatif sur la variable quantitative analysée. De plus, pour que les étapes en question soient compréhensibles nous allons les illustrer à travers d'un exemple d'application.

Avant de présenter les étapes à suivre pour réaliser une ANOVA à 1 facteur sous SPSS, rappelons d'abord que les hypothèses (H_0 et H_1) à tester sont comme suit :

$$H_0 : \;''\mu_1 = \mu_2 = ... = \mu_p = \mu'' \text{ contre } H_1 : \;''\exists\; i, j \in \{1, 2, .., p\} \text{ tel que } \mu_i \neq \mu_j ;'' ,$$

avec p est le nombre d'échantillons dont on dispose.

Remarque 3.1 *Pour plus de détails sur la méthode ANOVA 1, le lecteur peut se référer au polycopié du cours disponible sur la plateforme de Moodle.*

Exemple 3 *Nous souhaitons comparer quatre traitements, notés A, B, C et D. Nous répartissons par tirage au sort les patients, et nous leur affectons l'un des quatre traitements. Nous mesurons sur chaque patient la durée, en jours, séparant de la prochaine crise d'asthme. Les mesures sont reportées dans le tableau ci-dessous :*

Traitement A	*Traitement B*	*Traitement C*	*Traitement D*
36 ; 37 ; 35 ; 38 ; 41	*42 ; 38 ; 39 ; 42 ; 44*	*26 ; 26 ; 30 38 ; 34*	*42 ; 45 ; 50 ; 56 ; 58*

Pouvons-nous conclure, à un seuil de risque α, que le facteur traitement a une influence sur la durée séparant la prochaine crise d'asthme ? Dans le cas affirmatif, quels sont les traitements qui ont le même impact sur la durée moyenne entre deux crises d'asthme ?

3.1 ANOVA à seul facteur sous SPSS

La méthode de l'ANOVA à 1 facteur se réalise sous le logiciel SPSS en générale en effectuant les étapes suivantes :

Étape 1. *Introduction des données :* les différents échantillons doivent être introduit comme étant une seule variable (un seul échantillon), et d'ajouter une deuxième variable qui indique le numéro de l'échantillon dont l'observation X_i provienne c'est-à-dire de transformer nos données sous forme de couple (X_i, G_i), où X_i est une observation et G_i est numéro de l'échantillon qui contient cette observation. Ainsi pour les données de notre exemple on aura ce qui suit :

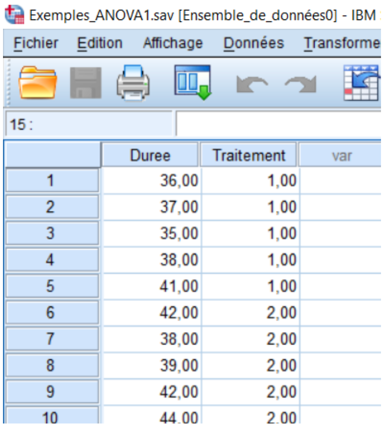


FIGURE 3.1: *Introduction des données Pour une ANOVA à une facteur sous SPSS.*

Étape 2 *Trouvez ANOVA à 1 facteur dans SPSS :* Pour localiser la méthode ANOVA à 1 facteur, dans le logiciel SPSS, il suffit de parcourir sur la barre de menu

Analyse → Comparer les moyennes → ANOVA à 1 facteur (voir figure 3.2).

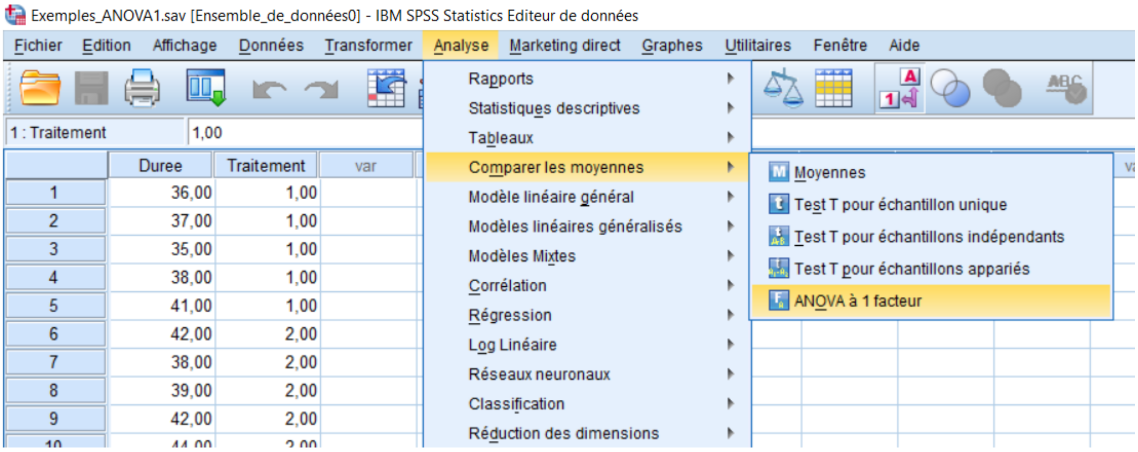


FIGURE 3.2: *Retrouver ANOVA à 1 facteur dans le logiciel SPSS.*

Étape 3 *Affectation des variables :* en cliquant sur " *ANOVA à 1 facteur*" (voir Etape 2) une fenêtre sera affichée dont la forme est présentée dans la figure 3.3 à gauche. Cette fenêtre contient trois espaces réservés pour les variables où :

La case A : Dans cette case on trouve la totalité des variables déclarées (dans notre exemple on n'a que deux variables, à savoir : la variable "Durée" et la variable "Traitement").

La case B : Cette espace est réservé pour la variable quantitative que nous désirons analyser. Dans notre exemple ça correspond à la variable "Durée".

La case C : Cette dernière case est réservée pour la variable qualitative (le facteur) qui indique le regroupement des observations. Dans notre exemple cette case est réservé à la variable de regroupement G_i c'est-à-dire à la variable "Traitement".

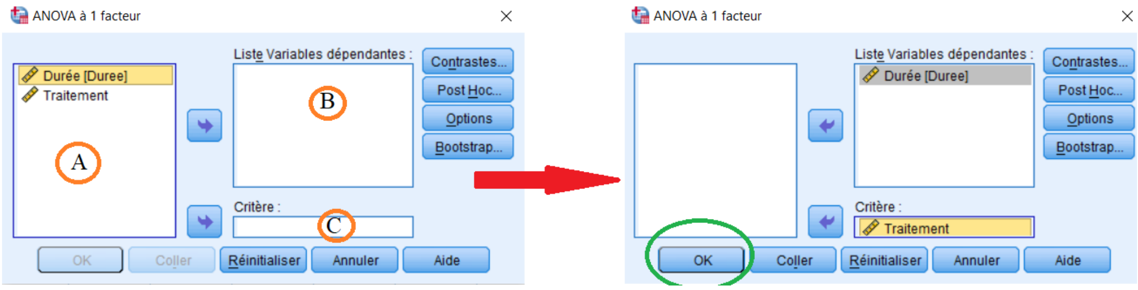
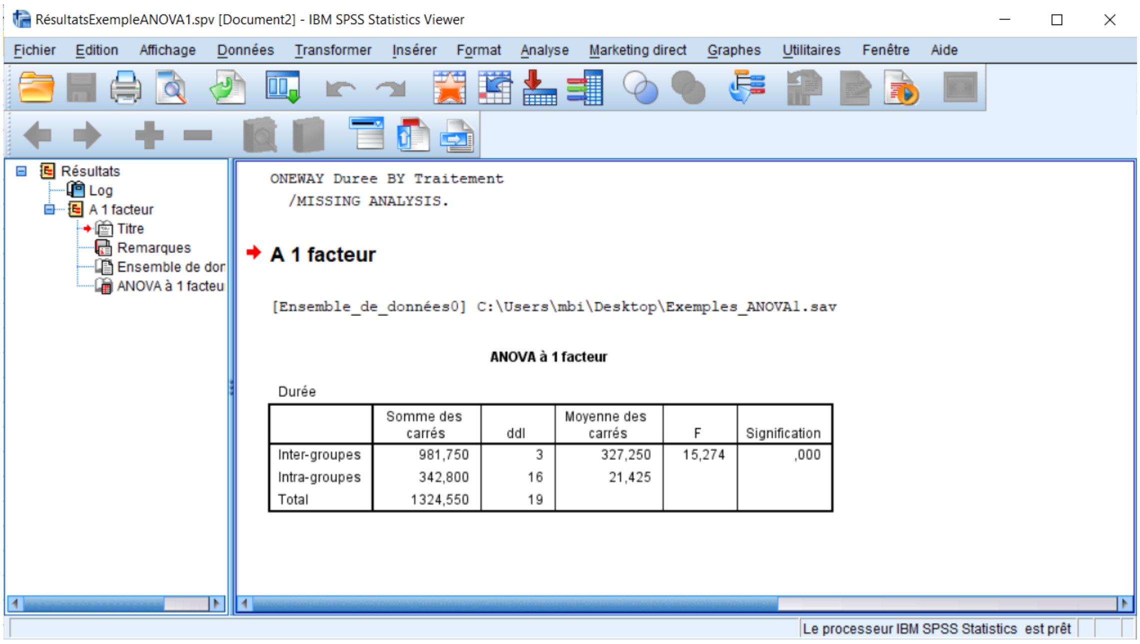


FIGURE 3.3: *Listes des variables et affectation des variables*

Étape 4 *Résultats : Résultats :* Une fois que l'étape 3 est achevée, il suffit de cliquer sur le Bouton "***OK***" (voir figure 3.3 à droite) alors dans toute une nouvelle fenêtre (fenêtre spéciale pour l'affichage des résultats) on aura les résultats de l'ANOVA à 1 facteur correspondant aux données qu'on sélectionnées dans l'étape 3. Dans le cas de notre exemple les résultats qu'on obtient sont présentés dans la figure 3.4.



	Somme des carrés	ddl	Moyenne des carrés	F	Signification
Inter-groupes	981,750	3	327,250	15,274	,000
Intra-groupes	342,800	16	21,425		
Total	1324,550	19			

FIGURE 3.4: *Fenêtre des résultats dans SPSS*

Étape 5 *Interprétation des résultats :* Afin d'en prendre une décision si le facteur à un effet significative ou non sur la variable quantitative analysée il suffit de comparer le seuil de risque α, fixé préalablement par l'analyste, avec la valeur " signification" affichée dans la dernière colonne de la table d'ANOVA à 1 facteur où on conclut :

$$\begin{cases} \text{Le facteur n'a pas un effet significatif } (H_0), & \text{si } \alpha < signification\,; \\ \text{Le facteur a un effet significatif } (H_1), & \text{si } \alpha \geq signification\,; \end{cases}$$

Ainsi, pour notre exemple on constate que si on fixe, par exemple, le seuil de risque à $\alpha = 1\%$, le facteur traitement à un effet significatif sur la durée séparant la prochaine crise d'asthme car $\alpha > signification$.

Remarque 3.2 *Pour les détails de la table d'ANOVA à 1 facteur le lecteur peut se référer au polycopie du cours.*

3.2 ANOVA à 1 facteur : Conditions et Classes homogènes

L'ANOVA en générale et à l'ANOVA 1 facteur en particulier ne permet de détecter que si toutes les moyennes sont les mêmes (H_0) ou si y a au moins une moyenne qui est différente des autres (H_1). Dans cette dernière situation, une question s'impose :

Comment savoir et déterminer les quelles de ces moyennes qui sont différentes et les quelles qui sont homogènes ?

✓ Si on ne rejette pas H_0, ce qui signifier que le facteur n'a pas d'effet sur la variables quantitative analyser. Dans ce cas on peut s'arrêter car tous les populations où les échantillons ont été prélevés ont significativement les mêmes moyennes.

✓ Si on rejette H_0, ce qui signifier qu'il y a au moins une moyenne qui est différente des autres. Dans ce cas, afin de déterminer les classes homogènes on dispose de plusieurs et différents tests (tests de comparaison multiples) tels : le test de Bonferroni, le test de Tukey, le test de Dunnett, le test de Sidak, le test de Scheffee,...

Mais ce qu'il faut retenir est que les tests de comparaison multiples dépendent de l'homogénéité de variances des différents échantillons. En effet, les tests de comparaisons multiples sont regroupés en deux catégories, à savoir : l'ensemble des tests qui sont utilisés sous l'hypothèse d'homogénéité de variances (exemple : Bonferroni, Tukey,...) et l'ensemble des tests qui sont utilisés sous l'hypothèse de variances inégales (exemple : T2 de Tamhan, C de Dunnett,...). Par conséquent, avant de procéder aux comparaisons multiples des moyennes, il est nécessaire de vérifier si les échantillons ont les mêmes variances ou non.

Ci-dessous une illustration des étapes à suivre sous SPSS pour vérifier l'hypothèse d'égalité de variances et la sélection des tests de comparaisons multiples.

3.2.1 Test d'homogénéité de variances

Reprenant les étapes 2 et 3 de l'ANOVA à 1 facteur citées précédemment dans la section 3.1 (" *sélectionner ANOVA à 1 facteur dans SPSS*" et " *Affectation des variables* ") mais avant de procéder à l'affichage des résultats c'est-à-dire avant de cliquer sur le bouton ***OK***, il faut réaliser ce qui suit :

1. sélectionner le Bouton ***Option***, qui se trouve sur la droite de la fenêtre. Une fois que qu'on clique sur ce dernier bouton, on aura une autre nouvelle fenêtre qui va apparaitre où il faut cocher la case " *Test d'homogénéité de variance*" (voir sur la figure 3.5).
2. cliquer sur le bouton ***Poursuivre***.
3. par la suite, cliquer sur le bouton ***OK*** pour afficher les résultats.

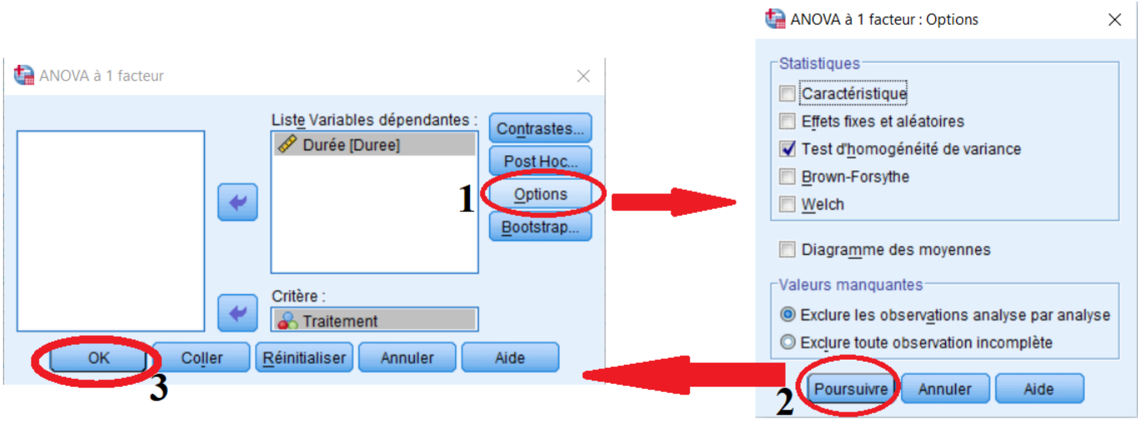
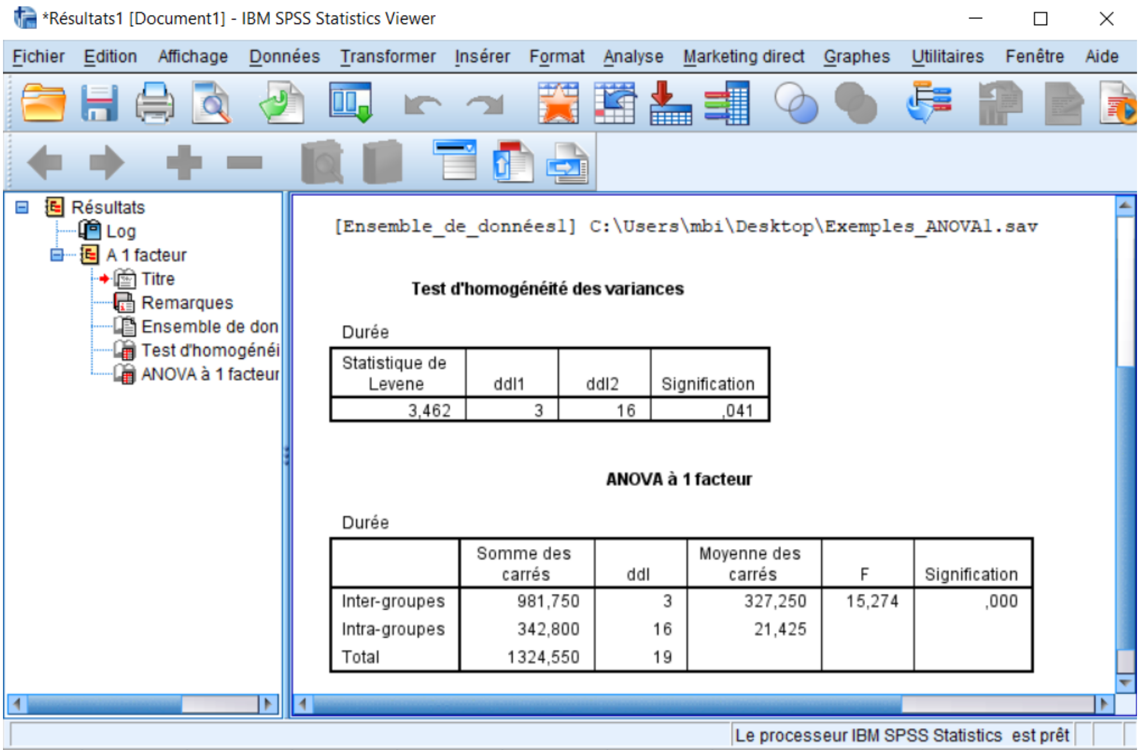


FIGURE 3.5: *Fenêtre des résultats de test d'homogénéité de variance.*

Si c'est trois étapes ont été réalisées convenablement sur nos données présentées dans l'exemple 1, les résultats qui seront affichés doivent être comme suit :



Test d'homogénéité des variances

Durée

Statistique de Levene	ddl1	ddl2	Signification
3,462	3	16	,041

ANOVA à 1 facteur

Durée

	Somme des carrés	ddl	Moyenne des carrés	F	Signification
Inter-groupes	981,750	3	327,250	15,274	,000
Intra-groupes	342,800	16	21,425		
Total	1324,550	19			

FIGURE 3.6: *Fenêtre des résultats de test d'homogénéité de variance et d'ANOVA à 1 facteur.*

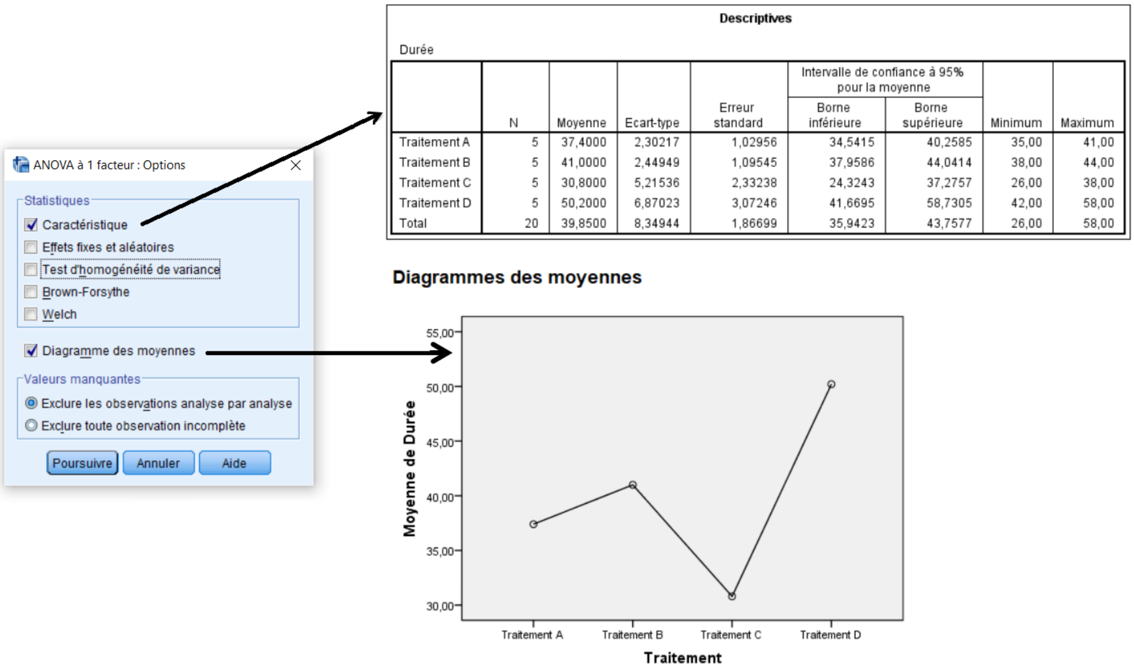
Afin de prendre une décision si les variances sont égales ou non on procède comme suit :

$$\begin{cases} Si\ \alpha < signification & \text{les variance sont égales ;} \\ Si\ \alpha \geq signification & \text{les variance sont inégales ;} \end{cases}$$

A partir des résultats obtenus sur notre exemple, on constate que pour un seuil de risque $\alpha = 5\%$, les variances sont significativement différentes car $\alpha > signification = 0.041$. Par contre, si on fixe le seuil de risque à une valeur $\alpha < 4\%$, on déduit que les variances sont égales $\alpha < signification = 0.041$.

Remarque 3.3 *Dans la fenêtre " ANOVA à 1 facteur : Options " (voir figures 3.5 et 3.7) si on coche la case " Caractéristiques " et la case " Diagramme des moyennes " on aura respective-*

ment les caractéristiques descriptives de chaque échantillons et le diagramme de la variation des moyennes d'un échantillon à un autre. Les résultats qu'on obtient sur les données notre exemple sont présentés dans la figure 3.7.



Descriptives

Durée

	N	Moyenne	Ecart-type	Erreur standard	Intervalle de confiance à 95% pour la moyenne		Minimum	Maximum
					Borne inférieure	Borne supérieure		
Traitement A	5	37,4000	2,30217	1,02956	34,5415	40,2585	35,00	41,00
Traitement B	5	41,0000	2,44949	1,09545	37,9586	44,0414	38,00	44,00
Traitement C	5	30,8000	5,21536	2,33238	24,3243	37,2757	26,00	38,00
Traitement D	5	50,2000	6,87023	3,07246	41,6695	58,7305	42,00	58,00
Total	20	39,8500	8,34944	1,86699	35,9423	43,7577	26,00	58,00

FIGURE 3.7: *Statistiques descriptives des échantillons et diagrammes des moyennes*

3.2.2 Test de comparaisons multiples (Post Hoc)

Pour réaliser le test de comparaisons multiples, on doit reprendre les étapes 2 et 3 de l'ANOVA à 1 facteur citées dans la section 3.1, mais avant de cliquer sur le bouton ***OK*** pour afficher les résultats, il faut cette fois-ci :

1. sélectionner le Bouton ***Post Hoc***. Une fois cliqué sur le bouton ***Post Hoc*** on aura une autre nouvelle fenêtre (voir figure 3.8) où cette dernière contient deux listes de tests de comparaisons multiples, à savoir : la liste des tests conditionnés par l'hypothèse de variances égales (voir 'A' dans la figure 3.8) et la liste des tests conditionnés par l'hypothèse de variances inégales (voir 'B' dans la figure 3.8). Ce qui fait, il suffit de cocher la case du test souhaité à réaliser toute en prenant en considération l'hypothèse imposée sur les variances.
2. fixer le seuil de risque α. Cette étape n'est pas obligatoire pour certains tests car leurs résultats sont indépendant de la probabilité α.
3. cliquer sur le bouton ***Poursuivre***.
4. En fin, cliqué sur le bouton ***OK*** pour afficher les résultats.

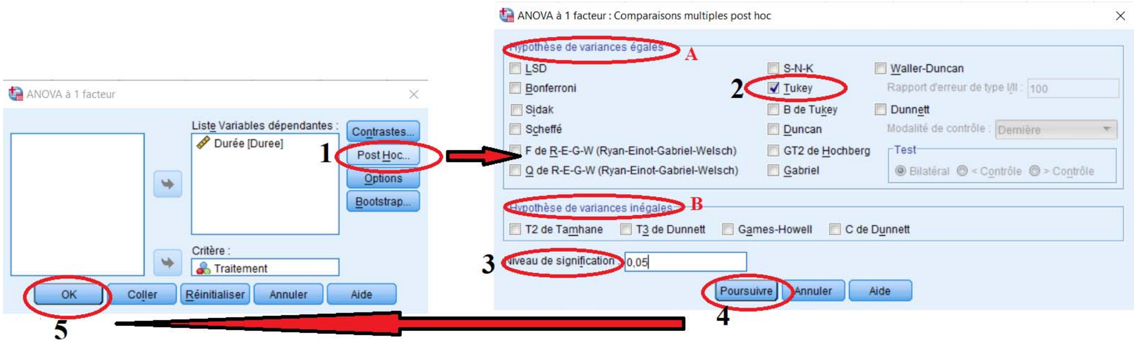
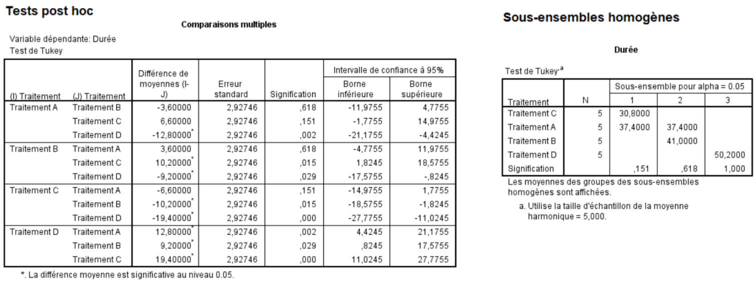


FIGURE 3.8: *Sélection d'un test de comparaisons multiples dans SPSS.*

Si on opte pour le test de comparaisons multiples de Tukey et un seuil de risque $\alpha = 5\%$ les résultats qu'on obtient qui serons affichés pour les données de notre exemple d'application sont présentés dans la figure 3.9.



Tests post hoc

Comparaisons multiples

Variable dépendante: Durée
Test de Tukey

(I) Traitement	(J) Traitement	Différence de moyennes (I-J)	Erreur standard	Signification	Intervalle de confiance à 95% Borne inférieure	Borne supérieure
Traitement A	Traitement B	-3,60000	2,92746	,618	-11,9755	4,7755
	Traitement C	6,60000	2,92746	,151	-1,7755	14,9755
	Traitement D	-12,80000*	2,92746	,002	-21,1755	-4,4245
Traitement B	Traitement A	3,60000	2,92746	,618	-4,7755	11,9755
	Traitement C	10,20000*	2,92746	,015	1,8245	18,5755
	Traitement D	-9,20000*	2,92746	,029	-17,5755	-,8245
Traitement C	Traitement A	-6,60000	2,92746	,151	-14,9755	1,7755
	Traitement B	-10,20000*	2,92746	,015	-18,5755	-1,8245
	Traitement D	-19,40000*	2,92746	,000	-27,7755	-11,0245
Traitement D	Traitement A	12,80000*	2,92746	,002	4,4245	21,1755
	Traitement B	9,20000*	2,92746	,029	,8245	17,5755
	Traitement C	19,40000*	2,92746	,000	11,0245	27,7755

*. La différence moyenne est significative au niveau 0.05.

Sous-ensembles homogènes

Durée

Test de Tukey[a]

Traitement	N	Sous-ensemble pour alpha = 0.05 1	2	3
Traitement C	5	30,8000		
Traitement A	5	37,4000	37,4000	
Traitement B	5		41,0000	
Traitement D	5			50,2000
Signification		,151	,618	1,000

Les moyennes des groupes des sous-ensembles homogènes sont affichées.

a. Utilise la taille d'échantillon de la moyenne harmonique = 5,000.

FIGURE 3.9: *Illustration des sous-ensembles homogènes fournis par le de Tukey.*

D'après le tableau "*sous-ensembles homogènes*" (voir figure 3.9 à droite), on remarque que pour un seuil de risque $\alpha = 5\%$, il existe trois sous-ensembles qui sont significativement homogènes, à savoir :

Sous-ensemble 1 : Traitement C et Traitement A ($\mu_C = \mu_A$).

Sous-ensemble 2 : Traitement A et Traitement B ($\mu_A = \mu_B$).

Sous-ensemble 3 : Traitement D.

On constate également que le *Traitement A* appartient à la fois au premier sous-ensemble et au deuxième sous-ensemble. Enfin, l'illustration graphique de ces sous-ensembles peut être schématisée comme suit :

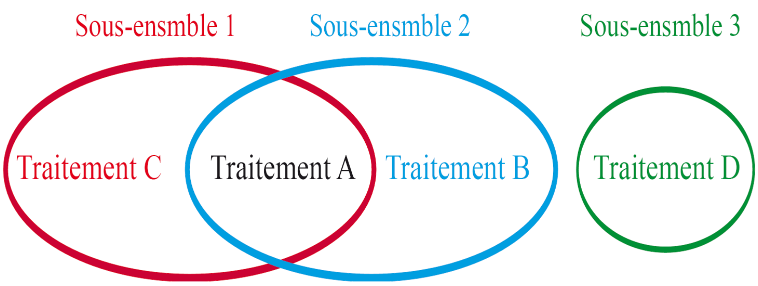


FIGURE 3.10: *Résultats du test de comparaisons multiples de Tukey*

Il est à signaler que

$$\left\{ \begin{array}{l} \mu_C = \mu_A \\ \mu_A = \mu_B \end{array} \right. \nRightarrow \mu_C = \mu_B \; ,$$

car ce n'est pas une comparaison purement mathématique.

Remarque 3.4

- *Dans l'exemple d'application, le choix du test de comparaisons multiples de Tukey pour* $\alpha = 5\%$ *n'est pas vraiment adéquat voir même faux. En effet, pour* $\alpha = 5\%$ *il est claire que les variances des échantillons sont inégales (voir figure 3.6) hors le test de Tukey est test qu'on peut utiliser dans le cas où l'hypothèse d'égalité de variances est vérifiée. Alors, pour utiliser ce test il est plus judicieux de fixer le seuil de risque à une valeur inférieure à* 4% *(afin que l'hypothèse d'égalité de variances soit vraie).*
- *En réalité le tableau* sous-ensembles homogènes *n'est que le résumé du premier tableau pour le seuil de risque* α *fixé préalablement.*
- *Pour certains tests on aura uniquement un seul tableau soit le premier tableau ou le deuxième tableau seulement mais pas les deux au même temps.*

Chapitre 4

Tableaux croisés et Test d'indépendance de χ^2

Introduction

Dans ce chapitre nous allons illustrer, à travers trois exemples d'application, comment réaliser le test d'indépendance de χ^2. En effet, dans un premier lieu, nous allons s'intéresser aux cas de données préalablement traitées et qui sont regroupées sous forme d'un tableau croisé. Dans un second lieu, l'objectif est de distinguer le cas d'utilisation des données brutes (sans un traitement préalable). En fin, nous allons présenter un exemple illustratif du cas de traitement d'indépendance de plus de deux variables mais ce dernier exemple sera laissé comme étant un exercice pour l'étudiant.

4.1 Test d'indépendance de χ^2 : cas de données regroupées

Exemple 4 *On s'intéresse à l'association entre le mode de vie, "seul" ou "en famille", et la présence ou l'absence d'une névrose. Dans un échantillon aléatoire d'individus d'une certaine population on a trouvé les effectifs ci-dessous :*

	Névrose		
Mode de vie	*Présente*	*Absente*	*Total*
En famille	*40*	*60*	*100*
Seul	*100*	*60*	*160*
Total	*140*	*120*	*260*

Peut-on rejeter, au seuil de α, l'hypothèse de non association entre le mode de vie et la présence d'une névrose ?

Pour répondre cette question, qui se fait via le test d'indépendance de χ^2, à l'aide du logiciel SPSS il faut suivre les étapes suivantes :

Étape 1. *Introduction des données :* les différents effectifs observés, N_{ij}, doivent être introduits comme étant une seule variable, et d'ajouter deux autres variables qui font référence aux variables analysées. Dans notre exemple, les deux variables correspondraient à la variable *Névrose* et la variable *Mode de vie.* Ainsi pour notre échantillon, les informations doivent être introduites tel qu'il est illustré dans la figure 4.1.

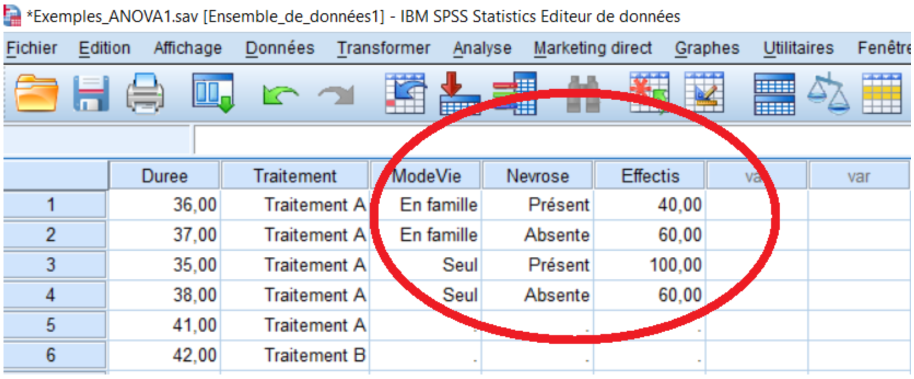


FIGURE 4.1: *Introduction des données pour le test de χ^2 : cas de données en Tableau croisé.*

Étape 2 *Pondération des observations :* Pour pondérées les observations il faut réaliser, successivement, les 04 étapes suivantes (voir figure 4.2) :

1. Cliquer sur le symbole de la ***Balançoire*** où une petite fenêtre va apparaitre,
2. Cocher la case "*Pondérées les observations par*",
3. Affecter la variable contenant les effectifs dans la case " *Variable d'effectif :*",
4. Cliquer sur le bouton ***OK*** pour terminer la pondération des variables.

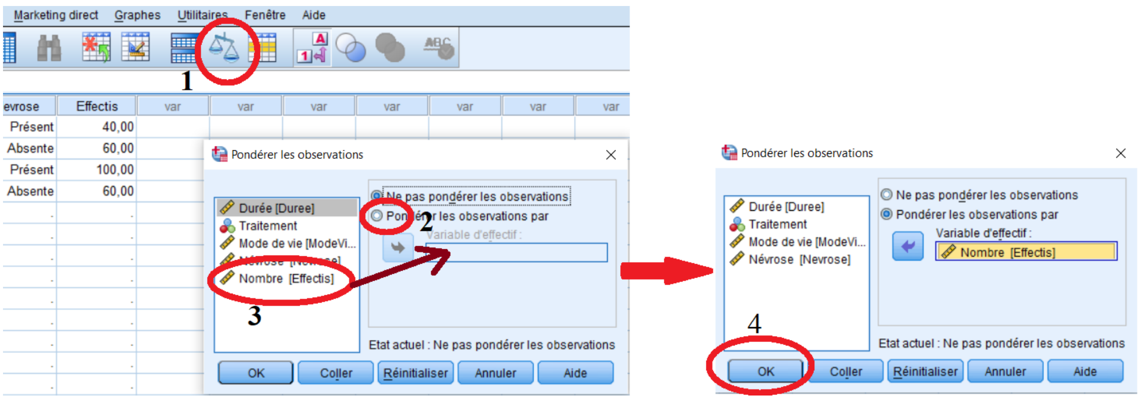


FIGURE 4.2: *Pondération des observations dans SPSS.*

Étape 3 *Tableaux croisés et test d'indépendance de χ^2 :*

– *Reconstruction du tableau croisé :* pour localiser la procédure de construction et reconstruction de tableaux croisés, il suffit de sélectionner sur la barre de menu ce qui suit :

Analyse → Statistiques descriptives → Tableaux croisés (voir figure 4.3).

Une fois qu'on clique sur ***Tableaux croisés*** une fenêtre va apparaitre et dans cette dernière, il suffit de sélectionner, dans la liste des variables, les deux variables que nous souhaitons à analyser et d'affecter l'une d'entre elles dans la case *ligne(s)* et l'autre dans la case *colonne(s)* et en fin de cliquer sur le bouton ***OK*** pour l'affichage des résultats (voir figure 4.3).

Il est à signaler que le choix de la variable qu'on doit affecter dans la ligne et la variable qui doit être affecter dans la colonne n'est pas important dans cette situation. L'inversement de l'emplacement des deux variables n'est que la transposition du tableau original des données et qui n'affecte pas les calculs ultérieur.

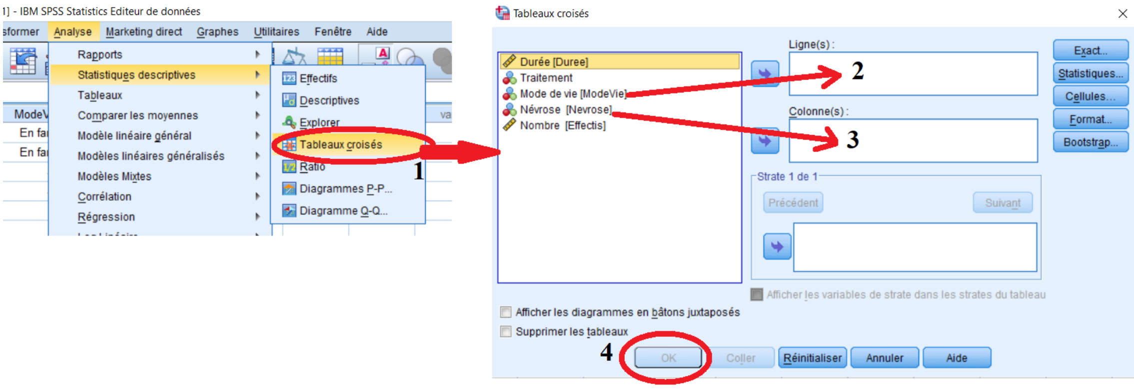


FIGURE 4.3: *Liste des variables et affectation des variables lignes et colonnes.*

Si on applique ces dernières étapes sur les données de notre exemple numérique les résultats qui serons affichés sont présentés dans la figure 4.4.

Récapitulatif du traitement des observations

	Observations					
	Valide		Manquante		Total	
	N	Pourcent	N	Pourcent	N	Pourcent
Mode de vie * Névrose	260	100,0%	0	0,0%	260	100,0%

Tableau croisé Mode de vie * Névrose

Effectif

		Névrose		Total
		Présent	Absente	
Mode de vie	En famille	40	60	100
	Seul	100	60	160
Total		140	120	260

FIGURE 4.4: *Résultats de reconstruction du tableau croisé.*

– *Test d'indépendance de χ^2 :*
Pour localiser le test d'indépendance de χ^2 dans le logiciel SPSS, il suffit de sélectionner sur la barre de menu ce qui suit :

Analyse → Statistiques descriptives → Tableaux croisés (voir figure 4.3).

Une fois qu'on clique sur ***Tableaux croisés*** une fenêtre va apparaitre et dans cette dernière, il suffit de sélectionner, dans la liste des variables, les deux variables que nous souhaitons à analyser et d'affecter l'une d'entre elles dans la case *ligne(s)* et l'autre dans la case *colonne(s)* (voir figure 4.3). Par la suite, il faut cliquer sur le bouton "***Statistiques***" pour qu'une deuxième fenêtre aille apparaitre. Dans cette dernière, il faut cocher la case qui correspond à la terminologie "*Chi-deux*" et de cliquer par la suite sur le bouton ***Poursuivre*** (voir figure 4.5).

FIGURE 4.5: *Sélection du test d'indépendance de χ^2 dans SPSS.*

Enfin, pour afficher les résultats, il ne reste qu'à cliquer sur le bouton ***OK***. Les résultats qui serons affiché pour notre exemple d'application sont présentés dans la figure 4.6.

Tests du Khi-deux

	Valeur	ddl	Signification asymptotique (bilatérale)	Signification exacte (bilatérale)	Signification exacte (unilatérale)
Khi-deux de Pearson	12,536[a]	1	,000		
Correction pour la continuité[b]	11,647	1	,001		
Rapport de vraisemblance	12,594	1	,000		
Test exact de Fisher				,001	,000
Association linéaire par linéaire	12,488	1	,000		
Nombre d'observations valides	260				

FIGURE 4.6: *Résultats du test d'indépendance de χ^2*

Étape 4 *Interprétation des résultats :* Afin d'en juger s'il existe un lien (dépendantes) ou non (indépendantes) entre les deux variables il suffit de comparer le seuil de risque α, fixé préalablement par l'analyste, avec la valeur "*signification*" où on décide selon la règle suivante :

$$\begin{cases} \text{Les deux variables sont indépendantes,} & \text{si } \alpha < signification\,; \\ \text{Les deux variables sont dépendantes),} & \text{si } \alpha \geq signification\,; \end{cases}$$

Pour un seuil de risque $\alpha = 1\%$, le fait que $\alpha > signification$ on conclut que la présence et l'absence d'une névrose chez un individu dépend de son mode de vie (voir figure 4.6).

Autres résultats : En plus des effectifs observés, N_{ij}, on peut également afficher d'autres résultats dans le tableau croisé tels : les effectifs théorique (attendus, n_{ij}), les écart ($N_{ij} - n_{ij}$, i.e. écart entre les effectifs observés et les effectifs théorique), le pourcentage des effectifs,... Pour ce faire, il suffit qu'à l'étape 3 (reconstruction du tableau croisé et test du χ^2) avant de cliquer sur le bouton ***OK*** pour l'affichage des résultats, de cliquer le bouton ***Cellules***. Ce dernier, nous permettra de visualiser une nouvelle fenêtre où il faut cocher les cases correspondantes aux paramètres que nous souhaitons à afficher. La figure 4.7 illustre un exemple d'affichage des effectifs observés (cocher la case A), les effectifs théoriques (cocher la case B) et l'écart entre ces deux effectifs (cocher la case C).

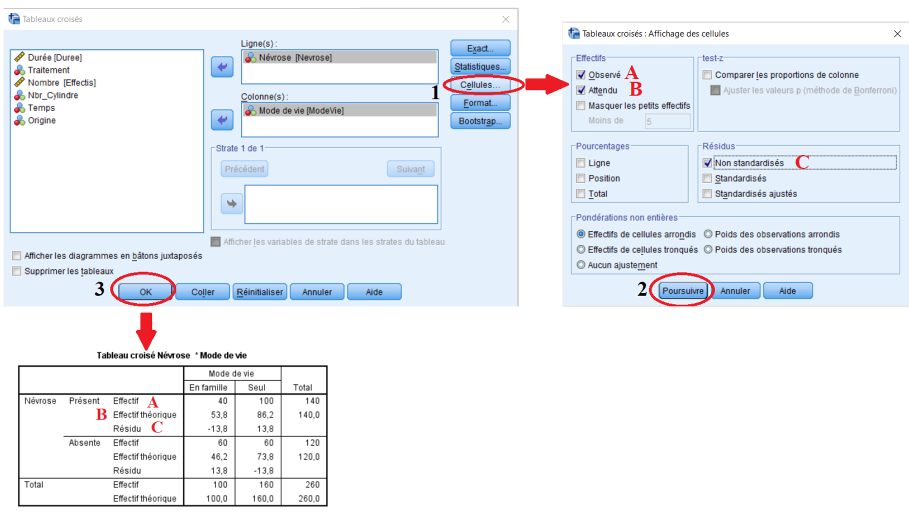


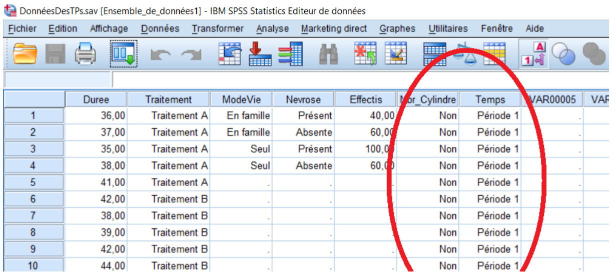
FIGURE 4.7: *Autres résultats : Plus de détails dans le tableau croisé.*

Remarque 4.1

- *Le test de χ^2 n'est pas l'unique test d'indépendance existants dans les statistiques, En effet, dans les statistiques plusieurs autres tests visent à vérifier l'existence d'une association entre les variables on peut citer par exemple : le test du rapport de vraisemblance, le test de Fisher, test Correction pour la continuité (utilisé seulement pour le cas de tableaux croisé de dimension 2×2),... La réalisation du test d'indépendance de χ^2 sous SPSS, visualise les résultats fournis par certains de ces tests (voir figure 4.6).*
- *Pour une autre éventuelle analyse statistiques dans le même fichier de données, il faut prendre en compte la pondération des observations effectuées au niveau de l'étape 2 c'est-à-dire il faut désactiver cette pondération ou de la changer et cela selon la nouvelle analyse prévue à réaliser.*

4.2 Test d'indépendance de χ^2 : cas de données brutes

Exemple 5 *On souhaite vérifier si le nombre de cylindre puissance (quatre cylindre ou autre) des voitures fabriquées entre 1970 et 1982 dépend de la période (période 1 : 1970–1974, période 2 : 1975–1979 et période 3 : 1980–1982) de son fabrication ou non. Les informations collectées sur 406 voitures sont introduite dans le logiciel SPSS tel qu'il est illustré dans la figure suivante.*



	Duree	Traitement	ModeVie	Nevrose	Effectis	...r_Cylindre	Temps	VAR00005	VAR
1	36,00	Traitement A	En famille	Présent	40,00	Non	Période 1	.	
2	37,00	Traitement A	En famille	Absente	60,00	Non	Période 1	.	
3	35,00	Traitement A	Seul	Présent	100,0	Non	Période 1	.	
4	38,00	Traitement A	Seul	Absente	60,0	Non	Période 1	.	
5	41,00	Traitement A	.	.	.	Non	Période 1	.	
6	42,00	Traitement B	.	.	.	Non	Période 1	.	
7	38,00	Traitement B	.	.	.	Non	Période 1	.	
8	39,00	Traitement B	.	.	.	Non	Période 1	.	
9	42,00	Traitement B	.	.	.	Non	Période 1	.	
10	44,00	Traitement B	.	.	.	Non	Période 1	.	

FIGURE 4.8: *Introduction des données brutes pour le test d'indépendance de χ^2.*

Dans ce cas, les données n'ont pas subi une analyse préliminaire (regroupement) d'où l'effectif de chaque observation est $N_{ij} = 1$. Pour réaliser le test d'indépendance dans ce type de situation en réalité on doit suivre les mêmes étapes que celles de la section 4.1 mais cette fois-ci on élimine l'étape 2 qui concerne la pondération des données c'est-à-dire on doit réaliser uniquement l'étape 1, l'étape 3 et l'étape 4. Les résultats qu'on obtient dans ce cas, sur nos données, sont présentés dans la figure 4.9.

Tableaux croisés

Récapitulatif du traitement des observations

	Observations					
	Valide		Manquante		Total	
	N	Pourcent	N	Pourcent	N	Pourcent
Nbr_Cylindre * Temps	406	100,0%	0	0,0%	406	100,0%

Tableau croisé Nbr_Cylindre * Temps

			Temps			Total
			Période 1	Période 2	Période 3	
Nbr_Cylindre	Non	Effectif	85	82	32	199
		Effectif théorique	64,7	76,0	58,3	199,0
		Résidu	20,3	6,0	-26,3	
	Oui	Effectif	47	73	87	207
		Effectif théorique	67,3	79,0	60,7	207,0
		Résidu	-20,3	-6,0	26,3	
Total		Effectif	132	155	119	406
		Effectif théorique	132,0	155,0	119,0	406,0

Tests du Khi-deux

	Valeur	ddl	Signification asymptotique (bilatérale)
Khi-deux de Pearson	36,739[a]	2	,000
Rapport de vraisemblance	37,874	2	,000
Association linéaire par linéaire	34,634	1	,000
Nombre d'observations valides	406		

a. 0 cellules (0,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 58,33.

FIGURE 4.9: *Résultats du test de χ^2 : cas des données brutes.*

D'après les résultats du test de χ^2, on constate que le nombre de cylindre dont une voiture est dotée dépend fortement de la période de fabrication de cette voiture (signification est 0.000).

4.3 Test d'indépendance de χ^2 : cas plus de deux variables

Reprenant l'exemple 5, mais supposons qu'on a introduit un troisième critère qui est le pays d'origine (les Etats unies d'Amérique, le Japon et l'Europe) de la fabrication de la voiture (voir figure 4.10 et pour plus de détails voir le fichier *DonnéesDesTPs.sav*). De plus, nous souhaitons également dans ce cas de vérifier si les trois variables retenues sont indépendantes ou au moins y a deux variables qui sont dépendantes.

Nbr_Cylindre	Temps	Origine	va
Non	Période 1	États unies	
Non	Période 1	États unies	
Non	Période 1	États unies	
Non	Période 1	États unies	
Non	Période 1	États unies	
Non	Période 1	États unies	
Non	Période 1	États unies	
Non	Période 1	États unies	
Non	Période 1	États unies	
Non	Période 1	États unies	

FIGURE 4.10: *Sélection du test d'indépendance de χ^2*

La résolution de ce problème se fait également à l'aide du test d'indépendance de chi^2.

Question : quelles sont les étapes à suivre, dans cette situation, sous SPSS pour réaliser le test d'indépendance de χ^2.

La reponse à cette question est laissée comme un exercice d'application pour l'étudiant.