

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

***BIOSTATISTIQUES***

Statistiques Appliquées à l'Expérimentation  
En Sciences Biologique  
sous SPSS

Préparé par :  
Dr. CHERFAOUI Mouloud

Université de Biskra, 2019/2020

# Table des matières

<b>1</b>	<b>L'analyse de la variance à 1 facteur sous SPSS</b>	<b>1</b>
	Introduction . . . . .	1
1.1	ANOVA à 1 facteur sous SPSS avec paramètres par défaut . . . . .	1
1.2	ANOVA à 1 facteur : Conditions et Classes homogènes . . . . .	4
1.2.1	Test d'homogénéité de variances . . . . .	4
1.2.2	Test de comparaisons multiples (Post Hoc) . . . . .	6
<b>2</b>	<b>Tableaux croisés et Test d'indépendance de <math>\chi^2</math></b>	<b>9</b>
	Introduction . . . . .	9
2.1	Test d'indépendance de $\chi^2$ : cas de données regroupées . . . . .	9
2.2	Test d'indépendance de $\chi^2$ : cas de données brutes . . . . .	13
2.3	Test d'indépendance de $\chi^2$ : cas plus de deux variables . . . . .	14

# 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 :

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

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

**Exemple 1** *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  $\alpha$ , 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 ?*

## 1.1 ANOVA à 1 facteur sous SPSS avec paramètres par défaut

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 :

	Duree	Traitement	var
1	36,00	1,00	
2	37,00	1,00	
3	35,00	1,00	
4	38,00	1,00	
5	41,00	1,00	
6	42,00	2,00	
7	38,00	2,00	
8	39,00	2,00	
9	42,00	2,00	
10	44,00	2,00	

FIGURE 1.1: Introduction des données Pour une ANOVA à une facteur sous SPSS.

**Remarque 1.1** Pour plus de détails sur les données voir le fichier **DonnéesDesTPs.sav** sur le site de l'université. Ceci reste valable pour la totalité des exemples traités dans ce document.

**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 1.2).

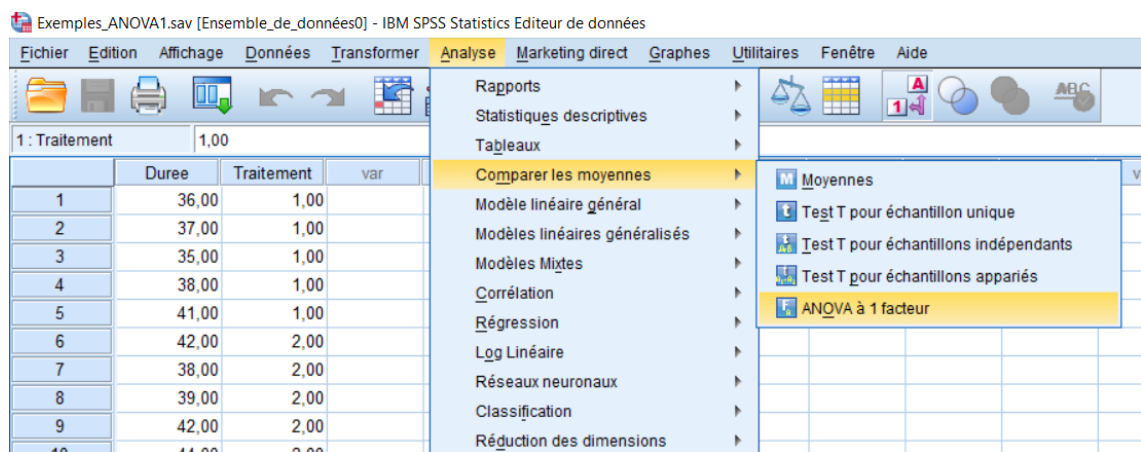


FIGURE 1.2: Retrouver ANOVA à 1 facteur dans le logiciel SPSS.

**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 1.3 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".

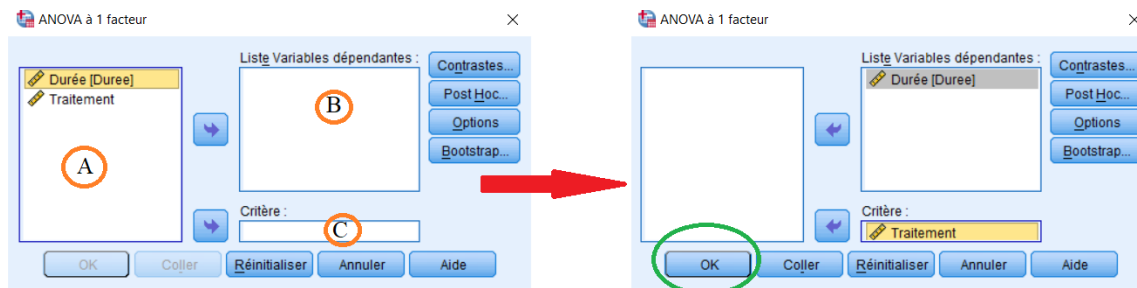


FIGURE 1.3: Listes des variables et affectation des variables

**Étape 4 Résultats :** Une fois que l'étape 3 est achevée, il suffit de cliquer sur le Bouton "OK" (voir figure 1.3 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 1.4.

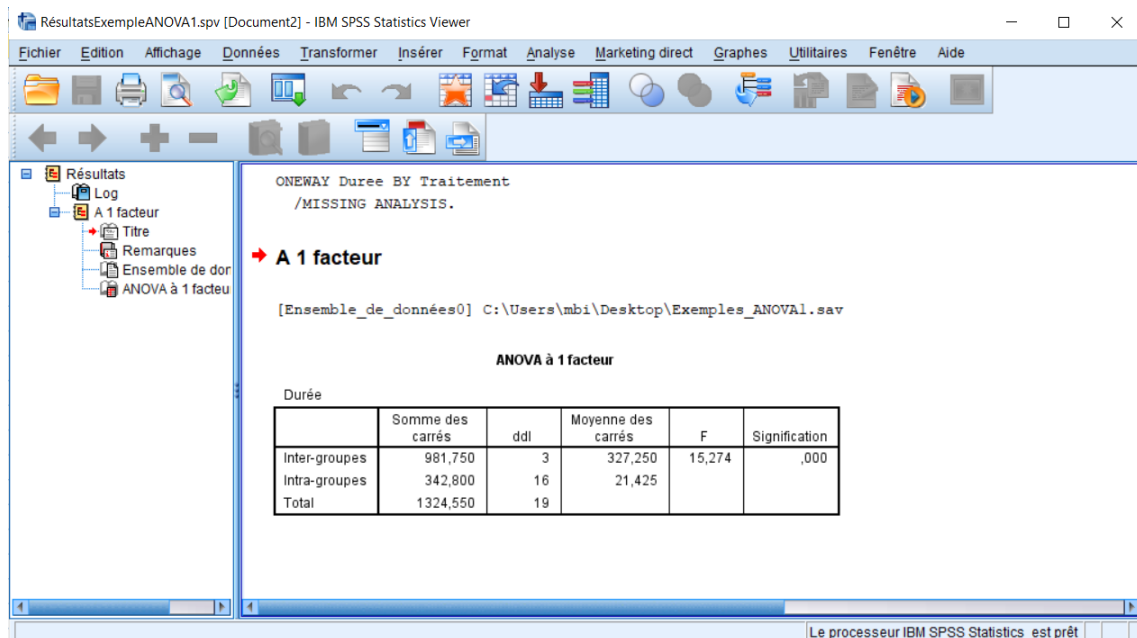


FIGURE 1.4: Fenêtre des résultats dans SPSS

**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  $\alpha$ , 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 :

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

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  $\alpha > \textit{signification}$ .

**Remarque 1.2** Pour les détails de la table d'ANOVA à 1 facteur le lecteur peut se référer au polycopie du cours.

## 1.2 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 signifie 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 signifie 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.

### 1.2.1 Test d'homogénéité de variances

Reprenant les étapes 2 et 3 de l'ANOVA à 1 facteur citées précédemment dans la section 1.1 (" *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 apparaître où il faut cocher la case " *Test d'homogénéité de variance*" (voir sur la figure 1.5).
2. cliquer sur le bouton **Poursuivre**.
3. par la suite, cliquer sur le bouton **OK** pour afficher les résultats.

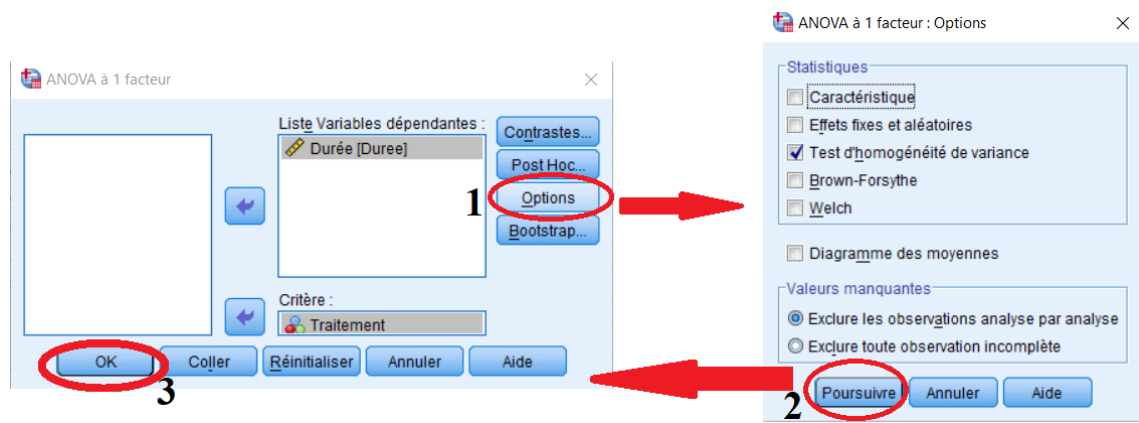


FIGURE 1.5: Fenêtre des résultats de test d'homogénéité de variance.

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 :

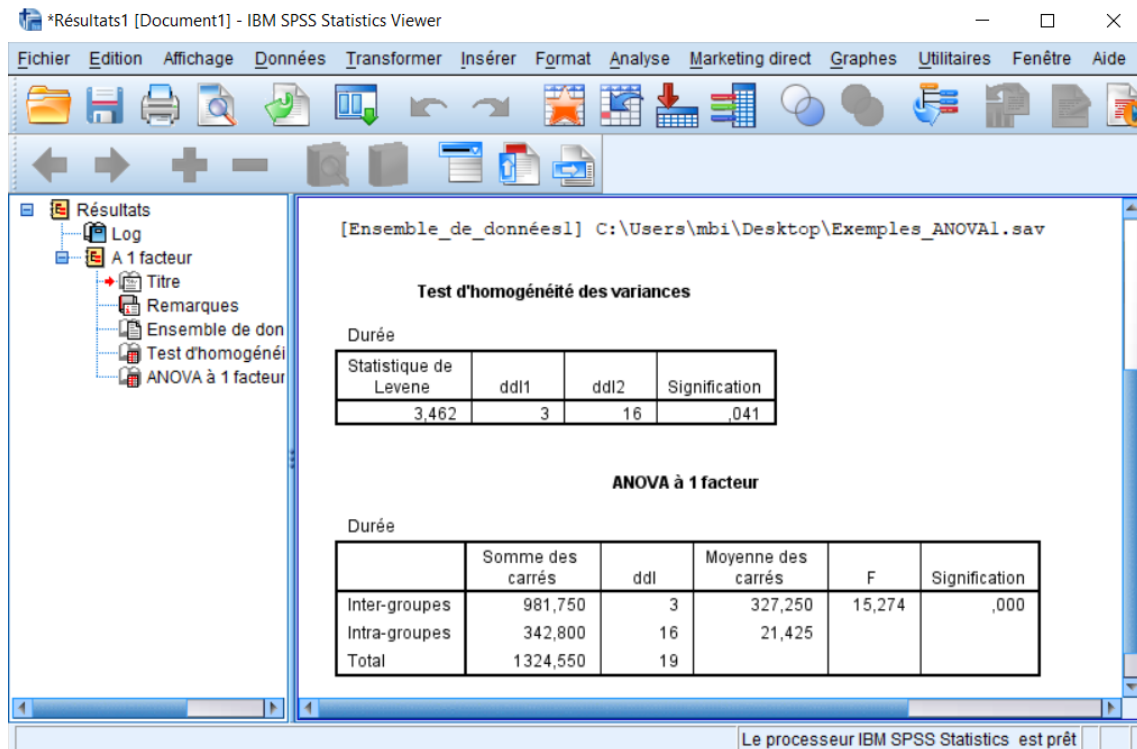


FIGURE 1.6: Fenêtre des résultats de test d'homogénéité de variance et d'ANOVA à 1 facteur.

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

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

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

**Remarque 1.3** Dans la fenêtre " ANOVA à 1 facteur : Options " (voir figures 1.5 et 1.7) si on coche la case " Caractéristiques " et la case " Diagramme des moyennes " on aura respective-

ment les caractéristiques descriptives de chaque échantillon 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 1.7.

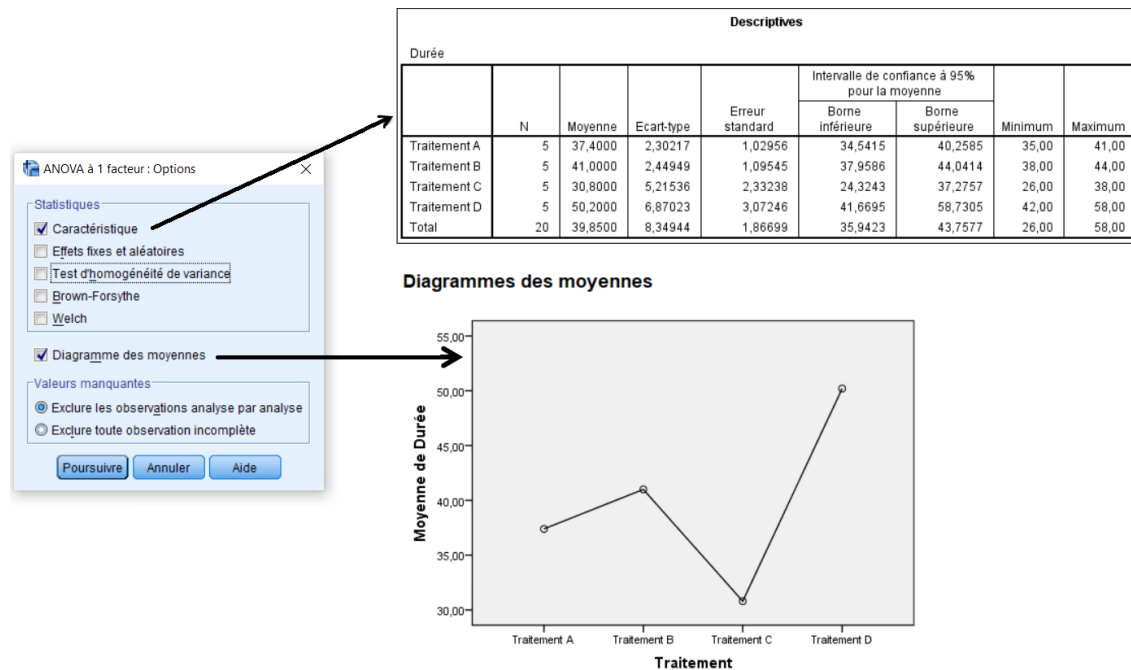


FIGURE 1.7: Statistiques descriptives des échantillons et diagrammes des moyennes

## 1.2.2 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 1.1, 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 1.8) 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 1.8) et la liste des tests conditionnés par l'hypothèse de variances inégales (voir 'B' dans la figure 1.8). 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  $\alpha$ . Cette étape n'est pas obligatoire pour certains tests car leurs résultats sont indépendant de la probabilité  $\alpha$ .
3. cliquer sur le bouton **Poursuivre**.
4. En fin, cliqué sur le bouton **OK** pour afficher les résultats.



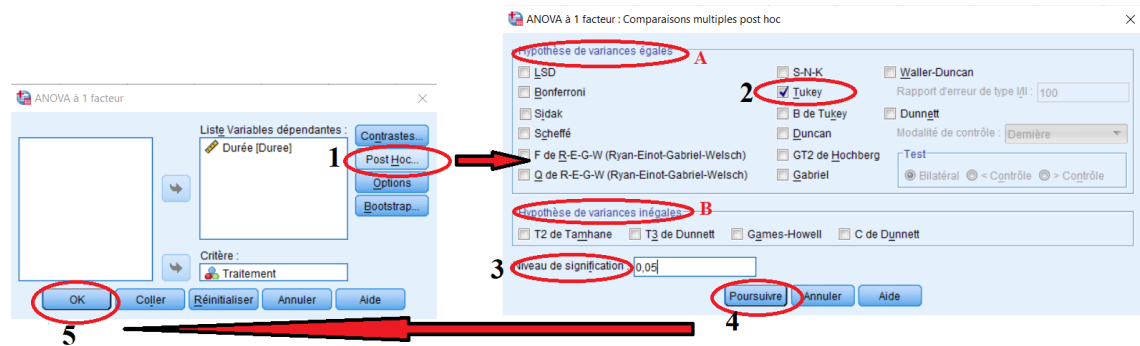


FIGURE 1.8: Sélection d'un test de comparaisons multiples dans SPSS.

Si on opte pour le test de comparaisons multiples de Tukey et un seuil de risque  $\alpha = 5\%$  les résultats qu'on obtient qui seront affichés pour les données de notre exemple d'application sont présentés dans la figure 1.9.

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 <sup>a</sup>	2,92746	,618	-11,9755	4,7755
Traitement C	Traitement D	-12,80000 <sup>a</sup>	2,92746	,002	-21,1755	-4,4245
Traitement B	Traitement A	3,60000	2,92746	,618	-4,7755	11,9755
Traitement B	Traitement C	10,20000 <sup>a</sup>	2,92746	,015	1,8245	18,5755
Traitement B	Traitement D	-9,20000 <sup>a</sup>	2,92746	,029	-17,5755	-,8245
Traitement C	Traitement A	-6,00000	2,92746	,151	-14,9755	1,7755
Traitement C	Traitement B	-10,20000 <sup>a</sup>	2,92746	,015	-18,5755	-1,9245
Traitement C	Traitement D	-19,40000 <sup>a</sup>	2,92746	,000	-27,7755	-11,0245
Traitement D	Traitement A	12,80000 <sup>a</sup>	2,92746	,002	4,4245	21,1755
Traitement D	Traitement B	9,20000 <sup>a</sup>	2,92746	,029	,8245	17,5755
Traitement D	Traitement C	19,40000 <sup>a</sup>	2,92746	,000	11,0245	27,7755

<sup>a</sup> La différence moyenne est significative au niveau 0.05.

Sous-ensembles homogènes				
Durée				
Test de Tukey <sup>a</sup>				
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.  
<sup>a</sup> Utilisez la taille d'échantillon de la moyenne harmonique = 5,000.

FIGURE 1.9: Illustration des sous-ensembles homogènes fournis par le de Tukey.

D'après le tableau "sous-ensembles homogènes" (voir figure 1.9 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 :

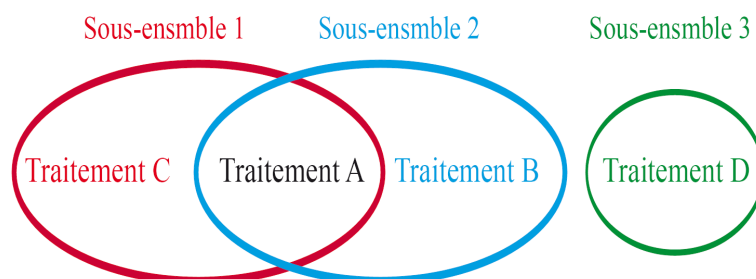


FIGURE 1.10: Résultats du test de comparaisons multiples de Tukey

Il est à signaler que

$$\begin{cases} \mu_C = \mu_A \\ \mu_A = \mu_B \end{cases} \not\Rightarrow \mu_C = \mu_B ,$$

car ce n'est pas une comparaison purement mathématique.

**Remarque 1.4**

- *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 clair que les variances des échantillons sont inégales (voir figure 1.6) hors le test de Tukey est un 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  $\alpha$  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.*

# Tableaux croisés et Test d'indépendance de $\chi^2$

## Introduction

Dans ce chapitre nous allons illustrer, à travers trois exemples d'application, comment réaliser le test d'indépendance de  $\chi^2$ . En effet, dans un premier lieu, nous allons nous 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.

## 2.1 Test d'indépendance de $\chi^2$ : cas de données regroupées

**Exemple 2** *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 :*

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

*Peut-on rejeter, au seuil de  $\alpha$ , 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  $\chi^2$ , 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 2.1.

	Duree	Traitement	ModeVie	Nevrose	Effectis	va	var
1	36,00	Traitement A	En famille	Présent	40,00		
2	37,00	Traitement A	En famille	Absente	60,00		
3	35,00	Traitement A	Seul	Présent	100,00		
4	38,00	Traitement A	Seul	Absente	60,00		
5	41,00	Traitement A					
6	42,00	Traitement B					

FIGURE 2.1: Introduction des données pour le test de  $\chi^2$  : 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 2.2) :

1. Cliquer sur le symbole de la **Balancoire** où une petite fenêtre va apparaître,
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.

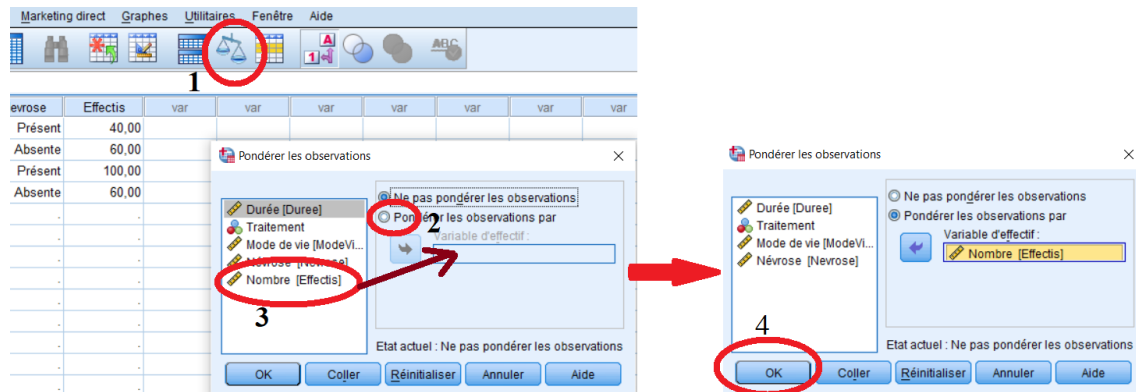


FIGURE 2.2: Pondération des observations dans SPSS.

**Étape 3** *Tableaux croisés et test d'indépendance de  $\chi^2$*  :

- *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 2.3).

Une fois qu'on clique sur **Tableaux croisés** une fenêtre va apparaître 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 2.3).

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.

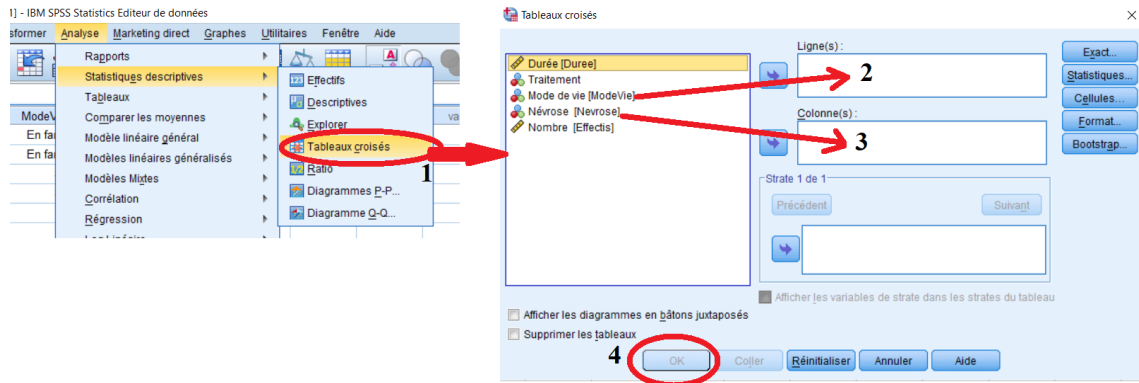


FIGURE 2.3: Liste des variables et affectation des variables lignes et colonnes.

Si on applique ces dernières étapes sur les données de notre exemple numérique les résultats qui seront affichés sont présentés dans la figure 2.4.

**Tableau croisé Mode de vie \* Névrose**

Récapitulatif du traitement des observations						Effectif					
	Observations						Névrose		Total		
	Valide		Manquante		Total		Présent	Absente			
	N	Pourcent	N	Pourcent	N	Pourcent					
Mode de vie * Névrose	260	100,0%	0	0,0%	260	100,0%	En famille	40	60	100	
							Seul	100	60	160	
							Total	140	120	260	

FIGURE 2.4: Résultats de reconstruction du tableau croisé.

– Test d'indépendance de  $\chi^2$  :

Pour localiser le test d'indépendance de  $\chi^2$  dans le logiciel SPSS, il suffit de sélectionner sur la barre de menu ce qui suit :

Analyse → Statistiques descriptives → Tableaux croisés (voir figure 2.3).

Une fois qu'on clique sur **Tableaux croisés** une fenêtre va apparaître 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 2.3). Par la suite, il faut cliquer sur le bouton "**Statistiques**" pour qu'une deuxième fenêtre aille apparaître. 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 2.5).

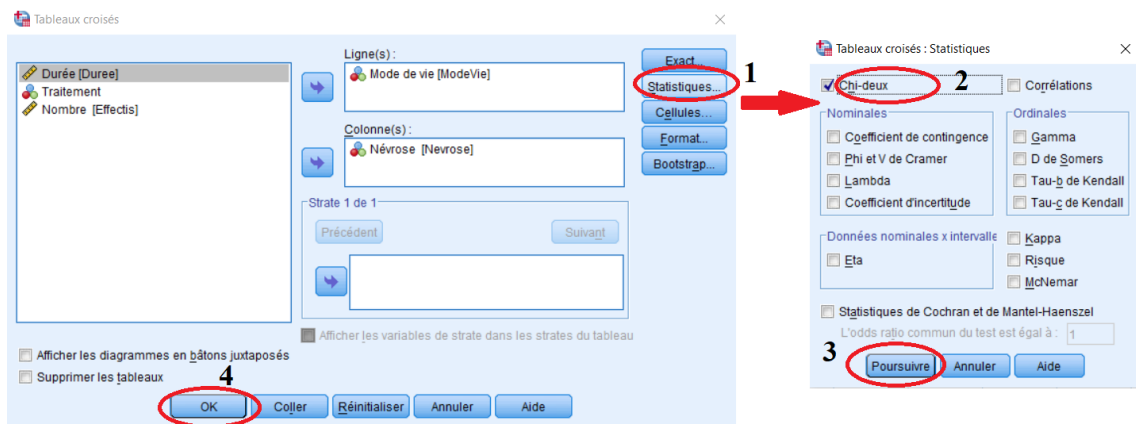


FIGURE 2.5: Sélection du test d'indépendance de  $\chi^2$  dans SPSS.

Enfin, pour afficher les résultats, il ne reste qu'à cliquer sur le bouton **OK**. Les résultats qui seront affichés pour notre exemple d'application sont présentés dans la figure 2.6.

	Valeur	ddl	Signification asymptotique (bilatérale)	Signification exacte (bilatérale)	Signification exacte (unilatérale)
Khi-deux de Pearson	12,536 <sup>a</sup>	1	,000		
Correction pour la continuité <sup>b</sup>	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 2.6: Résultats du test d'indépendance de  $\chi^2$

**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  $\alpha$ , fixé préalablement par l'analyste, avec la valeur "signification" où on décide selon la règle suivante :

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

Pour un seuil de risque  $\alpha = 1\%$ , le fait que  $\alpha > \textit{signification}$  on conclut que la présence et l'absence d'une névrose chez un individu dépend de son mode de vie (voir figure 2.6).

**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  $\chi^2$ ) 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 2.7 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).

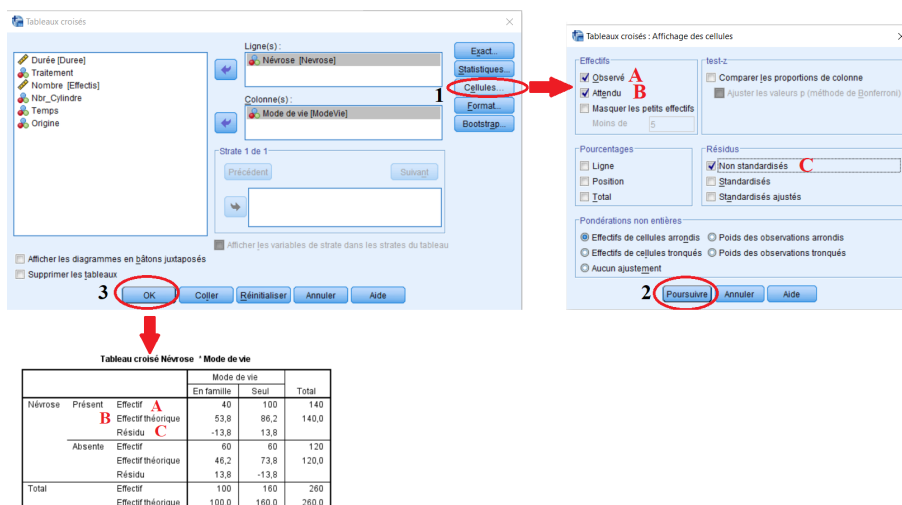


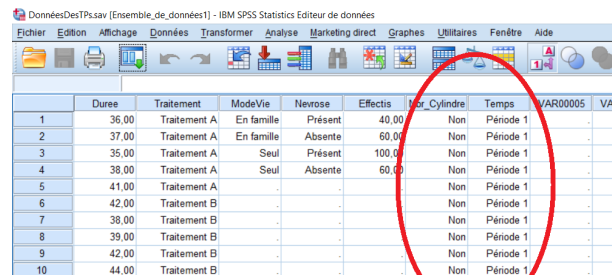
FIGURE 2.7: Autres résultats : Plus de détails dans le tableau croisé.

## Remarque 2.1

- Le test de  $\chi^2$  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  $2 \times 2$ ),... La réalisation du test d'indépendance de  $\chi^2$  sous SPSS, visualise les résultats fournis par certains de ces tests (voir figure 2.6).
- 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.

## 2.2 Test d'indépendance de $\chi^2$ : cas de données brutes

**Exemple 3** 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	Neurose	Effectifs	Nbr. 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	100.00	Non	Période 1		
4	38.00	Traitement A	Seul	Absente	60.00	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 2.8: Introduction des données brutes pour le test d'indépendance de  $\chi^2$ .

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 2.1 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 2.9.

**Tableaux croisés**

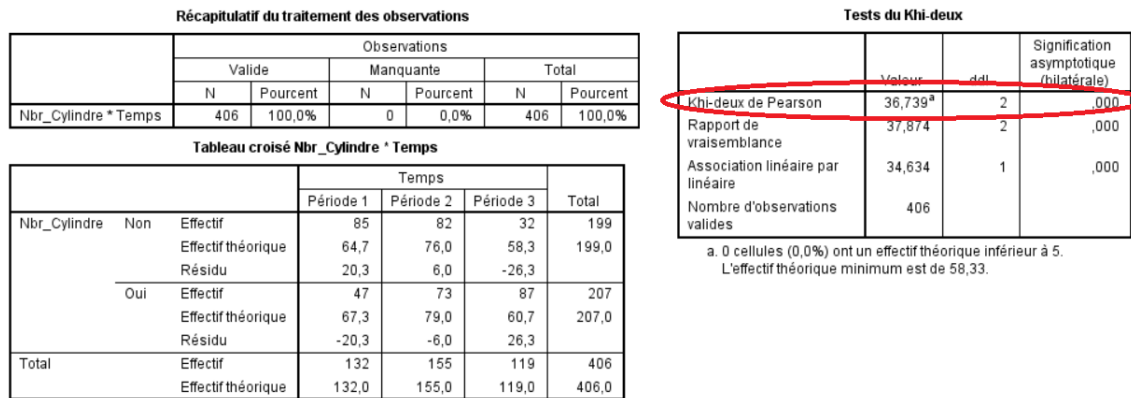


FIGURE 2.9: Résultats du test de  $\chi^2$  : cas des données brutes.

D'après les résultats du test de  $\chi^2$ , 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).

### 2.3 Test d'indépendance de $\chi^2$ : cas plus de deux variables

Reprenant l'exemple 3, 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 2.10 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	
Non	Période 1	États unies	

FIGURE 2.10: Sélection du test d'indépendance de  $\chi^2$

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  $\chi^2$ .

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