Revue Congolaise des Sciences Humaines et Sociales
Site web : http://recosh.cresh-rdc.org

ISSN :2957-6385 (print), ISSN : 2957-6393 (online)
https:/www. doi.org/10.59189/crsh102257

Analyse de Khi-carré : fondements, démarche, intérêt et limites

Khi-square analysis : foundations, approach, interest and limitations

Eugénie Kabali Hamuli[1] et José Mangalu Mobhe Agbada[2]

1 Professeure à l'Institut Supérieur Techniques Médicales de Kinshasa (ISTM/Kinshasa). e-mail : eugekab@yahoo.fr Tél : 0997260637
2 Professeur à l'Ecole des Sciences de la Population et du Développement (ESPD). Université de Kinshasa. e-mail : jose.mangalu@unikin.ac.cd Tél : 0997260637

Résumé. *En tant que l'une des techniques d'analyse des données les plus utilisées, le test de Khi-carré reste malheureusement très mal appliqué et surtout très mal interprété. Aussi, dans cet article, les auteur.e.s se propose de passer en revue, sous un éclairage nouveau et contextualisé, non seulement, les présupposés théoriques et les conditionnalités relatives à l'application de ce test, mais aussi, l'intérêt et les pièges à éviter dans son implémentation et dans l'interprétation des résultats qui en sont issus. Alors que la plupart d'ouvrages et autres articles sur la question se limitent principalement à l'étude de Khi-carré d'indépendance, dans le présent article, il a également d'autres types de test de Khi-carré comme le Khi-carré d'ajustement et le Khi-carré d'homogénéité. Par ailleurs, au-delà des discussions théoriques, des illustrations et des exemples concrets sont proposés dans le but de permettre aux lecteur.trice.s de se mettre en situation et d'appliquer la démarche théorique proposée. Cette façon de procéder parait d'autant plus importante que le test de Khi-carré constitue, dans la plupart des cas, un préalable à l'utilisation des tests statistiques les plus robustes comme la régression logistique, l'analyse discriminante, etc.*

Mots clés : Chi-carré, Fondements, Démarches, Intérêts, Limites et Pièges

Abstract. *As one of the most widely used data analysis techniques, the Chi-square test unfortunately remains very poorly applied and above all very misinterpreted. Also, in this article, the authors propose to review, in a new and contextualized light, not only the theoretical presuppositions and conditionalities relating to the application of this test, but also, the interest, limits and pitfalls to avoid in its implementation and in the interpretation of the results that come from it. While most books and other articles on the subject are mainly limited to the study of independence Chi-square, in this article it also has other types of Chi-square test such as fit Chi-square and homogeneity Chi-square. In addition, beyond the theoretical discussions, illustrations and concrete examples are proposed to allow readers to put themselves in a situation and apply the proposed theoretical approach. This approach seems more important since the Chi-square test is, in most cases, a prerequisite for the use of the most robust statistical tests such as logistic regression, discriminant analysis, etc.*

Key words: Chi-square, Foundations, Approaches, Interests, Limits and Pitfalls

OPEN ACCESS

Received: 19 november 2022
Accepted: 16 february 2022
available online: 9 june, 2023

I. Introduction

Dans cet article, il est question de traiter des fondements, de la démarche, de l'intérêt et des limites du test de Khi-carré. Le test de de Khi-carré (test de Khi-deux, test de Chi-deux ou X^2) constitue l'un des tests fondamentaux en matière d'analyse quantitative des données dans le domaine des sciences sociales. Dans notre pratique d'enseignement universitaire, nous avons été plusieurs fois témoins de difficultés que rencontrent les étudiant.e.s et autres chercheur.e.s juniors dans l'implémentation et surtout l'interprétation des résultats issus de ce test. Le principal objectif consiste à fournir à tou.te.s ceux/celles qui se sont déjà frotté.e.s à ce test les fondamentaux du test, à travers une prise en main étape par étape et en nous appuyant sur des exemples concrets tirés du contexte congolais.

En effet, le test de Khi-deux se propose d'étudier la relation entre deux variables catégorielles. Ce test est d'autant plus important qu'il constitue généralement une première étape pour des analyses un peu plus élaborées comme la régression logistique, l'analyse discriminante, l'analyse de correspondances, etc. Même si traditionnellement ce test est utilisé pour étudier les relations d'association ou d'indépendance entre deux variables catégorielles, on peut également l'utiliser pour étudier les différences de proportions entre les modalités d'une seule et même variable catégorielle sur un seul échantillon ou même la façon dont les individus d'un échantillon se répartissent entre les modalités d'une même variable catégorielle. On peut également recourir à ce test pour étudier la distribution d'une seule variable catégorielle sur deux ou plusieurs échantillons indépendants. Ainsi, sur le plan théorique, on peut distinguer globalement trois types de test de Khi-carré : le test de Khi-carré de conformité ou d'ajustement, le test de Khi-carré d'association ou d'indépendance et le test de Khi-carré d'homogénéité.

Concrètement, il sera question, dans un premier temps, d'aborder successivement les points suivants : les principes et les conditions d'utilisation du test de Khi-carré, la pose des hypothèses statistiques qui s'y rattachent, la présentation et l'interprétation des résultats. Dans un second temps, les fondements de chacun de trois types de test de Khi-carré seront développés en détail sur le plan théorique et illustré par des exemples concrets, suivi de l'interprétation des résultats. Enfin, il sera abordé les alternatives au test de Khi-carré, les qualités et les limites de ce test ainsi que les pièges à éviter.

II. Fondements et conditions du test de Khi-carré

Globalement le test de Khi-carré se propose de détecter l'existence des liens statistiquement significatifs entre variables qualitatives. Il vise ainsi à comparer les effectifs observés des effectifs qu'on dû observés (effectifs attendus ou effectifs théoriques) si les variables en présence n'entretenaient aucune relation entre elles. A l'instar d'autres tests statistiques, le test de Khi-carré obéît à une certaine conditionnalité pour valider son utilisation. Au nombre de ces conditions, il y a lieu de citer les suivantes :

1° Le test de Khi-deux porte sur des variables qualitatives ou catégorielles (nominales ou ordinales) ;

2° Chaque donnée ne doit s'insérer que dans une et une seule catégorie (modalité). Ces catégories doivent être mutuellement exclusives ;

3° Les données doivent être indépendantes ;

4° Les données doivent se présenter sous forme des fréquences (absolues ou relatives) ;

5° La fréquence théorique ou attendue de chaque catégorie, notamment dans un tableau de contingence, ne doit pas être inférieure à 5 pour un degré de liberté supérieur ou égal à 2 et inférieure à 10 pour un degré de liberté égal à 1 ;

III. Hypothèses statistiques associées au test de Khi-carré

Tous les tests statistiques sont fondés sur le test de l'hypothèse nulle (H0) que l'on cherche à rejeter. On procède pour cela à un raisonnement par absurde. En effet, ce que l'on cherche à démontrer est l'hypothèse alternative (H1), mais pour y arriver, l'on doit d'abord tester et rejeter l'H0. Ainsi, pour construire le test de l'H1, l'objet de la problématique doit d'abord être écrit. L'H0, l'objet du test, le contraire de l'H1, l'est ensuite (Ghewy, 2010).

La formulation de ces deux hypothèses (H0 et H1) est fonction du type de test de Khi-carré

choisi. Avec le test de Khi-carré d'ajustement, l'H0 postule l'absence de différence de la variable étudiée entre l'échantillon ou la distribution observée et une distribution théorique ou attendue de la même variable. Si la distribution observée est différente de la distribution théorique, on rejette H0 et on conserve H1. Les hypothèses sont ainsi libellées :

H_0 = la distribution observée est identique (n'est pas différente) de la distribution théorique ;

H_1 = la distribution observée est différente de la distribution théorique.

Avec le test de Khi-deux d'indépendance, soient deux variables qualitatives X1 et X2 issues d'un même échantillon, sous l'H0, la distribution de X1 devrait être différente de celle de X2. Si cette différence n'est pas observée, on rejette l'H0 et on conserve l'H1, c'est-à-dire, les variables X1 et X2 sont liées ou dépendantes. Les hypothèses sont ainsi libellées :

H_0 = les variables X1 et X2 sont indépendantes (ne sont pas liées) ;

H_1 = les variables X1 et X2 sont liées (sont dépendantes ou ne sont pas indépendantes).

Avec le test de Khi-carré d'homogénéité, l'H0 postule l'absence de différence de la distribution de la variable entre les différents échantillons issus de la même population ou des populations différentes. Si cette distribution est différente entre les différents échantillons, on rejette l'H0 et on conserve l'H1. Les hypothèses sont ainsi libellées :

H_0 = les distributions observées sont identiques (ne sont pas différentes) entre elles ;

H_1 = les distributions observées ne sont pas identiques (sont différentes) entre elles.

IV. Présentation et interprétation des résultats du test de Khi-carré

4.1. Présentation des résultats

La vérification de l'existence d'une relation entre deux variables catégorielles se réalise à travers le tableau de contingence. C'est une matrice qui croise les modalités de deux variables en présence et comptabilise le nombre d'observations caractérisées simultanément par chacune des combinaisons possibles de couples de modalités de ces deux variables. En dehors des colonnes et des lignes marginales des totaux, un tableau de contingence ou tableau croisé a autant de colonne que la variable placée en colonne a des modalités et autant de lignes que la variable placée en ligne a des modalités. Pour la faciliter la lecture et l'interprétation dans un tableau croisé à deux entrées, on place en ligne la variable supposée explicative et en colonne, la variable supposée expliquée .

Le test de Khi-deux permet de comparer les fréquences observées dans un tableau de contingence et les fréquences théoriques, c'est-à-dire les fréquences qu'on aurait dû observer s'il n'y avait pas de relation entre les deux variables d'étude. Les fréquences théoriques correspondent à l'hypothèse nulle (Ho) des statisticiens. La somme des écarts relatifs entre les fréquences observées et les fréquences théoriques produit une statistique appelée Khi-carré ou Khi-deux, dont la valeur sera d'autant plus élevée que : (Masuy-Stroobant, 2013, p.81)

L'écart entre les fréquences observées et les fréquences théoriques est grand ;

Le nombre d'unités d'analyse (taille de l'échantillon) est grand ;

Le nombre de modalités des variables (nombre des cellules du tableau de contingence) mises en relation est grand. En effet, plus ce nombre est élevé, plus nombreux seront également les écarts à cumuler, ce qui contribue ainsi à augmenter la valeur du Khi-carré.

Dépendant du type d'analyse de Khi-deux réalisé, les résultats peuvent se présenter soit sous la forme d'un tableau de contingence auquel on ajoute la valeur de Khi-deux calculé et sa probabilité associée (principalement pour les cas de Khi-carré d'indépendance ou d'homogénéité) ou plus simplement sous forme d'une ligne de commentaires dans le texte indiquant le Khi-carré et sa probabilité associée (principalement pour le cas du Khi-carré d'ajustement).

4.2. Règle de décision et interprétation des résultats

Une valeur élevée de Khi-carré indique que les fréquences observées s'écartent de l'H0, qui postulait une absence de relation entre les deux variables, mais encore faudra-t-il savoir à partir de quelle valeur et avec quel risque de se tromper.

Pour faciliter la prise de décision, la valeur du Khi-carré calculé sera comparée à la valeur correspondante dans la table de Khi-carré. Cette table constitue une distribution théorique du Khi-carré sous l'H0. Elle comporte une série de valeurs théoriques déclinées selon le degré de liberté du tableau de contingence et la probabilité de réalisation de l'H0. Ainsi, si le Khi-carré calculé est, en fonction du nombre de degré de liberté du tableau de contingence ad hoc (voir Encadré 1), supérieur à la valeur du Khi-carré de la table au niveau de risque spécifié (5 %, quelques fois 10 %) de probabilité que l'H0 soit vraie, rejeter l'H_0. Rejeter l'H0 avec un risque de se tromper de ≤ 5 %, correspond à ne pas rejeter l'H1 au niveau de signification de 5 %. Dans ce cas, on conclut, avec une probabilité ou un niveau de confiance de 95 %, à l'existence d'une relation entre les deux variables ou à la conformité de la distribution observée et la distribution théorique.

Plutôt que de confronter le Khi-carré calculé au Khi-carré de la table, on peut aussi se reporter à la probabilité associée à la distance Khi-carré. Celle-ci s'interprète comme une probabilité d'obtenir un échantillon au moins aussi éloigné de la situation d'indépendance que celui obtenu à travers le calcul (Khi-carré calculé). Si cette probabilité est faible et si les deux variables sont réellement indépendantes, il est peu probable d'obtenir un échantillon au moins aussi éloigné de la situation d'indépendance que celui obtenu empiriquement. Inversement, si la probabilité associée à la distance Khi-carré est élevée et si les deux variables sont réellement indépendantes, il est tout à fait probable d'obtenir un échantillon au moins aussi éloigné de la situation d'indépendance que celui obtenu empiriquement. Dès lors, l'hypothèse d'indépendance (H0) est tout à fait plausible (Martin, 2009).

Concrètement, cela consiste à confronter la probabilité (p) associée à la distance Khi-carré fournie par le logiciel statistique utilisé à la marge d'erreur initialement fixée (a), généralement 5 % ou exceptionnellement 10 %. Si p ≤ a, rejeter l'H0, on conclut que le test est significatif ; Si p > a, l'H_0 ne peut être rejetée, on conclut que le test n'est pas significatif.

Encadré 1 : Degré de liberté

En statistique, le degré de liberté (ddl) désigne le nombre de variables aléatoires qui ne peuvent être déterminées ou fixées par une équation. Il est généralement égal au nombre d'inconnues à estimer (dans une équation par exemple) moins le nombre de relations connues entre ces inconnues.

Par exemple l'équation du type : [X+Y=21] comporte deux inconnues X et Y. Le nombre de relation connue entre ces deux inconnues est de 1, c'est-à-dire la somme de [21]. En faisant 2 (nombre d'inconnues à estimer) moins 1 (nombre de relations connues entre les deux inconnues,) le nombre de degré de liberté est de 1. Intuitivement, cela peut être compris comme suit : Pour autant que la somme de cette équation reste égale à 21, on ne dispose que d'une seule possibilité (degré de liberté) de fixer arbitrairement la valeur d'une des inconnues, la valeur de l'autre inconnue sera automatiquement dérivée de la valeur donnée à l'autre inconnue. Ainsi, si on donne à X la valeur de 4, la valeur de Y ne peut être que de 17. De même, dans l'équation X+Y+Z=35, on dispose de deux degrés de liberté. On peut par exemple fixer arbitrairement les valeurs de X et de Y et la valeur de Z sera dérivée automatiquement des valeurs données aux deux premières inconnues pour autant que la somme totale reste égale à 35. Si on donne à X la valeur de 12, à Y la valeur de 18 et dans ce cas, la valeur de Z ne peut être qu'égale à 5.

Le même principe de raisonnement intuitif peut s'appliquer dans le cas d'un tableau de contingence.

Par exemple dans ce tableau de contingence

Tableau 1 : *Répartition des migrants congolais par sexe selon qu'ils ont ou non envoyé de l'argent à leurs ménages d'origine au moins une fois.*

Envoi des transferts	Sexe du migrant		Total
	Masculin	Féminin	
N'a pas transféré	C1	C2	564
A transféré	C3	C4	744
Total	733	575	1308

Source : *Mangalu (2011) à partir des données de l'enquête MAFE1-RDC, Juillet-août 2007*

Etant donné que les totaux marginaux sont connus, on ne dispose qu'un seul degré de liberté pour remplir ce tableau. Toutes les autres valeurs à compléter dans les autres cellules seront dérivées de cette valeur. Par exemple, si on place 375 dans la C4 (Cellule 4), on ne peut mettre que 220 dans C3, 344 dans C1 et 389 dans C3.

Pour généraliser, dans un tableau de contingence le nombre de degré de liberté est égal à C (nombre de colonnes internes) moins 1 multiplié par L (le nombre de lignes internes) moins 1, soit ddl = (Colonnes-1) (Lignes-1).

V. Typologie et fondements des tests de Khi-carré

Comme déjà indiqué, il existe plusieurs types de test de Khi-carré : le Khi-carré d'ajustement ou de conformité, le Khi-carré d'indépendance ou d'association et le Khi-carré d'homogénéité. Mais ils répondent tous globalement aux mêmes conditions et au même mode de raisonnement.

Les différences apparaissent principalement sur le nombre des populations (échantillons) impliquées, sur le nombre de variables en présence et sur la façon de poser les hypothèses nulle et alternative.

5.1. Fondements du test de Khi-deux d'ajustement ou de conformité

Le test de Khi-deux de conformité vise à déterminer si une distribution observée (série des fréquences absolues ou relatives observées sur une variable) à partir d'un échantillon est significativement différente d'une distribution de fréquences théoriques ou attendues (absolues ou relatives) sur cette même variable ou une distribution connue dans la population sous-jacente. En d'autres termes, il permet de voir si une distribution observée est conforme à la théorie. Deux possibilités s'offrent pour le choix de la distribution théorique : soit on se réfère à une distribution obéissant à la loi d'équiprobabilité, en attribuant à toutes les modalités de la variable la même fréquence absolue ou relative, soit on se réfère à une distribution théorique non-uniforme, en attribuant des fréquences absolues ou relatives différentes aux différentes modalités de la variable.

Dans le premier cas, l'hypothèse nulle testée postule que la distribution observée de l'échantillon n'est pas significativement différente de la distribution théorique de la loi d'équiprobabilité. Il s'agit d'une distribution avec égalité de fréquences théoriques- absolues ou relatives - pour toutes les modalités. C'est-à-dire, on devrait avoir le même nombre de personnes ou la même fréquence dans chaque sous-catégorie (modalité) de la variable. Cette fréquence théorique serait le quotient obtenu par la division de la taille l'échantillon (n), ou de 100 %, par le nombre de modalités de la variable (m). Le calcul de Khi-carré équivaudrait à comparer les fréquences observées et les fréquences théoriques. Si l'écart entre les deux est très important, il est probable que les différentes sous-catégories de la variable ne soient pas identiques, ce qui implique le rejet de l'hypothèse nulle. Cette conclusion sur le rejet de l'hypothèse nulle aurait été difficile à tirer en contemplant simplement les données brutes. D'où la nécessité de réaliser le test de Khi-deux.

Concrètement, les hypothèses statistiques sont formulées comme suit :

H_0 = il n'existe pas de différence statistiquement significative entre la distribution observée et la distribution théorique ;

H_1 = il existe une différence statistiquement significative entre la distribution observée et la distribution attendue.

La procédure de calcul du Khi-carré d'ajustement sous la loi d'équiprobabilité se réalise selon les séquences suivantes :

1° Calcul des fréquences théoriques (ft). Elles sont obtenues en divisant la taille de l'échantillon ou le 100 % par le nombre de modalités de la variable, soit :

$$ft = \frac{n}{m} \; ou \; \frac{100}{m}$$

Où

ft = fréquences théoriques ;

n= la taille de l'échantillon et

m = le nombre de catégories ou de modalités de la variable.

2° Calcul des écarts relatifs entre les fréquences observées (fo) et les fréquences théoriques (ft) pour chaque modalité de la variable et on élève chacun de ces écarts au carré puis on divise chacun de ces écarts élevés au carré par la fréquence théorique correspondance. C'est pour dégager l'importance relative de chaque carré d'écart par rapport l'effectif théorique de sa classe. On utilise pour cela la relation suivante :

$$écarts\ relatifs = \frac{(ft-fo)2}{ft}$$

3° Calcul du $\chi 2$ en sommant tous les écarts relatifs obtenus à l'étape précédente, par la relation suivante :

$$\chi 2 = \sum_{k=1}^{m} \frac{(ft-fo)2}{ft}$$

4° Détermination du nombre de degré de liberté (voir encadré 1). Pour le trouver, soustraire 1 du nombre de modalités de la variable ;

5° En fonction du nombre de degré de liberté et du niveau de confiance fixé (généralement 95 %) ou de la marge d'erreur acceptée (a)

(généralement 5 %), comparez le Khi-carré calculé au Khi-carré théorique ou de la table pour décider ou non du rejet de l'hypothèse nulle.

Dans le second cas, où les fréquences théoriques n'obéissent pas à la loi d'équiprobabilité, l'hypothèse nulle testée postule que la distribution observée de l'échantillon équivaut à la distribution théorique. La distribution théorique correspondant à la distribution de la variable d'étude au sein de la population ou en faisant référence à une théorie préexistante ou à défaut aux résultats des études antérieures. Ici, on ne devrait plus avoir le même nombre de cas ou la même fréquence dans chaque sous-catégorie (modalité) de la variable, certaines modalités étant surreprésentées que d'autres, en référence à ce que l'on sait de la distribution de la variable au sein de la population. Comme dans le cas précédent, le calcul de Khi-carré équivaudrait à comparer les fréquences observées et les fréquences théoriques. Si l'écart entre les deux est très important, il est probable que la distribution observée de la variable au sein de l'échantillon soit loin de correspondre à la distribution théorique de la même variable, ce qui implique le rejet de l'hypothèse nulle.

5.2. Fondements du test de Khi-deux d'association ou d'indépendance

L'analyse du Khi-carré d'association se propose d'étudier les relations entre les variables catégorielles issues d'un même échantillon ou d'une même population, prises deux à deux, en cherchant à examiner si et comment deux variables qualifiant simultanément chaque unité d'analyse varient conjointement. Il s'agira, en d'autres termes, de vérifier l'existence d'une relation d'association ou de dépendance entre les deux variables catégorielles tirées d'un même échantillon ou d'une même population. La vérification d'existence de cette relation d'association ou de cette dépendance se réalise à travers le tableau de contingence ou le tableau croisé. Comme déjà indiqué, pour faciliter la lecture et l'interprétation d'un tableau croisé à deux entrées, on place généralement en ligne la variable supposée explicative et en colonne, la variable supposée expliquée et les proportions ou les pourcentages sont générés dans le sens de la variable supposée indépendante.

Le test de Khi-deux permet ainsi de comparer les fréquences observées dans un tableau de contingence et les fréquences théoriques. Les fréquences observées (fo) sont celles issues de l'observation empirique ou du terrain, c'est la configuration de l'échantillon après la collecte des données. Les fréquences théoriques ou attendues (ft) correspondent aux fréquences qu'on aurait dues observées s'il n'y avait pas de relation ou d'association entre les deux variables. L'hypothèse nulle testée à ce niveau postule l'absence de relation ou d'association entre les deux variables.

Le test de Khi-deux d'association peut s'appliquer sur un tableau 2 X 2, c'est-à-dire, un tableau ayant deux lignes et deux colonnes (sans compter les lignes et colonnes marginales) en croisant deux variables ayant chacune deux modalités ou sur un tableau plus grand (2 X 3, 3 X 2, 3 X 3, etc.) .

La procédure de calcul du Khi-carré d'association se réalise selon les séquences suivantes :

1° Construire le tableau de contingence et calculer par la suite les différentes fréquences théoriques de chaque cellule. Il s'agit en d'autres termes, d'identifier la répartition proportionnelle des différentes modalités de la variable supposée explicative (profil ligne) et imaginer que s'il n'y a pas de lien entre les deux variables, les différentes modalités de la variable supposée expliquée (profil colonne) se répartiront proportionnellement de la même façon que les modalités de la variable explicative. Concrètement, les fréquences théoriques de chaque cellule se calculent selon la relation suivante :

$$ft = \frac{Effectif\ total\ de\ la\ ligne}{Effectif\ total\ de\ l'échantillon} x\ Effectif\ total\ de\ la\ colonne$$

Ces fréquences théoriques permettent de passer du tableau de contingence des données observées au tableau d'indépendance. Le tableau d'indépendance correspond à l'hypothèse nulle (H_0).

2° Calculer les écarts relatifs entre chaque fréquence observée (fo) et la fréquence théorique (ft) pour chaque cellule du tableau de contingence et élever chacun de ces écarts au carré, puis diviser chacun de ces écarts élevés au carré par la fréquence théorique

correspondante. On utilise la relation suivante :

$$écarts\ relatifs = \frac{(ft-fo)2}{ft}$$

3° Calculer le $\chi 2$ en additionnant tous les écarts relatifs obtenus à l'étape précédente, par la relation suivante :

$$\chi 2 = \sum_{n=1}^{k} \frac{(ft-fo)2}{ft}$$

4° Confronter la valeur de Khi-carré calculé à la valeur de référence sur la table de distribution théorique de Khi-carré.

Pour cette fin, on calcule le nombre de degré de liberté (ddl) et on fixe le seuil de signification acceptable. Dans le cas de tableaux de contingence, le nombre de degré de liberté est déterminé par la relation suivante :

$$ddl=(Colonnes-1)x\ (Lignes-1)$$

Le seuil de signification est généralement fixé à 5 %. Ces deux statistiques permettent d'entrer dans la table de la distribution théorique indiquant des valeurs critiques de Khi-carré au-delà desquelles on peut rejeter l'hypothèse nulle.

5.3. Fondements du test de Khi-deux de comparaison ou d'homogénéité

Le test de Khi-carré d'homogénéité est utilisé pour comparer la distribution d'une variable catégorielle à p modalités entre k échantillons de tailles n1, n2, … nk. Il s'agit de comparer deux ou plusieurs distributions observées sur des échantillons issus d'une même ou de plusieurs populations différentes. Les observations sont regroupées dans un tableau de contingence présentant autant de colonnes que d'échantillons observés (k colonnes). L'hypothèse nulle postule que les distributions observées sont identiques entre les échantillons observés. Comme dans les cas de Khi-deux d'association, il faut commencer par calculer les fréquences théoriques sous l'hypothèse nulle. Les principes, les procédures de calcul et les règles de décision seront quasiment identiques à ceux développés dans le cas de Khi-carré d'indépendance.

Les données utilisées dans ce chapitre pour illustrer ces différents tests de Khi-deux sont principalement issues d'une enquête réalisée à Kinshasa en juillet-aout 2007 dans le cadre du projet de recherche intitulé « Crise économique et migrations internationales en République Démocratique du Congo ». Ce projet s'inscrivait dans le cadre d'un vaste programme de recherche international dénommé FSP 2003-74 « Migrations internationales, recompositions territoriales et développement dans les pays du Sud ». Outre la R.D.C, 9 autres pays étaient impliqués dans ce projet.

Les données ont été collectées à la fois au niveau des ménages (ayant ou non des migrants) et des membres de ces ménages (migrants et non-migrants). Au niveau des ménages, les données ont été collectées sur la base d'un questionnaire ménage. Outre l'identification de tous les membres des ménages, présents et absents, le questionnaire ménage se proposait également de collecter les données sur les migrants du ménage et sur les frères et/ou sœurs des chefs de ménage et/ou de leurs conjoints qui habitent à l'étranger pour au moins 3 mois. En somme, l'enquête devrait permettre entre autres de répondre aux questions suivantes : Où vont les migrants congolais ? Les migrants congolais reçoivent-ils des aides de leurs ménages d'origine pour faciliter leurs migrations ? Quel est le profil des migrants les plus susceptibles de recevoir ces aides ? Les migrants congolais rapatrient-ils de l'argent et autres biens à leurs ménages d'origine ? Quelles sont les caractéristiques des migrants qui rapatrient de l'argent à leurs ménages d'origine ? Est-ce que la réception des transferts et le montant reçu agissent-ils sur les conditions de vie des ménages ? etc.

Au total, 945 ménages et 992 membres des ménages ont été enquêtés. Dans le cadre de cet article, seules les données ménages collectées à Kinshasa ont été exploitées. Sur les 945 ménages enquêtés, 478 ménages, soit environ 51 %, étaient directement concernés par la migration, avec au moins un membre ayant vécu ou vivant encore à l'étranger. Sur les 1 650 personnes ayant une expérience migratoire identifiées au sein des ménages enquêtés, 1 339 vivaient encore à l'étranger au moment de l'enquête.

Mais dans le cadre de cet article, nous répondrons principalement aux deux questions suivantes : Où vont les migrants congolais ? Quelles sont les caractéristiques des migrants qui rapatrient de l'argent et autres biens à leurs ménages

d'origine ? Pour la première question, il s'agira d'identifier le continent le plus susceptible d'être choisi par les migrants congolais lors de leurs premières migrations. Dans la deuxième, il s'agira d'étudier la probabilité pour les migrants congolais de rapatrier de l'argent et autres biens à leurs ménages d'origine en fonction de leurs caractéristiques socio-démographiques, mais pour le besoin de la cause, nous allons nous limiter seulement au sexe des migrants.

5.4. Application et illustration des tests de Khi-carré

5.4.1. Application du test de Khi-carré de conformité ou d'ajustement

Avec la première question, « Où vont les migrants congolais ? » et étant donné qu'il n'y a qu'une seule variable « Continent choisi lors de la première migration », le test indiqué est celui de Khi-deux de conformité ou d'ajustement, il s'agit de voir si une série d'effectifs observés diffère significativement d'une série d'effectifs théoriques ou attendus. Dans le cas d'espèce, il s'agit de voir si un continent donné avait plus de chance d'être choisi par les migrants congolais lors de leurs premières migrations. Si ce n'est pas le cas, on s'attendrait à ce que les proportions des migrants congolais qui choisissent l'un ou l'autre continent soient identiques ou à peu près identiques. La situation observée est reprise dans le Tableau 1 qui suit :

Tableau 1 : Répartition des migrants congolais selon les continents de la première migration

Continents	Effectif	Pourcentage
Afrique	1 011	61,7
Europe	552	33,6
Amérique	66	4,0
Asie	10	0,6
Total	1 639	100,0
Missing = 11		

Du Tableau 1, on peut déduire que la majorité des Congolais (près de 62 %) choisissent le continent africain lors de leurs premières migrations, près de 34 % choisissent l'Europe et les autres continents sont très marginaux. Signalons aussi que pour 11 migrants, le continent de la première migration n'a été indiqué, ce qui correspond aux données manquantes ou missing values. L'une des premières choses à faire est de regrouper les modalités des continents très faiblement représentées et de décider sur le sort des données manquantes . Cela donne le Tableau 1 bis suivant :

Tableau 1 : Répartition des migrants congolais selon les continents de la première migration

Continents	Effectif	Pourcentage
Afrique	1 011	61,7
Europe	552	33,6
Amérique	66	4,0
Asie	10	0,6
Total	1 639	100,0
Missing = 11		

On pourrait, à partir du Tableau 1 bis, répondre à la question principale en notant que c'est l'Afrique qui est le continent le plus susceptible d'être choisi par les migrants congolais lors de leurs premières migrations. Ce raisonnement ne tient pas à première vue dans la mesure où l'échantillon analysé ici n'est que l'un des échantillons possibles qu'on aurait dus tirés à partir de la population de la ville de Kinshasa et qu'il n'est pas exclu que les résultats observés ici ne soient vrais qu'à l'intérieur de l'échantillon analysé et non dans d'autres échantillons possibles, c'est ce qu'on qualifie d'erreur d'échantillon. Pour s'assurer que pareil résultat a également de chance d'être observé dans la population de référence, un test statistique est recommandé. Dans le cas d'espèce, c'est le test de Khi-carré d'ajustement qui est le plus indiqué.

La marche à suivre pour réaliser le test se décline en 6 principales étapes : vérification des conditions d'application de la méthode, la pose des hypothèses statistiques, le calcul de Khi-carré, le repérage du Khi-carré de la table, la confrontation du Khi-carré calculé et du Khi-carré de la table et enfin la conclusion à tirer.

1° Vérifier les conditions d'application du type de test choisi

Après avoir arrangé les données et régler les problèmes de données manquantes et données aberrantes éventuelles et décider du test à appliquer, il faille vérifier les conditions d'application du test choisi . Dans le cas d'espèce, il a été opté pour le test de Khi-deux d'ajustement ou de conformité étant donné que la distribution à analyser porte sur une seule variable, avec

trois modalités, dont on veut comparer la conformité ou l'ajustement par rapport à une distribution théorique.

2° Poser les hypothèses statistiques

En fonction des objectifs de la recherche, poser ou convertir les hypothèses de recherche en hypothèses statistiques à tester.

H0 = les migrants congolais choisissent de manière indifférenciée les continents de destination lors de leurs premières migrations. En d'autres termes, il y a une équiprobabilité dans le choix de continent de destination de la part les migrants congolais lors de leurs premières migrations ;

H1 = les migrants congolais ont une préférence pour un continent de destination quelconque lors de leurs premières migrations.

3° Calculer le Khi-carré sous l'hypothèse nulle

Si l'hypothèse nulle était vraie, on aura la même proportion des Congolais qui choisissent chacun des continents lors de leurs premières migrations. Se référant au résultat présenté dans le Tableau 1 bis, l'H0 correspondrait à diviser l'effectif total (1 639) ou la proportion totale (100 %) par le nombre de modalités de la variable (3), soit :

1639/3=546,3 , qui correspond à l'effet théorique ou attendu pour chaque modalité.

Ainsi, si le choix du continent de la première migration n'avait aucune préférence de la part des migrants, chacun de trois continents aurait été choisi par environ 546 migrants congolais, or on observe que ceci n'est pas le cas, certains continents ont été plus choisis que d'autres.

Il reste maintenant à calculer le Khi-carré et à le confronter au Khi-carré théorique.

Tableau 2 : Calcul des écarts absolus, des écarts relatifs et du Khi-carré

Continents	Effectif observé (fo)	Effectif théorique (ft)	Ecarts (ft-fo)	Ecarts au carré $(ft-fo)^2$	$(ft-fo)^2/ft$
Afrique	1 011	546,3	-464,7	215 946,1	395,3
Europe	552	546,3	-5,7	32,5	0,06
Autres continents	76	546,3	470,3	221 182,1	404,9
Total	1 639	1 639			800,2

En appliquant la formule, c'est-à-dire, en additionnant la dernière colonne, on trouve un X^2 = 800,2 à comparer avec le X^2 de la table.

4° Trouver le Khi-carré de la table

Pour trouver le X^2 de la table, commencer par fixer le seuil de signification et le nombre de degré de liberté. On fixe le seuil de signification à 95 %, avec une marge d'erreur de 5 %. Le nombre de degré de liberté correspond à 2 dans le cas d'espèce, soit nombre de modalités (3) moins 1. Avec ces deux entrées, le Khi-carré, tiré de la table de distribution théorique de Khi-carré est de 5,99.

5° Comparer le Khi-carré de la table au Khi-carré calculer et décider

Etant donné que le X^2 calculé est supérieur au X^2 de la table, on le rejette l'H0, ce qui implique le non-rejet de l'H1. Ainsi donc, on peut conclure que les migrants congolais choisissent de manière préférentielle l'Afrique comme le continent de leur première migration. On a plus de 95 % de chance que ce résultat tiré de l'échantillon soit également vrai pour l'ensemble de la population de la ville de Kinshasa.

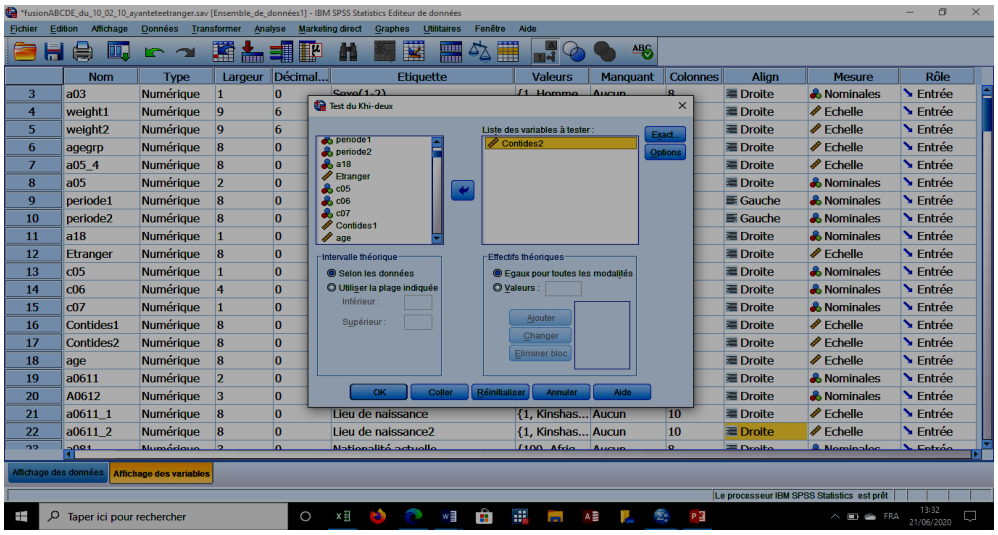
Par ailleurs, plutôt que de procéder à ces différents calculs manuellement, on peut recourir à un logiciel statistique spécialisé. Dans le cas d'espèce, on recourt à SPSS. Les données doivent être préalablement chargées sur SPSS . Procéder alors comme suit :

1° Cliquer sur Analyse Tests non paramétriques Boîte de dialogue valeurs anciennes Khi-carré

La boîte de dialogue ci-dessous s'affiche :

Figure 1 :

Test de Khi-carré d'ajustement sous SPSS



2° Placer la variable à analyser dans la zone « Liste des variables à tester », puis cliquer sur « Asymptotique » uniquement et cochez sur Exact, puis sur « Poursuivre » et enfin sur « OK » pour afficher le résultat. Le résultat SPSS se présente comme suit :

Tableau 3 :

Continent de première destination2

	Effectif observé	Effectif théorique	Résidu
Afrique	1011	546,3	464,7
Europe	552	546,3	5,7
Autres continents	76	546,3	-470,3
Total	1639		

Tableau 3 bis :

Tableau 3 bis :

Test

	Continent de première destination2
Khi-deux	800,172[a]
ddl	2
Signification asymptotique	,000

a. 0 cellules (0,0%) ont des fréquences théoriques inférieures à 5. La fréquence théorique minimum d'une cellule est 546,3.

p<0,0001 indique, si l'Ho est vérifiée, il y a moins d'une chance du 10 000 pour que le X^2 soit aussi grand que celui que nous avons trouvé. Autrement dit, c'est très improbable..

Comme il a été déjà indiqué, il est également possible de confronter les fréquences observées aux fréquences théoriques qui ne sont pas toutes égales. Le raisonnement reste parfaitement identique. Ce cas sera illustré par cet exemple-ci : Admettons qu'on ait observé au cours de l'année 2019, 15 004 naissances vivantes au sein d'une population donnée, dont 7 451 naissances masculines. La question que l'on voudrait vérifier ici est celle de savoir si la répartition de naissances selon le sexe dans ce pays au cours de l'année 2019 est conforme à la théorie démographique en la matière ?

La situation observée est présentée dans le Tableau 4 ci-dessous :

Tableau 4 : Répartition des naissances vivantes observées au cours de l'année 2019 selon le sexe

Tableau 4 : Répartition des naissances vivantes observées au cours de l'année 2019 selon le sexe

Naissances vivantes	Effectif	Pourcentage
Masculin	7 451	49,7
Féminin	7 553	50,3
Total	15 004	100,0

Il s'agit ici d'un cas où les effectifs attendus ne doivent pas être identiques ou équiprobables. On pose alors les hypothèses statistiques suivantes :

H0 = La distribution des naissances par sexe observées équivaut à la distribution théorique des naissances selon le sexe ;

H1 = La distribution des naissances par sexe observées est différente de la distribution théorique.

On procède à toutes les étapes développées ci-avant, notamment le calcul des effectifs théoriques sous l'hypothèse nulle. En effet, on sait à partir de la théorie démographique qu'il nait généralement environ 1,05 garçon pour 1 fille. C'est donc à partir de cette relation théorique que les effectifs théoriques ou attendus doivent être calculés.

Tableau 4 bis : Calcul des écarts absolus, des écarts relatifs et du Khi-carré

Tableau 4 bis : Calcul des écarts absolus, des écarts relatifs et du Khi-carré

Naissances vivantes	Effectif observé (fo)	Effectif théorique (ft)	Ecarts (ft-fo)	Ecarts au carré $(ft-fo)^2$	$(ft-fo)^2/ft$
Masculin	7 451	7 685	234	54 756	7,1250
Féminin	7 553	7 319	-234	54 756	7,4813
Total	15 004	15 004			14,6063

Comme la variable ne comporte que deux modalités, le nombre de degrés de liberté (ddl) ne peut être qu`égal à 1. Avec un ddl de 1 et une marge d'erreur de 5 %, le Khi-carré de la table équivaut à 3,84. Etant donné que le Khi-carré calculé (14,61) est supérieur au Khi-carré de la table (3,84), l'H0 est rejetée. En conclusion, la distribution des naissances par sexe observée en 2019 est statiquement différente de la distribution théorique des naissances par sexe.

Avec SPSS, on obtient les résultats suivants, qui restent d'ailleurs conformes à ceux obtenus procédant aux calculs manuels. Le résultat SPSS s'affiche dans le Tableau 5 ci-dessous :

Test du Khi-deux

Tableau : 5

Tableau : 5

Effectif

	Effectif observé	Effectif théorique	Résidu
Sexe Masculin	7451	7685,0	-234,0
Sexe féminin	7553	7319,0	234,0
Total	15004		

Tableau 5 bis :

Test

	Effectif
Khi-deux	14,606[a]
ddl	1
Signification asymptotique	,000

a. 0 cellules (0,0%) ont des fréquences théoriques inférieures à 5. La fréquence théorique minimum d'une cellule est 7319,0.

5.4.2. Application du test de Khi-deux d'association ou d'indépendance

Le test de Khi-carré d'association ou d'indépendance permet de vérifier l'existence d'une relation d'association ou de dépendance entre deux variables qualitatives ou catégorielles (nominales ou ordinales) issues d'une même population ou d'un même échantillon. Dans le cas d'espèce, on cherche à savoir s'il existerait un lien entre le sexe de migrants congolais résidant à l'étranger et leur probabilité de transférer de l'argent et des biens à leurs ménages d'origines. Rappelons que la question sur les transferts n'a été posée que pour les migrants résidant à l'étranger au moment de l'enquête et dont l'effectif s'élevait à 1 339. Les résultats observés à partir de l'enquête pour chacune des variables (sexe et probabilité de transférer) sont consignés dans les Tableaux 6 et 7 ci-dessous.

Tableau 6 : Répartition des migrants congolais résidant à l'étranger par sexe

Sexe	Effectifs	Pourcentages
Masculin	747	55,8
Féminin	591	44,2
Total	1338	100,0
Missing : 1		

Tableau 7 : Répartition des migrants congolais résidant à l'étranger qu'ils ont ou non transféré

Transfert	Effectifs	Pourcentages
N'a pas transféré	564	43,1
A transféré	744	56,9
Total	1308	100,0
Missing : 31		

Le Tableau 6 indique que près de 56 % de migrants congolais sont de sexe masculin. Parmi ces migrantes et migrants, près de 57 % ont rapatrié de l'argent ou d'autres biens matériels à leurs ménages d'origine au moins une fois depuis qu'ils sont à l'étranger (Tableau 7). Pour calculer le Khi-carré, il faut préalablement transformer les tableaux univariés 6 et 7 en tableau de contingence ou tableau croisé combinant les modalités de ces deux variables (Sexe et probabilité de transférer). Cette transformation donne le résultat repris dans le Tableau 8 ci-dessous :

Tableau 8 : Répartition des migrants par sexe selon qu'ils ont ou non-transféré de l'argent et autres biens à leurs ménages d'origine

Tableau 8 : Répartition des migrants par sexe selon qu'ils ont ou non-transféré de l'argent et autres biens à leurs ménages d'origine

Sexe	Transfert		Total
	N'a pas transféré	A transféré	
Masculin	344	389	733
Féminin	220	355	575
Total	564	744	1308

A partir de ce Tableau 8, procéder au calcul de Khi-carré selon la démarche suivante :

1° Vérifier les conditions d'application du test choisi ;

Après avoir arrangé les données et régler les problèmes liés aux données manquantes et aux données aberrantes éventuelles et décider du test à appliquer, il faille vérifier les conditions d'application du test choisi. Dans le cas d'espèce, étant donné qu'il s'agit des variables catégorielles ayant chacune deux modalités et que l'analyse consiste à vérifier les liens éventuels entre ces deux variables, le test de Khi-deux d'indépendance ou d'association est plus indiqué. Par ailleurs, le nombre d'observations et de cas dans chaque cellule du tableau est suffisant. Toutes les conditions d'application de la technique sont observées.

2° Poser les hypothèses statistiques

H0 = Les femmes transfèreraient autant que les hommes. En d'autres termes, la probabilité de transférer serait identique entre les femmes et les hommes ;

H1 = La probabilité de transférer des femmes serait différente de celle des hommes.

3° Calculer le Khi-carré sous l'hypothèse nulle

Le calcul du Khi-carré commence par le calcul des effectifs théoriques. Ils s'obtiennent selon la relation suivante :

$$ft = \frac{Effectif\ total\ de\ la\ ligne}{Effectif\ total\ de\ l'échantillon} x\ Effectif\ total\ de\ la\ colonne$$

Le premier effectif théorique (les hommes qui n'ont pas transféré) serait égal à : $ft_1 = \frac{733}{1308} * 564$: 316,1 ; le deuxième effectif théorique (les femmes qui n'ont pas transféré) correspond à : $ft_2 = \frac{733}{1308} * 564$: 247,9 ; le troisième effectif théorique (les hommes qui ont transféré) correspond à : $ft_3 = \frac{733}{1308} * 744$: 416,9 et le quatrième effectif théorique (les femmes qui ont transféré) correspond à : $ft_4 = \frac{575}{1308} * 744$: 327,1. Ceci produit le tableau que voici :

Tableau 9 : Distribution d'indépendance ou effectifs théoriques

Tableau 9 : Distribution d'indépendance ou effectifs théoriques

Sexe	Transfert		Total
	N'a pas transféré	A transféré	
Masculin	316,1	416,9	733
Féminin	247,9	327,1	575
Total	564,0	744,0	1308

A partir des effectifs théoriques, on calcule le Khi-carré.

Tableau 10 : Calcul des écarts absolus, des écarts relatifs et du Khi-carré

Tableau 10 : Calcul des écarts absolus, des écarts relatifs et du Khi-carré

Continents	fo	ft	(ft-fo)	$(ft-fo)^2$	$(ft-fo)^2/ft$
Hommes n'ayant pas transféré	344	316,1	-27,9	778,41	2,46254350
Femmes n'ayant pas transféré	220	247,9	27,9	778,41	3,14001614
Hommes ayant transféré	389	416,9	27,9	778,41	1,86713840
Femmes ayant transféré	355	327,1	-27,9	778,41	2,37973097
Total	1308	1308	-		9,84942901

Khi^2

La lecture que l'on peut faire des fréquences théoriques est la suivante : s'il n'y avait pas de lien entre le sexe et la probabilité de transférer (l'H0 non rejetée), on trouverait seulement 316 hommes qui n'ont pas transféré (mais on en observe plus, soit 344) et on trouverait 417 hommes qui ont transféré (mais on en observe moins, soit 389). De même, s'il n'y avait pas de lien entre les deux variables, on trouverait 248 femmes qui n'ont pas transféré (mais on en observe moins, soit 220) et on trouverait 327 femmes qui ont transféré (mais on en observe plus, soit 355). Ces chiffres suggèrent déjà l'existence d'un lien entre le sexe des migrants et leurs probabilités de transférer, mais seule la comparaison du Khi-carré calculé et celui de la table permettra de tirer des conclusions solides. En appliquant la formule, on trouve un X^2 = 9,84942901 que l'on doit maintenant comparer avec le X^2 de la table.

4° Trouver le Khi-carré de la table

Pour trouver le X^2 de la table, il faut fixer le seuil de signification et calculer le nombre de degré de liberté. On fixe le seuil de signification à 95 %, avec une marge d'erreur de 5 %. Le nombre de degré de liberté correspond à 1, soit ddl = (C-1) x (L-1) : (2-1) x (2-1) =1. Avec ces deux entrées, le Khi-carré de la table est égal à 3,84.

5° Comparer le Khi-carré de la table au Khi-carré calculé et décider

Le X^2 calculé (9,85) est supérieur au X^2 de la table (3,84), alors on rejette l'H0. Ainsi donc, on peut conclure, avec une marge d'erreur de

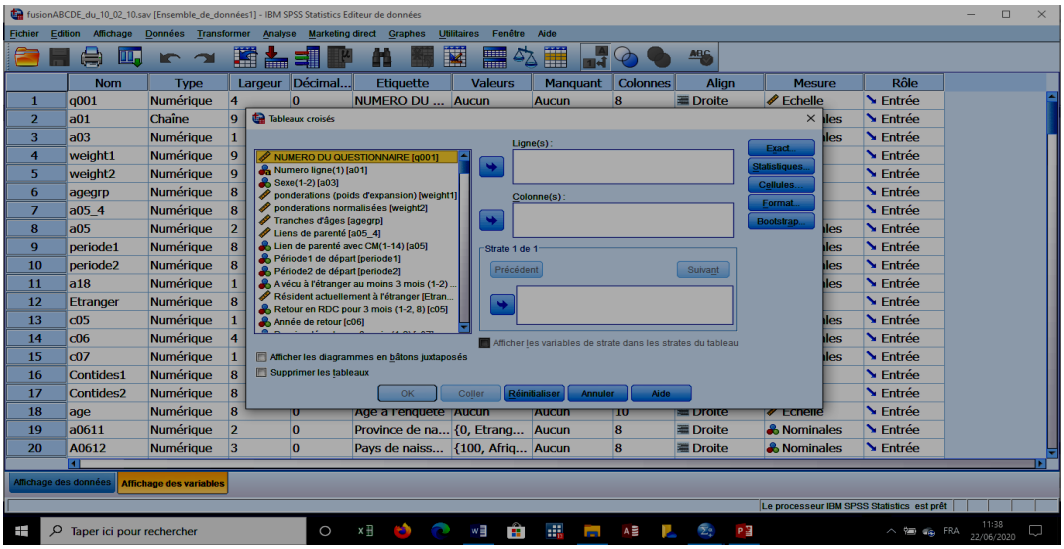
5 %, à l'existence d'un lien statistiquement significatif entre le sexe des migrants et leurs probabilités de transférer. Toutefois, le test de Khi-deux en lui-même ne permet pas de dire qui des femmes ou des hommes ont une probabilité plus élevée de transférer. Pour répondre à cette question, il faut se reporter au tableau de contingence, qui constitue l'un des résultats du test de Khi-carré. Confère Tableau 12 pour des commentaires appropriés. A noter que même si l'exemple utilisé ici a porté sur un tableau 2 X 2, il est également possible de traiter, selon la même logique, des tableaux plus grands, donc des variables avec plusieurs modalités (par exemple 6 X 4 ou 7 X 3). Toutefois, les interprétations deviennent plus délicates dans ce dernier cas.

Par ailleurs, plutôt que de calculer le Khi-deux à la main, on peut recourir à l'ordinateur et à certains logiciels statistiques, notamment SPSS, pour procéder au calcul de Khi-carré. Sous SPSS, procéder comme suit :

1° Cliquer sur : Analyse Statistiques descriptives Tableaux croisés

La boîte de dialogue ci-dessous s'affiche :

Figure 2 : Test de Khi-carré d'association sous SPSS



2° Déplacer la variable dépendante (Transfert) dans la zone Colonne et la variable indépendante (Sexe) dans la zone Ligne ;

3° Cliquer sur le bouton Statistiques et cocher sur Chi-deux, il est possible de demander d'autres statistiques comme le coefficient de contingence, le Phi et V de Cramer, Lambda, le coefficient d'incertitude, etc. et sur Poursuivre dès que vous avez fini ;

4° Cliquer sur Cellule. L'option Observé dans le groupe Effectif est déjà sélectionnée par défaut, mais il est aussi possible de demander les effectifs attendus en sélectionnant le bouton correspondant. N'oublier pas de demander la production de pourcentage dans le même sens que la variable supposée explicative. Dans le cas d'espèce, il s'agit de pourcentage en ligne et cliquer sur Poursuivre dès que vous avez fini ;

5° Cliquez sur OK pour afficher les résultats.

SPSS, comme d'autres applications statistiques, produit beaucoup des résultats dont certains ne seront même pas utilisés. Parmi les résultats produits par SPSS, il y a notamment ce qui suit :

Tableau 11 : Récapitulatif du traitement des observations

	Observations							
	Valide		Manquante		Total			
	N	Pourcent	N	Pourcent	N	Pourcent		
Sexe (1-2) * Envoi des transferts	1308	97,7%	31	2,3%	1339	100,0%		

Tableau 11 bis : Tableau croisé Sexe * Envoi des transferts

			N'a pas transféré	A transféré	Total
			Envoi des transferts		
Sexe (1-2)	Homme	Effectif	344	389	733
		% compris dans Sexe	46,9%	53,1%	100,0%
	Femme	Effectif	220	355	575
		% compris dans Sexe	38,3%	61,7%	100,0%
Total		Effectif	564	744	1308
		% compris dans Sexe	43,1%	56,9%	100,0%

Tableau 11 ter : Test du Khi-deux

	Valeur	ddl	Signification asymptotique (bilatérale)	Signification exacte (bilatérale)	Signification exacte (unilatérale)
Khi-deux de Pearson	**9,875[a]**	**1**	**,002**		
Correction pour la continuité[b]	9,524	1	,002		
Rapport de vraisemblance	9,908	1	,002		
Test exact de Fisher				,002	,001
Association linéaire par linéaire	9,867	1	,002		
Nombre d'observations valides	1308				

a. 0 cellules (0,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 247,94.

b. Calculé uniquement pour un tableau 2x2.

6° Commenter les résultats et décider

La première chose à regarder dans ce résultat est le dernier tableau intitulé Test de Khi-deux (Tableau 11 ter). Ce tableau présente notamment la valeur de Khi-carré et la probabilité associée ainsi que les autres statistiques. C'est à partir de ce tableau que l'on répond à la question de départ sur le lien éventuel entre les deux variables. Le Khi-carré calculé est de 9,875, avec une probabilité associée de 0,002 et un degré de liberté de 1. Etant donné que la probabilité associée (p) de 0,002 est inférieure à la marge d'erreur (a) de 0,05 (5 %), on rejette l'H0.

Le deuxième tableau, intitulé Tableau croisé Sexe * Envoi des transferts (Tableau 11 bis) complète les commentaires ci-dessus, notamment en identifiant entre les catégories (Masculin et Féminin) de la variable supposée « explicative » (Sexe), laquelle a une probabilité plus élevée d'envoyer de l'argent et d'autres biens aux ménages d'origine . Ainsi, en examinant le Tableau 10 bis, on note que sur les 100 % d'hommes, seuls 53,1 % ont envoyé des transferts aux ménages d'origine alors que cette proportion est de près de 62 % chez les femmes. On conclut donc que les femmes ont une probabilité plus élevée de rapatrier les transferts à leurs ménages d'origine que les hommes au seuil de 95 %. En d'autres termes, on est à plus de 95 % sûr que les résultats observés sur cet échantillon puissent se reproduire au sein de la population de référence.

Le premier tableau, intitulé Tableau récapitulatif de traitement des observations (Tableau 11) permet juste de se faire une idée sur le nombre d'observations impliquées dans l'analyse ainsi que sur le nombre d'éventuelles observations exclues. Dans un tableau croisé, une observation n'est prise en compte dans l'analyse que si et seulement si elle dispose des données à la fois sur les deux variables impliquées. Si elle manque des données sur l'une ou l'autre de deux variables ou sur les deux à la fois, elle est automatiquement exclue de l'analyse. Ainsi, sur les 1 339 migrants vivant à l'étranger identifiés au sein des ménages enquêtés, seuls 1 308 ont répondu simultanément aux deux questions, les 31 autres n'ont pas de données sur l'une ou sur les deux variables à la fois. L'examen des tableaux univariés permet de se faire une idée exacte sur ces missings.

Toutefois, ce ne sont pas tous ces tableaux produits par le logiciel utilisé qui sont présentés dans un rapport de recherche. Seuls les tableaux portant des renseignements utiles et réarrangés par l'analyste sont à présenter. Dans le cas d'espèce, seul le tableau croisé, auquel on ajoute des indications sur le test de Khi-carré et la significativité de ce test, est présenté dans un rapport. Concernant la significativité, plutôt que de présenter des chiffres, il est usuel de les présenter sous forme d'étoile. Une convention veut que l'on attribue trois étoiles (***) aux résultats avec une probabilité inférieure à 1 % (0,01), deux étoiles (**) aux résultats ayant une probabilité inférieure à 5 % (0,05) et une étoile (*) aux résultats ayant une probabilité inférieure à 10 % (0,1). Généralement, tous les résultats ayant une probabilité supérieure à 5 % sont jugés non-significatifs. Toutefois, on peut, pour des phénomènes rares, admettre jusqu'à 10 %.

Tableau 12 : Répartition des migrants par sexe selon qu'ils ont ou non-transféré de l'argent et autres biens à leurs ménages d'origine

Sexe	Transfert				Total
	N'a pas transféré		A transféré		
	Eff.	%	Eff.	%	
Masculin	344	46,9	389	53,1	733
Féminin	220	38,3	355	61,7	575
Total	564	43,1	744	56,9	1308
Khi-carré = 9,875***					

Il est également possible de ne présenter dans le tableau que des effectifs relatifs, la taille de l'échantillon étant présentée dans la même ligne que le Khi-carré et sa probabilité associée.

5.4.3. Application du test de Khi-carré d'homogénéité

Le test de Khi-carré d'homogénéité est utilisé pour comparer la distribution d'une variable qualitative à p modalités entre k échantillons de tailles n1, n2, … nk. Comme dans les cas de Khi-deux d'association, il faut commencer par vérifier les conditions d'application de la technique et par calculer les fréquences théoriques sous l'hypothèse nulle. Les principes, les procédures de calcul et les règles de décision seront développés dans les lignes qui suivent. Par exemple, si l'on cherche à savoir si la structure par âge de la population féminine congolaise en âge de procréer est restée la même ou a évolué, d'une part, entre 2001 et

2010 et, d'autre part, entre 2001 et 2018 ; on peut utiliser les données des enquêtes MICS 2001, MICS2010 et MICS2018. Le Tableau 13 présente la répartition de la population féminine de la R.D.C en âge de procréer par grands groupes d'âges tirée de chacune de ces enquêtes.

Tableau 13 : Proportion de la population féminine congolaise en âge de procréer par grands groupes d'âge de 2001 à 2018.

Groupe d'âge	MICS2001		MICS2010		MICS 2018	
	Effectif	%	Effectif	%	Effectif	%
15-19 ans	2 933	23,6	2 732	21,3	5 076	23,3
20-34 ans	6 075	49,0	6 681	52,0	10 861	49,9
35-49 ans	3 401	27,4	3 437	26,7	5 819	26,7
Total	12 409	100,0	12 850	100,0	21 756	100,0

La première question à laquelle on tentera de répondre est celle-ci : la structure par âge de la population congolaise en âge de procréer de 2001 (Pf15-49_2001) est-elle identique à celle de 2010 (Pf15-49_2010) ? Avec cette question, il s'agirait de procéder à une comparaison portant sur une même variable (proportion par âge de la population féminine en âge de procréer) sur deux échantillons indépendants (Pf15-49_2001 et Pf15-49_2010). L'analyse de Khi-carré de d'homogénéité répond à ces types de questions.

A l'instar des autres types de tests de Khi-carré développés ci-dessus, la démarche à suivre est la suivante :

1° Vérifier les conditions d'application du test choisi

Après avoir arrangé les données et régler les problèmes liés aux données manquantes et aux données aberrantes éventuelles et décider du test à appliquer, il faille vérifier que les conditions d'application du test de Khi-carré d'homogénéité sont remplies. Etant donné qu'il s'agit ici de comparer les proportions issues de deux populations (échantillons) différentes, le test de Khi-deux d'homogénéité est le plus indiqué. On a également observé que le nombre de cas est suffisamment important pour permettre une analyse solide.

2° Poser les hypothèses statistiques

Pour ce qui est du test de Khi-carré d'homogénéité, les hypothèses statistiques peuvent se libeller comme suit :

H0 = La structure par âge de la population féminine congolaise en âge de procréer de 2001 est identique à celle de 2010, soit Pf15-49_2001 = Pf15-49_2010.

H1 = La structure par âge de la population féminine congolaise en âge de procréer de 2001 est différente de celle de 2010, soit Pf15-49_2001 ≠ Pf15-49_2010.

3° Calculer les effectifs ou les fréquences théoriques sous l'hypothèse nulle

Pour rappel, les effectifs ou fréquences théoriques ou attendus correspondent aux effectifs ou fréquences qu'on observerait si la structure de la population féminine congolaise en âge de procréer de 2001 était identique à celle de 2010 (H0). Elles se calculent comme indiqué précédemment. Ces fréquences sont présentées dans le Tableau 15 ci-dessous :

Groupe d'âge	MICS2001	MICS2010	Total
15-19 ans	2 782,05	2 881,95	5 665
20-34 ans	6 266,64	6 489,35	12 756
35-39 ans	3 359,31	3 478,69	6 638
Total	12 409	12 850	25 259

4° Calculer le Khi-carré sous l'hypothèse nulle

L'utilisation de la formule de Khi-carré implique qu'on calcule préalablement les écarts absolus par rapport aux effectifs observés, à élever ces écarts au carré et à calculer par la suite les écarts relatifs en divisant chaque écart absolu élevé au carré à l'effectif théorique correspondant. Le Khi-carré correspond alors à la somme de tous ces écarts relatifs.

Tableau 16 : Les écarts absolus par rapport aux effectifs observés (ft-fo)

Groupe d'âge	MICS2001	MICS2010
15-19 ans	-149,9530	149,9530
20-34 ans	191,6457	-191,6457
35-39 ans	-41,6927	41,6927

Tableau 17 : Les écarts absolus par rapport aux effectifs élevés au carré $(ft-fo)^2$

Groupe d'âge	MICS2001	MICS2010
15-19 ans	22 485,8924	22 485,8924
20-34 ans	36 728,0783	36 728,0783
35-49 ans	1 738,2848	1 738,2848

Tableau 18 : Les écarts relatifs $((ft-fo)^2/ft)$

Groupe d'âge	MICS2001	MICS2001	Total
15-19 ans	8,0796	7,8023	15,8819
20-34 ans	5,8609	5,6597	11,5206
35-39 ans	0,5175	0,4997	1,01715
Total	14,5480	13,9617	28,4197

5° Trouver le Khi-carré de la table

Pour trouver le $^X2$ de la table, fixer le seuil de signification et calculer le nombre de degré de liberté. On fixe le seuil de signification à 95 %, avec une marge d'erreur de 5 %. Le nombre de degré de liberté vaut à 2, soit ddl = (C-1) x (L-1) : (2-1) x (3-1) =2. Avec ces deux entrées, le Khi-carré de la table est égal à 5,99.

6° Comparer le Khi-carré de la table au Khi-carré calculé et décider

Le X^2 calculé (28,42) est supérieur au $^X2$ de la table (5,99), on rejette alors l'H0 et on accepte l'H1. Ainsi, on peut conclure, avec une marge d'erreur de 5 %, que la structure par âge de la population féminine de la R.D.C en âge de procréer de 2001 est différente de celle de 2010. Cette structure a donc évolué entre les deux dates. Mais, malheureusement, avec le Khi-carré, il n'est pas facile d'identifier dans quels sens s'est réalisé cette évolution, ni quels sont les groupes d'âges spécifiques responsables de cette évolution. Pour cette fin, on recourt à d'autres types d'analyse comme l'analyse de différence des moyennes ou le t-test.

Comme dans les exemples précédents, ici également on peut recourir à l'ordinateur et à l'application SPSS pour réaliser le test de Khi-carré d'homogénéité. La démarche reste identique à celle développée pour le cas du test de Khi-carré d'indépendance. Toutefois, pour raison de commodité, on demande de placer la variable indiquant les différents échantillons en colonne (ici, il s'agit des différentes enquêtes à comparer : MICS2001 et MICS2010) et la variable étudiée (Ici âge ou groupes d'âges) en ligne et calculer les pourcentages en colonne. SPSS produit les résultats suivants :

Tableau 19 : Récapitulatif du traitement des observations

	Observations					
	Valide		Manquante		Total	
	N	Pourcent	N	Pourcent	N	Pourcent
Groupes d'âges * Enquêtes source	25259	98,5%	382	1,5%	25641	100,0%

Tableau 19 bis : Tableau croisé Groupes d'âges * Enquêtes source

		MICS2001 MICS2010	Enquêtes source		Total
			MICS2001	MICS2010	
Groupes d'âges	15-19 ans	Effectif	2 933	2 732	5 665
		% compris dans Enquêtes source	23,6%	21,3%	22,4%
	20-34 ans	Effectif	6 075	6 681	12 756
		% compris dans Enquêtes source	49,0%	52,0%	50,5%
	35-49 ans	Effectif	3401	3437	6 838
		% compris dans Enquêtes source	27,4%	26,7%	27,1%
Total		Effectif	12 409	12850	25 259
		% compris dans Enquêtes source	100,0%	100,0%	100,0%

Tableau 19 ter : Tests du Khi-deux

	Valeur	ddl	Signification asymptotique (bilatérale)
Khi-deux de Pearson	**28,420[a]**	2	**,000**
Rapport de vraisemblance	28,423	2	,000
Association linéaire par linéaire	3,767	1	,052
Nombre d'observations valides	25259		

a. 0 cellules (0,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 2783,05.

C'est le tableau croisé (Tableau 19 bis) qui sera présenté dans le rapport après avoir fait l'objet de quelques modifications de forme. Il pourrait se présenter comme suit :

Tableau 20 : Evolution (%) de la structure par âge de la population féminine de la RDC en âge de procréer entre MICS2001 et MICS2010

Groupes d'âge	MICS2001	MICS2010	Total
15-19 ans	23,6	21,3	22,4
20-34 ans	49,0	52,0	50,5
35-49 ans	27,4	26,7	27,1
Total	100,0 % (12409)	100,0 % (12850)	100,0 % (25259)
Khi-carré : 28,42****			

On pourrait également chercher à comparer la structure de la population féminine en âge de procréer de 2001 à celle de 2018, en répondant à la même question. On produit ici directement le tableau de contingence issu de SPSS, en indiquant le Khi-carré et sa probabilité associée.

Tableau 21 : Evolution (%) de la structure par âge de la population féminine de la RDC en âge de procréer entre 2001 et 2018

Groupes d'âge	MICS2001	MICS2018	Total
15-19 ans	23,6	23,3	22,4
20-34 ans	49,0	49,9	49,6
35-49 ans	27,4	26,7	27,0
Total	100,0 % (12409)	100,0 % (21756)	100,0 % (34165)
Khi-carré : 3,0774 ; *p*= 0.215			

Avec un Khi-carré de 3,0774 et sa probabilité associée de 0,215, supérieur à la marge d'erreur acceptable de 0,05, on ne rejette pas l'hypothèse nulle. Donc, il n'existe pas de différences statistiquement significatives entre la structure par âge de la population féminine congolaise en âge de procréer de 2001 et celle de 2018. Les deux structures sont quasiment identiques.

A noter que, comme on a pu probablement s'en rendre compte, le test de Khi-deux d'homogénéité et celui de l'indépendance sont quasiment identiques. Ils se différencient seulement du fait que la première traite d'une seule variable portant sur deux ou plusieurs échantillons différents alors que la seconde traite de deux variables portant sur deux ou plusieurs catégories d'un même échantillon.

VI. Intérêt, limites et pièges à éviter

Le test de Khi-deux est très facile à exécuter, tant manuellement qu'en recourant à un logiciel spécifique. Il est également très utile pour décrire une relation de dépendance entre deux variables catégorielles et sa compréhension est intuitivement facile. Toutefois, le test de Khi-deux n'indique pas l'intensité ni le sens de lien entre variables. Le test de Khi-carré reste également très sensible à la taille de l'échantillon. Plus la taille de l'échantillon est grande, plus les différences entre proportions risquent d'être significatives .

Le test de Khi-carré ne permet pas non plus de hiérarchiser les relations entre variables en identifiant celles qui sont fortement liées et celles qui le sont un peu moins. Il n'indique simplement que la confiance qu'il est possible d'accorder à l'hypothèse d'indépendance (H0). En effet, le test de Khi-carré ne rend compte que d'un effet général (dit omnibus). La conclusion qu'un pourcentage est différent entre un groupe A et un autre groupe B n'est possible que pour des tableaux 2 X 2. Pour des tableaux plus grands que 2 X 2, il n'est pas possible d'interpréter si la différence significative à l'origine de l'effet se situe entre les groupes A et B, B et C, ou A et C. Dans ce cas, il faut créer après coup (ex post-hoc) autant des tableaux de contingence que des combinaisons possibles de croisements 2 X 2 entre les modalités, et appliquer idéalement une correction de la p-valeur type ajustement de Bonferroni (Broc et Caumeil, 2018, p. 148). Par ailleurs, le test de Khi-carré ne se limite qu'à l'étude des variables prises deux à deux. Elle est complètement inefficace pour intégrer une troisième variable.

En outre, à l'instar d'autres analyses descriptives, l'existence d'une relation statistique entre deux variables induite par le Khi-carré ne signifie pas forcément que cette relation ait du sens empiriquement ou sociologique . Rappelons également, et non sans intérêt, qu'une relation ou une association statistique n'implique pas nécessairement une causalité, même s'il n'y a pas de causalité sans relation et sans association statistique préalable ! (Fox, 1999). En effet, le test de Khi-carré n'indique pas si une variable est à l'origine (la cause) de la variation d'une autre. Il se limite à décrire une relation ou une association entre variables sans l'expliquer. C'est ainsi que le test de Khi-carré constitue généralement une première étape avant d'aller vers des analyses explicatives plus poussées. Par exemple, il est inutile de procéder à une régression logistique entre deux variables si l'analyse de Khi-deux conclut à l'absence des liens entre elles. Qu'on se rappelle également les problèmes que peut poser l'intrusion d'une troisième variable, notamment les effets d'interaction et de confusion .

Etant donné que l'analyse de khi-carré constitue un cas particulier de l'analyse de différence de proportions, outre le problème lié aux faibles effectifs et à une multitude des modalités des variables d'analyse, un piège à éviter est celui lié au sens du calcul des proportions ou de pourcentages dans un tableau croisé. En effet, des proportions ou pourcentages calculés dans un mauvais sens peuvent aboutir à des conclusions erronées et partant, biaiser complètement l'interprétation des résultats. La règle d'or à retenir ici est que, quelle que soit la position des variables dans le tableau de contingence, les proportions ou les pourcentages doivent toujours être calculés dans le sens de la variable supposée indépendante. L'exemple repris ci-dessous illustre la mauvaise conclusion susceptible d'être tirée des pourcentages calculés dans le mauvais sens. Nous reprenons pour cela l'exemple du Tableau 12 portant sur le lien entre le sexe des migrants congolais et le fait de transférer de l'argent et d'autres biens à leurs ménages d'origine.

Tableau 12 : Répartition des migrants par sexe selon qu'ils ont ou non transféré de l'argent et autres biens à leurs ménages d'origine

Sexe	Transfert				Total
	N'a pas transféré		A transféré		
	Eff.	%	Eff.	%	
Masculin	344	64,9	389	52,1	733
Féminin	220	38,3	355	61,7	575
Total	564	100,0	744	56,9	1308
Khi-carré = 9,875***					

Tableau 12 bis : Répartition des migrants par sexe selon qu'ils ont ou non transféré de l'argent et autres biens à leurs ménages d'origine

Sexe	Transfert				Total
	N'a pas transféré		A transféré		
	Eff.	%	Eff.	%	
Masculin	344	61,0	389	52,3	733
Féminin	220	39,0	355	47,7	575
Total	564	100,0	744	100,0	1308
Khi-carré = 9,875***					

Dans le Tableau 12, les pourcentages ont été calculés dans le bon sens, en ligne, dans le sens de la variable « indépendante » alors que dans le Tableau 12 bis, les pourcentages ont été calculés dans le mauvais sens, dans le sens de la variable « dépendante ». Les interprétations à tirer des cellules de ces deux tableaux divergent. En effet, alors que dans le Tableau 12, les femmes sont proportionnellement plus nombreuses (près de 62 %) à avoir envoyé de l'argent et/ou d'autres biens à leurs ménages d'origine que les hommes (53,1 %) ; dans le Tableau 12 bis en revanche, ce sont les hommes qui paraissent proportionnellement plus nombreux (52 %) à avoir transféré de l'argent et/ou d'autres biens à leurs ménages d'origine que les femmes (près de 48 %).

De même, avec le Tableau 12 bis, on n'est même plus à mesure de connaître la proportion de migrants qui ont transféré indépendamment du sexe, alors que cette information est disponible dans le Tableau 12 (près de 57 %). Ce problème est dû à la surreprésentation des hommes par rapport aux femmes parmi les migrants, soit 56 % contre 44 %. Cette surreprésentation fait que les effectifs des hommes dépassent ceux des femmes dans toutes les cellules du tableau et de ce fait, influent sur le calcul des pourcentages sexes séparés. Pour éviter pareil désagrément, il est donc conseillé de calculer les pourcentages dans le sens de la variable « indépendante », même si les interprétations doivent se faire dans le sens de la variable « dépendante ».

Enfin, dans la réalité sociale, il n'est pas seulement question de conclure à l'existence d'une relation entre les deux variables qualitatives, encore faudrait-il chercher à déterminer, comme dans les cas de variables continues, l'intensité et le sens de cette relation. Pour cette fin, on peut compléter le test de Khi-deux avec l'analyse de risques relatifs (RR), l'analyse des rapports des cotes (Odds ratios ou OR) ou la production de certaines statistiques associées telles que le coefficient Phi, pour des tableaux 2 x 2, le coefficient de contingence pour des tableaux de n'importe quelle taille mais dont le nombre de lignes égale au nombre de colonnes, le V de Cramer pour des tableaux de n'importe quelle dimension. Ces différentes statistiques s'interprètent comme de coefficient de corrélation. Voir notamment Ghewy (2010), Masuy-Stroobant (2013), Dancey, C. et Reidy (2016) pour des amples détails.

VII. Conclusion

Le χ^2 est un test d'hypothèses basé sur la comparaison des valeurs observées et des valeurs théoriques ou attendues. La distribution théorique est celle qui correspond à l'hypothèse nulle (hypothèse d'égalité). Parce qu'il est facile à comprendre et à calculer, le Khi-carré est une forme courante de test d'hypothèses. On distingue généralement trois types de χ^2 à savoir celui de conformité ou d'ajustement, celui d'indépendance et celui d'homogénéité.

Le test de χ^2 de conformité ou d'ajustement est un test qui interroge la manière dont se répartissent les individus entre les modalités d'une seule variable catégorielle. On cherche à voir si une distribution d'effectifs observés est la même qu'une distribution théorique à priori. La distribution théorique peut être équiprobable lorsque les fréquences attendues sont égales dans les différentes modalités de la variable ou alors les effectifs théoriques ne sont pas tous égaux. Dans le cas de deux variables qualitatives, on cherche à savoir s'il existe une association (lien) entre variables, on teste alors l'indépendance des variables (test de X^2 d'indépendance). Les données sont présentées

tout simplement dans un tableau qui montre comment une caractéristique dépend d'une autre (tableau de contingence). Le test de Khi-carré d'homogénéité est utilisé pour comparer la distribution d'une variable catégorielle entre plusieurs échantillons. Avec ce test, on cherche à vérifier si les différents échantillons proviennent de la même population-mère, s'ils sont homogènes.

Il convient de noter que si le test de Khi-deux d'homogénéité traite d'une seule variable portant sur deux ou plusieurs échantillons différents indépendants, celui d'ajustement traite également d'une seule variable sur un seul échantillon tandis que le test d'indépendance traite quant à lui de deux ou plusieurs variables d'un même échantillon.

Le travail du χ^2 consiste toujours de comparer les effectifs théoriques et les effectifs observés. Sans assistance de logiciel on procède de la manière suivante :

- Faire apparaître les effectifs observés et théoriques
- Calculer la différence entre les effectifs théoriques et les effectifs observés
- Elever les différences au carré
- Diviser chacune de ces différences élevées au carré avec l'effectif théorique correspondant
- Additionner tous les quotients obtenus à l'étape 4 pour trouver le χ^2
- Déterminer le nombre de degré de liberté (ddl) et le seuil de signification (α)
- Comparer le χ^2 calculé à la valeur du χ^2 théorique dans la table
- Interpréter les résultats et décider du rejet ou non de l'hypothèse nulle de conformité, d'indépendance ou d'homogénéité

Comme il pose peu d'hypothèses quant à la population sous-jacente, on classe habituellement le test de khi-deux dans les tests non paramétriques. Cependant, avant chaque test d'hypothèse du χ^2, il est important de vérifier que :

- Les variables sont qualitatives ou catégorielles (ou des variables métriques ou ordinales dégradées en variables nominales) ;
- Les effectifs théoriques sont supérieurs à 5 dans chaque cellule (on peut admettre que moins de 25 % des cellules ne remplissent pas cette condition) ;
- C'est dans le cas de grandes valeurs du χ^2 que l'on peut conclure, un χ^2 trop petit n'est jamais considéré comme significatif.

Références bibliographiques

Broc, G. et Caumeil, B. (2018). Analyse de données. Louvain-la-Neuve : Deboeck Supérieur.

Dagnelie, P. (1998). Statistique théorique et appliquée. Tome 2. Inférence statistique à une et à deux dimension. Paris et Bruxelles : De Boeck et Larcier.

Dancey, C. P et Reidy, J. (2016). Statistiques sans maths pour les psychologues (2è édition française). Louvain-la-Neuve : Deboeck Supérieur.

Fox, W. (1999). Statistiques sociales. (3ième édition) (L. Imbeau, Traduction). Bruxelles : DeBoeck Université. (Travail original publié en 1998).

Genin, M. (2015). Le test de χ2. Notes de cours polycopiées. Université de Lille 2. Droit et Santé.

Ghewy, P. (2010). Guide pratique de l'analyse de données. Avec applications sous IBM SPSS Statistics et Excel. Questionnez, analysez…et décidez. Bruxelles : De Boeck.

Larose, D. T et Larose, D. C. (2018). Data mining. Découverte de connaissances dans les données. 2ième édition. (T. Vallaud, Traduction). Paris : Vuibert. (Travail original publié en 2014).

Mangalu, M.A. (2011). Migrations internationales, transferts des migrants et bien-être des ménages d'origine. Cas de la ville de Kinshasa. Louvain-la-Neuve : Presses

Universitaires de Louvain.

Mangalu, M.A.J., Mbo, V. J. et Djoke, L. (2016). Notes du cours de statistique appliquée au développement : deuxième licence en sciences de développement. Syllabus, Institut Facultaire de Développement.

Martin, O. (2009). L'enquête et ses méthodes. L'analyse de données quantitatives (2è édition). Paris : Armand Colin.

Masuy-Stroobant, G. (2013). Deux variables qualitatives. In Masuy-Stroobant, G. et Costa, R. Analyser les données en sciences sociales. De la préparation des données à l'analyse multivariée. Bruxelles : P.I.E. Peter Lang, pp.77-93.

Py, B. (2007). La statistique sans formule mathématique. Comprendre la logique et maîtriser les outils. Paris : Pearson Education.

Wiley, J. (1991). Statistiques. Economie-Gestion-Sciences-Médecine. 4ième édition. (T.H. Wonnacott & R.J. Wonnacott, Traduction). Paris : Economica.