CHAPITRE 6

Le test du chi-carré

Dans ce chapitre, nous traiterons des tests de signification statistique – c'est-à-dire ce sur quoi l'on se base pour décider si la relation que nous avons découverte à partir de données d'échantillon est susceptible de se retrouver aussi dans la population de laquelle provient l'échantillon. Nous porterons un intérêt tout spécial au test de signification du chi-carré pour des relations décrites par un tableau bivarié. Par moments, ce chapitre demandera de votre part un certain effort d'abstraction. Aussi devrez-vous peut-être le relire sérieusement plusieurs fois.

Après ce chapitre vous pourrez :

1. Expliquer la logique des tests de signification statistique.
2. Expliquer et donner des exemples d'hypothèses « nulles ».
3. Expliquer la différence entre les erreurs de type I et les erreurs de type II et donner des exemples de ces deux types d'erreurs.
4. Accomplir et interpréter des tests du chi-carré à l'aide de tableaux de données.
5. Lire un tableau de chi-carré et interpréter les valeurs du chi-carré en termes de signification statistique.
6. Connaître les difficultés qui peuvent survenir lorsque l'on se sert du test du chi-carré à partir de fréquences anticipées inférieures à 5. Résoudre ce problème en excluant ou en regroupant des valeurs.
7. Expliquer la différence entre la signification statistique et la signification « réelle » ; expliquer la relation qui existe entre la signification statistique et la taille de l'échantillon et l'intensité de la relation ; et être prudent dans l'interprétation des résultats de test de chi-carré quand les N sont très élevés.
8. Savoir, ici aussi, que l'association n'implique pas la causalité.

6.1 La logique des tests de signification statistique

Les méthodes d'analyse de tableau dont nous avons traité au chapitre précédent répondent à cette question : Y a-t-il une relation entre les deux variables pour les données que nous examinons ? Or si nous découvrons qu'une relation existe et si les données proviennent d'un échantillon plutôt que d'une population entière, nous devons poser une seconde question : Y a-t-il une relation dans cette population plus vaste de laquelle fut extrait l'échantillon ?

Après tout, il se peut bien que soit dû au hasard le fait qu'un ensemble particulier de cas montre une relation entre deux variables alors que, en réalité, il n'y a pas de telle relation à l'intérieur de la population prise comme un tout. Peut-être, par exemple, n'y a-t-il pas réellement de relation entre le niveau d'instruction et l'attitude face à la désobéissance civile à l'intérieur de la population de tous les Américains adultes. En effet, il est possible que l'échantillon du General Social Survey ou que les 50 cas du sous-échantillon que nous avons selectionné soient par hasard des répondants parmi lesquels les plus instruits tendent à suivre leur propre conscience et les moins instruits à obéir aux lois. Comme nous l'avons constaté avec l'exemple du QI dans la section 4.5, si les sondeurs du GSS avaient choisi 1606 répondants différents, peut-être n'aurions-nous pas trouvé de relation entre le niveau d'instruction et l'attitude face à la désobéissance civile.

(À partir d'ici, le raisonnement devient un peu plus complexe. Concentrez-vous donc sérieusement en lisant le reste de cette section, spécialement les deux prochains paragraphes. Aussi, lisez-les plus d'une fois.)

Ce qu'il nous faut, c'est une façon de déterminer la probabilité (c'est-à-dire les chances) de découvrir une relation dans notre échantillon quand il n'y en a pas dans notre population. Si cette probabilité est petite (disons une probabilité de 1 sur 20), et si nous découvrons une relation dans notre échantillon, nous pourrons conclure qu'il existe probablement une relation dans la population. En effet, on a peu de chance de trouver une relation dans notre échantillon s'il n'en existe pas dans la population. Si nous en avons trouvé une, il est donc probable qu'il y ait également une relation dans notre population. Les statisticiens usent du terme ***hypothèse nulle*** (que l'on note souvent H_0) pour indiquer la supposition voulant qu'il n'y ait pas de relation dans notre population. Pour employer le jargon statistique, nous ***rejetons l'hypothèse nulle*** affirmant l'absence de relation dans notre population. Il existe probablement une relation dans la population, et nous disons de cette relation qu'elle est ***statistiquement significative.***

À l'inverse, si les chances de trouver une relation dans l'échantillon alors qu'il n'y en a pas dans la population sont élevées (supérieures à 1 sur 20), nous ne pouvons croire en toute confiance à l'existence d'une relation dans la population plus vaste. La relation que nous avons découverte à l'intérieur de l'échantillon ne devait peut-être son existence qu'au hasard seul. Dans un tel cas, nous ***ne rejetons pas l'hypothèse nulle*** affirmant l'absence de relation dans la population. Il pourrait très bien n'y avoir aucune relation dans la population, et nous disons de la relation qu'elle n'est *pas* significative statistiquement.

La probabilité que l'on retrouve grâce au seul hasard une relation à l'intérieur d'un échantillon, donc en dépit de l'absence de relation dans la population d'où provient l'échantillon, est nommée le ***niveau de signification*** de la relation. Le niveau de signification est exprimé par un nombre de 0 à 1,00. Il s'agit là de la façon conventionnelle d'exprimer les probabilités. Une ***probabilité*** indique les chances que quelque chose survienne – dans le cas présent, les chances que nous découvrions une relation à l'intérieur d'un échantillon sans qu'il y ait semblable relation dans la population.

Plus précisément, une probabilité que l'on multiplie par 100 (afin de se débarrasser des décimales) indique le nombre de fois qu'un événement est susceptible de survenir à la suite de 100 « essais ». Une probabilité de 0 signifie que les chances que se produise un événement quelconque sont inexistantes – c'est donc impossible ; cela ne surviendra jamais (comme la semaine des quatre jeudis). Quelque chose qui survient 1 fois sur 100 a une probabilité de 0,01. Quelque chose qui survient 1 fois sur 20 a une probabilité de 0,05. Une probabilité de 0,5 signifie qu'un événement est susceptible de se produire 50 fois sur 100 – comme obtenir « face » en tirant à pile ou face. Une probabilité de 1,00 signifie que quelque chose est absolument certain – un coup sûr, un événement qui surviendra 100 fois sur 100 (il y a bien peu d'événements de ce genre dans la vie, n'est-ce pas ?). Donc plus la probabilité de quelque chose est faible, moins ce quelque chose est susceptible de se produire. Plus la probabilité est élevée, plus il est susceptible de se produire.

J'ai mentionné que nous concluons à la signification statistique lorsque les chances de trouver par hasard la relation observée sont inférieures à 1 sur 20. Il y a une raison qui justifie mon choix de 1 sur 20 comme point limite pour décider de la signification statistique. Par convention, les chercheurs considèrent qu'une probabilité égale ou inférieure à 0,05 (à 1 sur 20) est raisonnable pour conclure à la signification statistique. Ils expriment cette signification statistique par l'expression $p < 0,05$ (p indiquant « probabilité »). Évidemment les chercheurs expriment de plus faibles niveaux de signification statistique, donc des probabilités plus faibles, lorsqu'ils en découvrent. (Des niveaux de signification aussi faibles sont communément arrondis à 0,01 ou à 0,001 et, par conséquent, exprimés $p < 0,01$ ou $p < 0,001$.) Toutefois, c'est la plupart du temps le niveau 0,05 que nous utilisons comme limite de la signification statistique. Si la probabilité de trouver la relation que nous avons découverte est égale ou inférieure à 0,05, nous pouvons croire avec une confiance raisonnable que la relation existe « réellement » et qu'elle n'est pas simplement l'effet du hasard. Tout compte fait, une probabilité égale ou inférieure à 0,05 laisse peu de chances de découvrir une relation dans un échantillon alors qu'il n'y en a pas dans la population. Aussi pouvons-nous, dans ce cas, rejeter l'hypothèse nulle qui affirme l'absence de relation.

Gardez cependant à l'esprit qu'il n'y a rien d'absolu ou d'éternel dans ce seuil de 0,05. C'est un niveau de signification dont l'usage n'est justifié que par la convention. Ce ne sont pas tous les chercheurs ni tous les statisticiens qui acceptent cette convention. En fait, certains le nomment avec sarcasme le seuil « sacro-saint » de 0,05. Ce seuil n'est qu'une création de la culture ; il n'a pas été révélé par une quelconque divinité des statistiques. Si l'on en croit l'Ancien Testament, Dieu est probablement, Lui-même, plus exigeant. En vérité, le seuil de 0,05 a été « révélé » (pour parler de la sorte) par Sir Ronald Fisher en 1925 dans son fameux livre *Statistical Methods for Research Workers*. Beaucoup de disciplines emploient un niveau de signification beaucoup plus élevé que 0,05. Par exemple, les physiciens qui étudient les particules subatomiques appliquent une probabilité beaucoup plus exigeante de 0,0001 ou moins avant de reconnaître la détection d'une nouvelle particule. Mais, nous, les chercheurs en sciences sociales, nous nous contentons d'un niveau de 0,05.

Le fait que la signification statistique repose sur une probabilité implique que nous ne pouvons jamais être absolument certains de faire le bon choix lorsque nous rejetons ou conservons H_0. Après tout, nous ne pouvons jamais savoir avec certitude si, à partir d'informations provenant d'un échantillon, deux variables sont reliées dans la population. Seules des données de population pourraient nous donner cette certitude. Les erreurs sont toujours possibles lorsque nous travaillons avec des données d'échantillon parce qu'il se peut que notre échantillon ne soit pas représentatif. Même avec des techniques d'échantillonnage aléatoire on peut obtenir un échantillon non représentatif (même si en général, cela ne se produit pas).

Lorsque nous rejetons l'hypothèse nulle alors qu'elle est vraie, nous commettons ce que les statisticiens appellent une ***erreur***

probabilité que nous commettions une erreur de type I en rejetant l'hypothèse nulle. Le niveau de signification est parfois appelé le ***niveau alpha*** et est représenté par le symbole α. Au seuil de 0,05, nous serons dans l'erreur environ 1 fois sur 20.

Mais nous commettons également une erreur lorsque nous ne rejetons pas l'hypothèse nulle et que celle-ci est, en réalité, fausse. Démontrant aux yeux de tous leur habilité à compter, même en chiffres romains, les statisticiens ont nommé cela une ***erreur de type II*** ou ***erreur bêta***. Le tableau 6.1 résume ces deux types d'erreurs. (H_0 indique l'hypothèse nulle.)

Tableau 6.1. Erreurs de type I et de type II

Décision concernant H_0	Si H_0 est vraie	Si H_0 est fausse
Rejeter H_0	Erreur de type I	Pas d'erreur
Conserver H_0	Pas d'erreur	Erreur de type II

Puisque la probabilité d'une erreur de type I est α, il est tentant de conclure que la probabilité d'une erreur de type II soit $1 - \alpha$. Eh bien, c'est faux : un α de 0,05 ne veut pas dire que la probabilité d'une erreur bêta soit 0,95. Les erreurs bêta sont un peu plus compliquées que cela, et vont au-delà de ce que nous pouvons discuter dans cet ouvrage. Il est toutefois vrai que les erreurs de type I et de type II sont inversement liées et qu'en réduisant les chances d'une erreur de type I on augmente les chances d'une erreur de type II, et vice-versa.

Que nous décidions de rejeter ou de ne pas rejeter l'hypothèse nulle, il n'existe aucune manière d'éliminer complètement la possibilité de commettre une erreur. Nous pouvons cependant déterminer la probabilité de commettre une erreur de type I lorsque nous rejetons l'hypothèse nulle puisque cette probabilité est égale au niveau de signification statistique.

C'est à ce moment que le concept de distribution d'échantillonnage, dont nous avons pris connaissance à la section 4.5, s'avère particulièrement utile. La distribution d'échantillonnage nous permet de connaître le niveau de signification – c'est-à-dire la probabilité de trouver une relation dans notre échantillon alors que, dans la population, il n'y a pas de relation entre les variables. Par conséquent, la distribution d'échantillonnage permet de décider rationnellement de la signification statistique de la relation. La distribution d'échantillonnage et la méthode pour déterminer la signification statistique qui sont les plus appropriées dépendent de la nature des données que nous analysons. Pour des données disposées en tableau, les statisticiens ont conçu un test de signification largement répandu, appelé test du chi-carré, qui repose sur la distribution d'échantillonnage du chi-carré. Voyons comment fonctionne ce test.

2 Le test du chi-carré

Le ***chi-carré***, représenté par la lettre grecque χ^2 et que l'on prononce « ki-carré », est un nombre qui compare les fréquences observées dans un tableau bivarié aux fréquences auxquelles on devrait s'attendre s'il n'y avait pas du tout de relation entre les deux variables dans la population (les fréquences « anticipées »). La distribution du chi-carré est une distribution d'échantillonnage très utile. En supposant que, dans la population, les deux variables ne soient pas reliées, on peut déterminer la probabilité d'obtenir un chi-carré de telle ou telle taille.

Voici la logique du ***test du chi-carré***. Si les ***fréquences observées*** dans un tableau bivarié construit à partir de données d'échantillon sont similaires aux fréquences auxquelles on s'attendrait si, dans la population, il n'y avait pas de relation entre les deux variables (les fréquences anticipées), on peut alors croire qu'il n'y a vraiment pas de relation dans la population (et courir le risque de commettre une erreur de type II). Par contre, si les fréquences observées dans un tableau sont très différentes de celles auxquelles on s'attendrait si, dans la population, les variables n'étaient pas reliées, on rejette alors la supposition selon laquelle il n'y a pas de relation entre les variables. Par conséquent, on conclut qu'il y a probablement une relation entre les variables (et on court le risque de commettre une erreur de type I).

Mais comment trouve-t-on les ***fréquences anticipées*** ? Prenons un exemple hypothétique pour faciliter les calculs. Considérons de nouveau les 50 cas portant sur la relation Niveau d'instruction → Attitude face à la désobéissance civile que l'on a présentée à la section 5.1. La distribution bivariée de fréquences de cette relation est présentée au tableau 6.2.

Tableau 6.2. Attitude face à la désobéissance civile selon le niveau d'instruction (en fréquences)

	Niveau d'instruction
Désobéissance	Moins que le

S'il n'y avait pas de relation entre l'instruction et l'attitude face à la désobéissance civile, on devrait trouver dans chaque catégorie de la variable « niveau d'instruction » la même proportion de répondants qui suivent leur conscience. De même on devrait trouver pour chaque niveau d'instruction la même proportion de gens qui obéissent aux lois. C'est ce qu'on veut dire quand on dit qu'il n'y a pas de relation entre les variables : la **proportion de gens moins instruits qui suivent leur conscience** est la **même que la proportion de diplômés de l'école secondaire qui suivent leur conscience**, et ces deux proportions sont égales à la proportion des personnes qui ont un niveau d'instruction postsecondaire qui suivent leur conscience. Si ces proportions sont les mêmes, les pourcentages seront les mêmes et une comparaison des pourcentages d'une catégorie de la variable indépendante à l'autre montrera qu'il n'existe aucune différence. En fait chacune des distributions de la variable « Désobéissance civile » à l'intérieur des catégories du niveau d'instruction serait proportionnelle à la distribution totale de la variable « Désobéissance civile » qui se trouve être la distribution marginale de la colonne de droite du tableau 6.2. (S'il vous plaît, relisez ce paragraphe !)

Il est facile de trouver les fréquences anticipées, c'est-à-dire les fréquences auxquelles on s'attendrait s'il n'y avait pas de relation entre les variables. Considérons d'abord les fréquences anticipées auxquelles on devrait s'attendre dans la première rangée. Puisque 29/50 de tous les répondants suivent leur conscience, nous nous attendons à ce que 29/50 des 10 répondants ayant un niveau d'instruction inférieur au secondaire suivent leur conscience. C'est-à-dire : $\left(\frac{29}{50}\right)10 = 5{,}80$ répondants. De même, puisque 29/50 de tous les répondants suivent leur conscience, nous nous attendons à ce que 29/50 des 24 répondants qui sont titulaires d'un diplôme d'études secondaires suivent leur conscience. C'est-à-dire $\left(\frac{29}{50}\right)24 = 13{,}92$ répondants. En appliquant le même raisonnement pour les répondants qui ont une éducation postsecondaire, on obtient $\left(\frac{29}{50}\right)16 = 9{,}28$ répondants.

Ainsi, pour chaque cellule, tout ce que nous avons à faire est :

1. de diviser le total de la rangée par le grand total N ;

En formule mathématique, cela s'exprime comme suit :

$$f_a = \left(\frac{\text{Total de la rangée}}{N}\right) \text{Total de la colonne}$$

lorsque f_a = fréquence anticipée d'une cellule

N = le nombre total de cas

C'est ainsi que nous trouvons les fréquences que l'on devrait observer dans chaque cellule s'il n'y avait pas de relation entre les variables.

Appliquons la formule afin de trouver le nombre de personnes les moins instruites qui suivraient leur conscience s'il n'y avait pas de relation entre les variables :

$$f_a = \left(\frac{\text{Total de la rangée}}{N}\right) \text{Total de la colonne}$$

$$= \left(\frac{29}{50}\right) 10$$

$$= (0{,}58)\ (10)$$

$$= 5{,}80$$

Donc, il y aurait 5,80 cas (plutôt que les 4 cas que nous avons observés) dans la cellule du bas la plus à gauche s'il n'existait pas de relation entre le niveau d'instruction et l'attitude face à la désobéissance civile.

On peut aussi trouver le nombre anticipé de diplômés du secondaire qui suivent leur conscience :

$$f_a = \left(\frac{29}{50}\right) 24$$

$$= (0{,}58)\ (24)$$

$$= 13{,}92$$

S'il n'y avait pas de relation entre les variables, nous aurions 13,92 cas (plutôt que les 13 cas observés) dans la cellule « Secondaire-Conscience ».

Je vous laisse le soin de calculer la fréquence anticipée des répondants ayant une éducation postsecondaire qui disent qu'il faut suivre sa conscience. (Je compte sur vous pour faire en sorte d'obtenir une fréquence anticipée de 9,28.)

Le tableau 6.3 présente les fréquences pour cette relation, avec les fréquences anticipées entre parenthèses :

Tableau 6.3. Attitude face à la désobéissance civile selon le niveau d'instruction (en fréquences et fréquences anticipées)

	Niveau d'instruction			
Désobéissance civile	Moins que le secondaire	Secondaire	Postsecondaire	Total
Conscience	4	13	12	29
	(5,80)	(13,92)	(9,28)	
Obéir aux lois	6	11	4	21
	(4,20)	(10,08)	(6,72)	
Total	10	24	16	50

Additionnez les fréquences anticipées et vous verrez que les **totaux correspondent aux distributions marginales**. Par exemple, 5,80 + 13,92 + 9,28 = 29, et 5,80 + 4,20 = 10. Autrement dit, les fréquences anticipées reproduisent les distributions marginales originales. De plus – vous n'en serez pas surpris si vous avez suivi la logique de ce que nous avons fait –, si nous calculons les fréquences anticipées en pourcentages, nous observons qu'il n'y a pas de relation entre les variables.

Tableau 6.4. Attitude face à la désobéissance civile selon le niveau d'instruction basé sur les fréquences anticipées (en pourcentages)

	Niveau d'instruction			
Désobéissance civile	Moins que le secondaire	Secondaire	Postsecondaire	Total
Conscience	58	58	58	58
Obéir aux lois	42	42	42	42
Total	100	100	100	100
(N)	(10)	(24)	(16)	(50)

C'est ainsi que cela doit être puisque, pour calculer les fréquences anticipées, nous supposons justement qu'il n'y a pas de relation entre les variables.

Nous sommes maintenant prêts à calculer le chi-carré. Voici la formule que nous utiliserons :

$$\chi^2 = \Sigma \frac{(f_o - f_a)^2}{f_a}$$

lorsque χ^2 = chi-carré

f_o = fréquence observée de chaque cellule

f_a = fréquence anticipée de chaque cellule

Vous vous souvenez que le signe Σ (sigma) veut dire « additionner tout ce qui suit ». Ainsi pour trouver le chi-carré calculez d'abord la fréquence anticipée de chacune des cellules et servez-vous ensuite de la formule :

1. soustrayez la fréquence anticipée de la fréquence observée pour chaque cellule ;
2. élevez au carré chacune de ces différences ;
3. divisez chaque différence au carré par la fréquence anticipée de la cellule ;
4. additionnez tous ces chiffres pour toutes les cellules.

Le résultat est le χ^2. N'oubliez pas d'élever au carré la différence dans chaque cellule (étape 2) avant de diviser par la fréquence anticipée de cette cellule.

En examinant le numérateur (c'est-à-dire la partie supérieure) de la formule, vous pourrez constater que plus la différence entre la fréquence observée et la fréquence anticipée est grande, plus le χ^2 est grand. Puisque le calcul des fréquences anticipées repose sur la supposition voulant qu'il n'y ait pas de relation, des différences plus grandes entre les fréquences observées et les fréquences anticipées indiquent une relation plus forte (c'est-à-dire une relation qui s'éloigne d'une relation nulle). Ainsi plus la relation est forte (pour un nombre de cas donné), plus le chi-carré est élevé.

En appliquant la formule du chi-carré aux fréquences observées et anticipées de la relation Instruction → Désobéissance civile, on obtient le résultat suivant :

$$\chi^2 = \Sigma \frac{(f_o - f_a)^2}{f_a}$$

$$= \frac{(4-5{,}80)^2}{5{,}80} + \frac{(6-4{,}20)^2}{4{,}20} + \frac{(13-13{,}92)^2}{13{,}92} + \frac{(11-10{,}08)^2}{10{,}08} + \frac{(12-9{,}28)^2}{9{,}28} + \frac{(4-6{,}72)^2}{6{,}72}$$

$$= \frac{3{,}2400}{5{,}80} + \frac{3{,}2400}{4{,}20} + \frac{0{,}8464}{13{,}92} + \frac{,8464}{10.08} + \frac{7{,}3984}{9{,}28} + \frac{7{,}3984}{6{,}72}$$

$$= 0{,}5586 + 0{,}7714 + 0{,}0608 + 0{,}0840 + 0{,}7972 + 1{,}1010$$

$$= 3{,}373$$

Vous trouverez peut-être qu'il est plus simple de calculer le χ^2 à l'aide d'un tableau de calcul tel que le tableau 6.5.

Tableau 6.5. Calcul du chi-carré pour l'attitude face à la désobéissance civile selon le niveau instruction

f_o	f_a	$f_o - f_a$	$(f_o - f_a)^2$	$(f_o - f_a)^2 / f_a$
4	5,80	–1,80	3,2400	0,5586
6	4,20	1,80	3,2400	0,7714
13	13,92	–0,92	0,8464	0,0608
11	10,08	0,92	0,8464	0,0840
12	9,28	2,72	7,3984	0,7972
4	6,72	-2,72	7,3984	1,1010
Total				$\chi^2 = 3{,}3730$

Ici les statisticiens ont fait une chose merveilleuse pour nous. Ils ont estimé la distribution d'échantillonnage du chi-carré. C'est-à-dire qu'ils ont calculé la probabilité d'obtenir un chi-carré au moins aussi grand qu'une certaine valeur si, dans la population de laquelle fut tiré l'échantillon, il n'y a pas de relation entre les deux variables. Ce calcul implique que les données que nous analysons proviennent d'un échantillon aléatoire, postulat qui est respecté dans les sondages rigoureux, tel le General Social Survey.

Cette probabilité dépend en partie des ***degrés de liberté (dl)*** du chi-carré. Rappelez-vous que nous avons traité des degrés de liberté lorsque nous avons examiné la variance et l'écart-type pour des données d'échantillon à la section 4.1. Pour le chi-carré le nombre de degrés de liberté équivaut à (r – 1) (c – 1), lorsque r et c sont le nombre de rangées et de colonnes dans le tableau (en omettant la rangée et la colonne « Total »). Donc, dl = 1 pour un tableau 2 x 2, dl = 2 pour un tableau 2 x 3, dl = 3 pour un tableau 2 x 4, ainsi de suite. Pour le tableau 2 x 3 présenté plus haut :

$$
\begin{aligned}
dl &= (r-1)(c-1) \\
&= (2-1)(3-1) \\
&= (1)(2) \\
&= 2
\end{aligned}
$$

La distribution du chi-carré est donnée au tableau 1 de l'appendice. Le tableau donne la valeur minimale du chi-carré nécessaire pour obtenir un résultat statistiquement significatif aux seuils 0,05, 0,02, 0,01 et 0,001. Pour utiliser le tableau trouvez d'abord les degrés de liberté dans la colonne de gauche puis, en allant vers la droite, voyez quelle valeur minimale de chi-carré vous devez obtenir pour rejeter l'hypothèse nulle à un niveau donné de signification. Voici les rangées du tableau de chi-carré pour 2 degrés de liberté :

	Probabilité			
dl	0,05	0,02	0,01	0,001
.	.	.	.	.
2	5,991	7,824	9,210	13,815
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

Ainsi, avec 2 degrés de liberté, il nous faut un chi-carré d'au moins 5,991 pour obtenir un résultat qui soit statistiquement significatif au seuil de signification 0,05, un chi-carré d'au moins 7,824 pour que le résultat soit statistiquement significatif au seuil 0,02, un chi-carré d'au moins 9,210 au seuil 0,01, et un chi-carré d'au moins 13,815 au seuil 0,001.

Notre chi-carré est de seulement 3,373, **moins que le chi-carré correspondant au seuil 0,05. Nous ne pouvons donc pas rejeter l'hypothèse nulle selon laquelle il n'y a pas de relation entre l'instruc**tion et l'attitude face à la désobéissance civile dans la population d'où provient notre échantillon hypothétique de 50 cas. Nous aurions pu trouver les fréquences que nous observons dans nos données d'échantillon simplement par hasard, donc malgré l'absence de relation entre les deux variables dans la population. Comme la relation que nous avons mesurée dans notre échantillon aurait pu être due au hasard, nous sommes peu enclins à rejeter l'hypothèse nulle voulant qu'il n'y ait pas de relation dans la population. Dans le jargon statistique, « nous échouons dans notre tentative de rejeter l'hypothèse nulle ». Prenez note qu'il ne nous arrive jamais d'« accepter » l'hypothèse nulle. **Le raisonnement statistique procède plutôt par falsification, c'est-à-dire en rejetant l'opposé d'une idée plutôt que d'accepter l'idée elle-même**. C'est de cette façon que fonctionne la science et ce n'est pas par hasard que cette méthode donne de si bons résultats.

Bien sûr, nous pourrions faire une erreur (une erreur de type II) en ne rejetant pas l'hypothèse nulle si elle était vraiment fausse. Mais la probabilité de découvrir la relation que nous observons dans nos données d'échantillon sans qu'il y ait de relation dans la population est tout de même supérieure à 1 sur 20.

Avec l'habitude, il est facile de tenir pour acquises des statistiques telles que le chi-carré et d'oublier leur sens. J'espère que vous vous souviendrez que le chi-carré est une statistique au sens technique du terme. C'est-à-dire que c'est une caractéristique d'un échantillon, comme une moyenne, un écart-type ou un pourcentage peuvent être les caractéristiques d'un échantillon. Rappelez-vous également que les tableaux de chi-carré en annexe décrivent une distribution d'échantillonnage. Ils décrivent la distribution d'une statistique (le chi-carré) en indiquant la probabilité associée avec tous les résultats d'échantillon possibles (c'est-à-dire tous les chi-carrés possibles). Pour les deux variables que nous analysons, imaginez tous les tableaux bivariés possibles pour un N donné, chacun ayant un chi-carré associé. Le tableau de chi-carré en annexe montre la distribution de ces chi-carrés en établissant la probabilité d'obtenir chacun d'eux s'il n'y a pas de relation dans la population. Nous déterminons alors où se trouve le chi-carré de notre relation dans cette distribution. Ainsi nous découvrons la possibilité d'obtenir notre chi-carré s'il n'y avait pas de relation dans la population. Cette **probabilité constitue le niveau de signification statistique de la relation de notre échantillon**.

Incidemment, les chercheurs indiquent souvent le chi-carré et la probabilité qui y est associée juste au-dessous du tableau, comme cela est fait dans le tableau 6.6. Ici cette probabilité est supérieure à 0,05, ce qui indique que la relation n'est pas statistiquement significative. L'abréviation « n.s. » signifie « non significatif ». L'abréviation « p » signifie « probabilité ». Elle est suivie de la probabilité correspondant à la valeur du chi-carré. Quand la relation est statistiquement significative, on indique les probabilités de la façon suivante : $p < 0,05$, $p < 0,01$, $p < 0,02$ ou $p < 0,001$, selon le niveau de signification du χ^2. Il arrive souvent aussi qu'une mesure d'association comme celles que nous allons voir au chapitre suivant soit rapportée à la suite des résultats du test du chi-carré.

Tableau 6.6. Attitude face à la désobéissance civile selon le niveau d'instruction (en pourcentages)

Désobéissance civile	Niveau d'instruction: Moins que le secondaire	Secondaire	Postsecondaire
Conscience	40	54	75
Obéir aux lois	60	46	25
Total	100	100	100
(N)	(10)	(24)	(16)

$\chi^2 = 2{,}00$; n.s.

Les tests de signification statistique comme le test du chi-carré sont des outils très utiles dans l'analyse des données. Ils nous disent jusqu'à quel point nous pouvons croire que nos résultats reflètent la réalité plutôt que seulement l'effet du hasard du processus d'échantillonnage. Mais tous les outils doivent être utilisés avec précaution. (Vous êtes-vous déjà donné un coup de marteau sur le pouce ?) Les trois sections qui suivent décrivent les principales précautions qu'il faut prendre lorsque l'on se sert du test du chi-carré.

6.3 Les problèmes causés par des fréquences anticipées inférieures à 5

J'ai volontairement choisi de petits nombres dans l'exemple qui précède pour que nous puissions concentrer notre attention sur la logique et le fonctionnement du test du chi-carré. Mais je dois admettre que ces petits nombres posent un sérieux problème. En effet, il faut toujours être très prudents lorsque l'on emploie le test du chi-carré alors que la fréquence anticipée d'une des cellules est inférieure à 5. Cela est particulièrement vrai quand les fréquences anticipées de plus de 20 % des cellules sont aussi faibles. (L'exemple Instruction → Désobéissance civile s'approche dangereusement de cette limite, avec une fréquence anticipée inférieure à 5 dans une des six cellules, c'est-à-dire 16,7 % du total des cellules.) Nous pouvons indiquer la probabilité correspondant au chi-carré, mais nous ne pouvons guère nous y fier. La raison de cela est que la distribution du chi-carré est supposée être continue, le chi-carré pouvant prendre n'importe quelle valeur. Mais, lorsque les fréquences anticipées sont petites, les valeurs possibles du chi-carré sont mathématiquement limitées, ce qui viole le postulat de continuité sur lequel repose le test. Par conséquent, quand les fréquences anticipées sont inférieures à 5 nous inscrivons, en guise de niveau de signification, la mention « ne s'applique pas » (en abrégé « n.a. »). D'autres tests (comme le test exact de Fisher) peuvent être utilisés lorsque les fréquences anticipées sont petites. Ces autres tests sont décrits dans des manuels de statistiques plus avancées cités en bibliographie.

Comme les fréquences anticipées dépendent des fréquences marginales, c'est seulement lorsque les valeurs de l'une des variables (ou des deux) ont très peu de cas que se pose ce problème des faibles fréquences anticipées. Voilà qui devrait nous suggérer quelques solutions. Songez à fusionner les valeurs moins fréquentes de façon à ce que les fréquences anticipées atteignent au moins 5. Ou bien si la variable est nominale vous pouvez simplement exclure les catégories à l'origine des petites fréquences anticipées. Ces solutions au problème des faibles fréquences anticipées sont les mêmes que celles qui ont été présentées à la section 5.7 pour remédier aux problèmes causés par des pourcentages reposant sur de faibles N. Comme toujours, n'excluez ou ne fusionnez des fréquences que si cela a un sens et vous aide à atteindre les objectifs de votre recherche.

6.4 Signification statistique et signification « réelle »

Une autre mise en garde à propos du test du chi-carré (ou de n'importe quel autre test de signification statistique) : un résultat *statistiquement* significatif n'est pas nécessairement un résultat « *réellement* » significatif à l'aune de la question étudiée. Nous qualifions une relation de « réellement » significative seulement lorsqu'elle est passablement forte. Qui se préoccupe d'une relation basée sur une différence de quelques points de pourcentage seulement, même si elle est généralisable à la population (c'est-à-dire statistiquement significative) ? Or la signification statistique ne dépend pas seulement de l'intensité de la relation mais aussi du nombre de cas dans l'échantillon. Cela est logique puisque nous avons plus confiance en des généralisations reposant sur un grand échantillon que sur un petit échantillon. Mais cela veut dire que, si nous avons un nombre suffisant de cas, une relation très faible au point d'être triviale sera statistiquement significative. De même une grande différence entre les pourcentages ne sera pas significative si elle est basée sur un nombre très faible de cas.

Examinons, en guise d'exemple, les trois tableaux hypothétiques 6.7a, b et c qui décrivent la relation entre la démarche et le caquetage de canard, d'une part, et le fait d'être un canard, d'autre part. L'hypothèse que nous testons correspond au vieux cliché qui veut que seul un canard ait une démarche de canard et caquette comme un canard, une hypothèse qui contredit le fameux proverbe « L'habit ne fait pas le moine ». Évidemment nous nous intéressons aux nombres et non à cette hypothèse farfelue bien qu'une telle hypothèse puisse nous rappeler que parfois la signification statistique n'est pas aussi « significative » qu'on le voudrait. Dans chaque tableau on observe la même différence de 20 points de pourcentage mais des niveaux de signification passablement différents en fonction du nombre de cas. Pour seulement 50 cas (tableau 6.7a), la différence de 20 points de pourcentage n'est pas significative statistiquement. Pour 100 cas (6.7b), la même différence de 20 points est statistiquement significative au seuil 0,05. Et pour 500 cas (6.7c), la même différence est significative au seuil 0,001. Cela est normal car nous sommes plus confiants qu'une relation existe « vraiment » lorsqu'elle est basée sur un plus grand nombre de cas. Même une grande différence, si elle repose sur un faible nombre de cas, ne peut pas être généralisée avec confiance. Et même une très petite différence sans importance peut être généralisée à la population si elle est découverte dans un échantillon suffisamment grand.

Tableau 6.7a.
Fait d'être un canard selon la démarche et le caquetage pour 50 cas (en pourcentages)

Est un canard	Marche et caquette comme un canard ? Oui	Non
Oui	60	40
Non	40	60
Total	100	100
(N)	(25)	(25)

$\chi^2 = 2{,}00$; n.s.

Tableau 6.7b.
Fait d'être un canard selon la démarche et le caquetage pour 100 cas (en pourcentages)

Est un canard	Marche et caquette comme un canard ? Oui	Non
Oui	60	40
Non	40	60
Total	100	100
(N)	(50)	(50)

$\chi^2 = 4{,}00$; $p < 0{,}05$

Tableau 6.7c.
Fait d'être un canard selon la démarche et le caquetage pour 500 cas (en pourcentages)

Est un canard	Marche et caquette comme un canard ? Oui	Non
Oui	60	40
Non	40	60
Total	100	100
(N)	(250)	(250)

$\chi^2 = 20{,}00$; $p < 0{,}001$

Attention donc aux prétentions de certains chercheurs selon lesquelles telle ou telle relation est « statistiquement significative » lorsque le N est très grand. Cette prétention peut être justifiée mais la relation peut être faible au point d'être triviale. Il faut garder à l'esprit que la signification statistique signifie simplement qu'une relation mise à jour dans des données d'échantillon peut être généralisée avec confiance à la population entière. C'est une information très utile à propos de la relation, mais elle n'est pas suffisante pour dire de la relation qu'elle est importante ou intéressante.

6.5 Les tests de signification sur des données de population

Dans ce chapitre nous avons appliqué le test du chi-carré à des données d'échantillon en vue de déterminer si nous pouvions généraliser avec confiance à la population une relation découverte dans un échantillon. C'est la principale application des statistiques inférentielles comme le test du chi-carré. Nous faisons des tests de signification statistique pour décider si nous pouvons généraliser des relations mises au jour dans des données d'échantillon comme le General Social Survey américain à la population entière. Les tests de signification nous permettent d'éliminer, avec un certain niveau de confiance, la possibilité que nos résultats d'échantillon soient dus au hasard des procédures d'échantillonnage.

Néanmoins, vous verrez souvent des recherches où l'on utilise des tests de signification statistique même pour des données de population telles que les 50 nations les plus peuplées, les 50 États américains ou bien encore les 13 provinces et territoires canadiens. Nous n'avons pas besoin de faire des tests de signification pour généraliser des données de population. En effet, si nous avons des données pour toute la population, on connaît donc notre population et il n'est nul besoin de généraliser. Avec des données de population nous sommes certains à 100 % (c'est-à-dire p = 100) que la relation discernée dans les données est valable pour l'ensemble de la population. On fait toutefois des tests de signification pour vérifier la probabilité qu'une relation trouvée dans les données de population soit due au hasard.

Par exemple, si le test de chi-carré d'une relation bivariée retracée dans des données de population indique que la relation est statistiquement significative au seuil 0,05, on peut considérer que l'association entre les variables n'est pas due au hasard. Il est peu probable d'obtenir une relation bivariée lorsque les cas sont distribués aléatoirement dans les cellules. La relation n'est donc probablement pas uniquement due au hasard. Il y a vraisemblablement une raison expliquant la relation et en tant que chercheurs scientifiques nous devons expliquer pourquoi ces variables sont associées.

Vous devez toutefois savoir que votre instructeur peut avoir une autre opinion sur l'utilité des tests de signification pour des données de population et voudra peut-être en discuter en classe.

6.6 Résumé du chapitre 6

Voici ce que nous avons vu dans ce chapitre :

- Les tests de signification statistique nous renseignent sur la probabilité de trouver une relation aussi forte que celle que l'on a découverte dans l'échantillon quand il n'y a pas de relation dans la population de laquelle est tiré cet échantillon.
- Dans les tests de signification statistique, l'hypothèse nulle affirme généralement que, dans la population, il n'y a pas de relation entre les deux variables.
- Le seuil de signification 0,05 est celui qui est généralement utilisé dans les tests de signification statistique. Autrement dit, si la probabilité de trouver une relation dans un échantillon quand il n'y a pas de relation entre les deux variables dans la population est inférieure à 0,05, on rejette l'hypothèse nulle voulant qu'il n'y ait pas de relation.
- Rejeter une hypothèse nulle vraie est une erreur de type I. Ne pas rejeter une hypothèse nulle fausse est une erreur de type II.
- Le niveau de signification est la probabilité de commettre une erreur de type I.
- Le test de signification du chi-carré est pertinent pour l'analyse de tableau bivarié.
- Le test du chi-carré compare les fréquences observées avec les fréquences auxquelles on s'attendrait s'il n'y avait pas de relation entre les variables (les fréquences anticipées).
- Le test du chi-carré n'est pas approprié lorsque les fréquences anticipées sont inférieures à 5. Dans une telle situation, envisagez d'exclure ou de regrouper les valeurs ayant de faibles fréquences pour en augmenter les fréquences anticipées.
- La signification statistique dépend à la fois de l'intensité de la relation et de la taille de l'échantillon (le nombre de cas dans l'échantillon).
- Qu'une relation soit statistiquement significative ne signifie pas qu'elle soit importante ou intéressante. Même une relation faible peut être statistiquement significative si elle est basée sur un nombre suffisant de cas.
- Les tests de signification sont souvent utilisés pour les données de population en vue de mesurer la possibilité que la relation soit due au hasard ou à un processus aléatoire.

Principaux concepts et mesures

Termes et idées

test de signification statistique
hypothèse nulle
rejet de l'hypothèse nulle
échec du rejet de l'hypothèse nulle
seuil de signification statistique
probabilité
erreur de type I (erreur alpha)
seuil alpha
erreur de type II (erreur bêta)
chi-carré
test du chi-carré
fréquence observée
fréquence anticipée
degré de liberté (dl)

Symboles

H_0

χ^2

f_a

f_o

dl

$p < 0{,}05$; $p < 0{,}01$; et $p < 0{,}001$

Prob. = n.a.

n.s.

Formules

$$f_a = \left(\frac{\text{Total de la rangée}}{N}\right) \text{Total de la colonne}$$

$$\chi^2 = \Sigma \frac{(f_o - f_a)^2}{f_a}$$

$$dl = (r - 1)(c - 1)$$