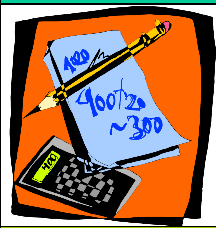
Approche quantitative



Le chi carré

Les objectifs pédagogiques

➡ Définir le test du chi carré
➡ Déterminer la nature des données propres au chi carré
➡ Savoir calculer le chi carré
➡ Savoir déterminer les fréquences théoriques
➡ Connaître les conditions d'application du chi carré
➡ Savoir formuler une hypothèse concernant le chi carré
➡ Savoir interpréter les résultats du chi carré

Le sommaire

A. Comment définit-on le test du chi carré ?
B. La nature des données propres au chi carré
C. L'écart entre les fréquences théoriques et les fréquences observées
D. L'écart entre deux échantillons tirés de la même population
E. Peut-on manipuler à volonté les fréquences théoriques?
F. Une distinction nécessaire
G. Comment calcule-t-on les degrés de liberté du chi carré ?
H. Comment formule-t-on une hypothèse concernant le chi carré ?
I. Quelles sont les exigences de base du chi carré ?
J. Comment interprète-t-on les résultats du chi carré ?
K. Un exemple complet de chi carré
L. Je fais mes exercices
M. Les sources bibliographiques

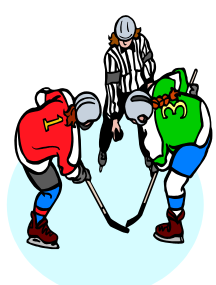
Le chi carré (*prononcé* « ki carré »)

A. Comment définit-on le test du chi carré ?

Le chi carré est un test statistique conçu pour déterminer si **la différence entre deux distributions de fréquences** est attribuable à l'erreur d'échantillonnage (le hasard) ou est suffisamment grande pour être statistiquement significative.

Si la différence entre les deux distributions est réduite, l'hypothèse nulle sera acceptée. Si la différence est grande, l'hypothèse nulle sera rejetée. Dans ce dernier cas, on parlera d'une différence statistiquement significative parce que l'écart entre les deux distributions est trop important pour être expliqué par le hasard seulement : une différence réelle existe donc.

Les données recueillies auprès d'un groupe sous forme de fréquences lors de votre recherche peuvent être perçues comme provenant d'un échantillon parmi une multitude d'autres échantillons. Si un nombre infini d'échantillons était ainsi formé pour comparer les échantillons un à un à une distribution théorique, il en résulterait des différences infinies qui pourraient être portées dans un tableau. Dans ce tableau, on constaterait que les grandes différences sont beaucoup moins nombreuses que les petites. Aussi la chance de trouver une grande différence entre deux distributions est-elle mince. **Une grande différence est attribuable à une source *autre* que l'échantillonnage. On parle alors d'une différence significative.**



Nous verrons plus loin qu'il est possible de comparer entre elles deux distributions observées, et non seulement une distribution observée à une distribution théorique. Dans ce cas, l'hypothèse nulle énoncera **qu'il n'y a pas de relation significative entre deux groupes** quant à une variable quelconque ou **qu'il n'y a pas de relation significative entre deux variables**. Ces deux façons de formuler une hypothèse détermine l'orientation que prendra l'interprétation des résultats.

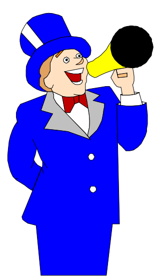
Bien qu'il soit un test non paramétrique, le test du chi carré permet de vérifier une hypothèse de recherche.

B. La nature des données propres au chi carré

Le chi carré est un test statistique non paramétrique qui convient à des fréquences, donc à des proportions, à des pourcentages et à des probabilités. Un phénomène quelconque peut être mesuré selon sa fréquence d'occurrence, sa durée, son intensité ou en fonction d'autres caractéristiques.

Par exemple, la participation des élèves d'une école à des activités parascolaires peut être mesurée selon le nombre d'élèves participants (fréquence), la proportion de filles par rapport au nombre de garçons, la satisfaction retirée de cette participation, le nombre moyen d'heures par semaine consacrées par chacun à ces activités, et bien d'autres variables encore.

La plupart des analyses statistiques connues portent sur des données pour lesquelles il est possible de calculer **une valeur de tendance centrale ou de variabilité**. Il en est ainsi notamment pour l'analyse de variance, la régression multiple et l'analyse factorielle. Peu de tests statistiques ont été conçus pour analyser des variables mesurant la fréquence d'occurrence d'un phénomène. À vrai dire, il n'en faut pas plusieurs non plus, car la fréquence, de par sa nature, impose des limites Le chi carré est justement ce test qui traite des fréquences.



Une variable comporte deux ou plusieurs catégories de réponses. Votre recherche terminée, vous désirez savoir si le nombre de cas à chacune des catégories dépasse ou non un critère quelconque. Vous avez même formulé une hypothèse qui prédit un nombre de cas par catégorie. À l'aide du chi carré vous calculez l'écart entre ce que vous avez obtenu et ce que vous souhaitiez obtenir.

Par ailleurs, de nouveaux tests statistiques puissants, dont la matière de base est la fréquence, font leur apparition. Cette famille d'analyses connue sous l'appellation ***Multiple Frequency Analysis*** gagne de la popularité tout simplement parce que de nombreuses situations en recherche requièrent ce type d'analyse.

Soyons pratiques, le chi carré s'applique surtout à des variables nominales et ordinales comportant un nombre restreint de catégories ou de niveaux.

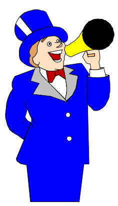
Une variable peut comprendre deux ou plusieurs valeurs. Le chi carré peut s'appliquer à une variable comportant plusieurs valeurs ou catégories. L'âge, par exemple, peut comprendre les catégories suivantes :

1	1 à 10 ans	4	31 à 40 ans
2	11 à 20 ans	5	41 à 50 ans
3	21 à 30 ans	6	51 à 60 ans

En admettant que vous désirez savoir s'il y a une relation entre le sexe des membres de votre échantillon et leur âge, vous obtiendrez alors le tableau croisé suivant :

Sexe	Catégories d'âge					
	1-10	11-20	21-30	31-40	41-50	51-60
Féminin						
Masculin						

Rien ne vous empêche d'utiliser le chi carré pour vérifier s'il y a relation ou non entre les deux variables même si l'une des variables comporte plusieurs catégories. Advenant un résultat significatif, ça devient onéreux et éprouvant d'interpréter ces résultats lorsque le nombre de cellules dans un tableau de contingence est assez élevé comme dans l'exemple précédent.



Le nombre de catégories de réponses qu'il est possible d'analyser à l'aide du chi carré peut être élevé. Cependant, même si, en pratique, le chi carré peut être exécuté sur une variable comportant de nombreuses catégories, les difficultés d'interprétation des résultats font qu'il est préférable de réduire le plus possible le nombre de catégories.

Pour en apprendre davantage sur la nature des variables, se rapporter au module intitulé **Les variables** sur le site Web suivant :

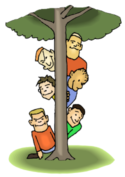
http://www3.umoncton.ca/cdem/longd/

Que le nombre de catégories soit petit ou grand, il reste que le chi carré compare le nombre de cas observés dans une catégorie au nombre de cas observés dans une autre catégorie ou au nombre de cas prévu. **Dans ce dernier cas**, il compare une distribution observée à une distribution théorique définie par une table de valeurs calculées. **Dans le premier cas**, il compare entre elles deux distributions observées.

Cette distinction est importante.

C. L'écart entre les fréquences théoriques et les fréquences observées

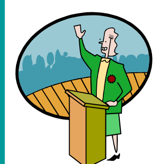
Le test du chi carré n'est pas vraiment différent des autres tests statistiques. On obtient une valeur statistique, appelée chi carré, qu'on rapporte à une table théorique de probabilité qui permet de décider si cette valeur se situe à l'intérieur ou à l'extérieur des limites fixées par le hasard de l'échantillonnage. Si la valeur du chi carré ne dépasse pas la valeur prévue l'hypothèse nulle est acceptée. Par contre, si cette valeur est plus grande, l'hypothèse nulle est rejetée.

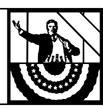
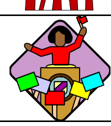


Vous voulez, par exemple, répondre à la **question de recherche** suivante : Le taux d'absentéisme des élèves est-il uniforme à chaque jour de la semaine ? L'**hypothèse nulle** peut être formulée de la manière suivante : *Le taux d'absentéisme chez les élèves sera le même à chaque jour de la semaine*.

Soyez alertés au fait suivant. Comme pour les autres tests statistiques, plus votre échantillon d'élèves sera nombreux, moins la différence entre le taux d'absentéisme entre chacun des jours de la semaine devra être grande pour rejeter l'hypothèse nulle. Pour s'en convaincre, il s'agit de scruter la table des probabilités théoriques du chi carré.

Dans une enquête téléphonique, avant la tenue d'élections provinciales, vous interrogez 330 personnes quant à leur préférence envers les trois candidats en lice. Rien ne vous indique au préalable qu'un candidat est avantagé par rapport aux autres. La probabilité théorique de chacun des candidats d'être vainqueur aux élections est donc de 33,3 %. Votre sondage terminé, vous établissez les fréquences observées. Le tableau suivant résume bien la situation.



Candidats	Fréquences théoriques	Fréquences observées	Différence
	f_t	f_o	$(f_t - f_o)$
	110	62	-48
	110	77	-33
	110	191	81

Tout porte à croire que la candidate (la dernière dans le tableau) l'emportera sur les deux autres candidats à la prochaine élection. À la fin du module, des exercices sont proposés, dont celui de calculer le chi carré à partir des données du tableau précédent. Ainsi, nous saurons (avant les électeurs…!) lequel d'entre eux deviendra député.

D. L'écart entre deux échantillons tirés de la même population

Au lieu de comparer des fréquences observées à des fréquences théoriques, nous pouvons les comparer à d'autres fréquences observées. Les statisticiens appellent cette analyse **test d'indépendance entre deux variables**.

Si le chi carré n'est pas significatif, les deux variables sont considérées comme indépendantes; s'il est significatif, les deux variables entretiennent une relation, une forme d'association. Néanmoins, cette dépendance ne doit pas être considérée comme une relation de causalité.

Pour exécuter le test du chi carré, les données doivent être incorporées à un tableau de contingence où les rangées sont identifiées par *r*, les colonnes par *c*, et les totaux des rangées et des colonnes comme les *fréquences marginales*. Une variable peut être associée aux rangées aussi bien qu'aux colonnes. Il n'y a pas véritablement de règles à ce sujet.

Supposons qu'une variable comporte 2 catégories de réponse et que l'autre variable en contient 4, votre tableau de contingence comprendra 8 cellules et vous en parlerez comme un tableau 2 x 4 ou 4 x 2. Le premier chiffre se rapporte au nombre de rangées, tandis que le deuxième se rapporte au nombre de colonnes (*r* x *c*).

? Comment calcule-t-on les fréquences théoriques d'un tableau de contingence ? ?

Les fréquences théoriques des cellules d'un tableau de contingence sont calculées selon les lois de la probabilité. Si les deux variables en question sont statistiquement indépendantes, il s'ensuit que la probabilité qu'elles existent ensemble dépend du produit de leur probabilité marginale propre. Alors, la fréquence théorique d'une cellule se calcule ainsi :

$$e_{ij} = R_i C_j / T$$

Légende: R_i représente le total de la rangée
C_j représente le total de la colonne
T représente le total global de toutes les cellules du tableau

	C_1	C_2	C_3	C_4	T
R_1	r_1c_1	r_1c_2	r_1c_3	r_1c_4	r_1
R_2	r_2c_1	r_2c_2	r_2c_3	r_2c_4	r_2
T_c	c_1	c_2	c_3	c_4	$r_{ij}c_{ij}$

rangée colonne fréquences marginales cellule

Appliquons ces nouvelles connaissances à des exemples concrets.

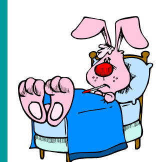
La direction d'une école veut savoir si le nombre d'élèves absents est plus élevé le vendredi que les autres jours de la semaine. La direction se demande aussi si les absences sont plus fréquentes chez les garçons que chez les filles. Pendant une semaine les absences ont été comptabilisées. Voici les résultats.

Journée	Garçons %	Filles %	Différence %
Lundi	10	19	-9
Mardi	8	11	-3
Mercredi	16	6	10
Jeudi	14	12	2
Vendredi	28	22	6

Comme nous le verrons plus loin, si le chi carré s'avère significatif, la direction d'école ne saura pas si le rejet de l'hypothèse nulle signifie qu'il y a une différence significative entre les garçons et les filles pour une ou plusieurs journées de la semaine quant au nombre d'absences. N'ayez crainte, les nouveaux tests modernes permettent des comparaisons fines de ce genre.

Les exemples employés jusqu'à présent sont incomplets et n'ont servi qu'à montrer un aspect particulier du chi carré. Prenons un autre exemple en y ajoutant, cette fois-ci, le calcul du chi carré.

L'animatrice d'une garderie s'inquiète du fait que plusieurs des enfants attrapent la grippe. Elle soupçonne, comme leurs parents d'ailleurs, que le taux de grippe à sa garderie est plus élevé que chez les enfants qui ne fréquentent aucune garderie. Elle va donc comparer les 32 enfants de sa garderie à 32 autres enfants qui demeurent à la maison. Voici les résultats.



État de l'enfant	Fréquences observées		
	Pas de grippe	Grippe	Total
Garderie	20	12	32
Maison	28	4	32
Total	48	16	64

État de l'enfant	Fréquences prévues		
	Pas de grippe	Grippe	Total
Garderie	24	8	32
Maison	24	8	32
Total	48	16	64

État de l'enfant	Proportions obtenues/prévues		
	Pas de grippe	Grippe	Total
Garderie	0,625/0,75	0,375/0,25	32/0,50
Maison	0,875/0,75	0,125/0,25	32/0,50
Total	48/0,75	16/0,25	64/1,00

Avant d'élaborer sur la formule statistique du chi carré voyons pourquoi on trouve les valeurs 24, 8 et 32 dans le tableau précédent des fréquences prévues.

Le nombre total d'enfants qui fréquentent la garderie est de 32, tout comme ceux qui sont élevés à la maison. En supposant qu'au cours des trois dernières années, le Ministère de la santé rapporte qu'en moyenne 1 enfant sur 4 attrape la grippe à chaque année, nous allons fixer les fréquences théoriques en conséquence. La deuxième partie du tableau précédent montre que 8 enfants sur 32 sont susceptibles d'attraper la grippe. Cette fréquence théorique attendue est aussi valable pour les enfants de la garderie que de la maison.

Rien ne nous empêche d'imposer des pourcentages ou des prévisions à notre guise pour, ensuite, comparer ces pourcentages théoriques aux pourcentages obtenus.

La formule qui permet de calculer le chi carré est la suivante :

$$X^2 = \Sigma \frac{((o_i - t_i) - .5)^2}{t_i} \quad \textit{(sommation } i=1 \longrightarrow k\textit{)}$$

Légende : o = fréquence observée
t = fréquence théorique

Remarque. À cause du degré de liberté unique, il est nécessaire d'apporter une correction pour la continuité (-0,5).

$X^2 =$	$\frac{((20-24)-,5)^2}{24}$	+	$\frac{((28-24)-,5)^2}{24}$	+	$\frac{((12-8)-,5)^2}{8}$	+	$\frac{((4-8)-,5)^2}{8}$	=

$X^2 =$	**0,84**	+	**0,51**	+	**0,44**	+	**2,53**	=4,32

$X^2 = 4,32$, dl = 1, $p < 0,05$ (approximativement $p = 0,04$).

D'après ces résultats, le taux de grippe à la garderie est significativement plus élevé chez les enfants de la garderie que chez les enfants de la maison.

Vous avez déjà compris que :
➡ plus la valeur du chi carré est élevée,
➡plus l'écart est grand entre les valeurs observées et les valeurs prévues ou théoriques,
➡ plus les chances sont grandes que le chi carré soit significatif
➡et plus l'hypothèse nulle risque d'être rejetée.

La plupart des statisticiens recommandent d'éviter d'utiliser le chi carré lorsque le nombre de fréquences est peu élevé. Hugues et Grawoig (1971) insistent pour qu'aucune cellule du tableau de contingence ne contienne moins de 5 fréquences et que le total des fréquences ne soit pas inférieur à 50.

E. Peut-on manipuler à volonté les fréquences théoriques ?

La plupart des logiciels modernes, tels que SPSS, offre un test de chi carré non paramétrique qui permet de préciser la fréquence théorique de chaque catégorie de la variable. Pour chacune de ces catégories vous pouvez indiquer la fréquence théorique désirée.

On demande à un psychologue industriel de déterminer si le nombre d'arrêts de travail dans une usine particulière au cours de la nuit est considérable ou non. Il n'a aucune donnée provenant d'une usine semblable. Par contre, il dispose d'amplement de données pour le groupe de travailleurs de jour. Il recueille donc les données de ces deux groupes couvrant la dernière semaine de travail. Le tableau suivant montre comment les données ont été juxtaposées en prévision de l'analyse du chi carré.

Catégories	Jour	Nuit
Lundi	11	13
Mardi	6	9
Mercredi	13	18
Jeudi	8	13
Vendredi	19	17

En fait, le tableau sert à montrer qu'il est possible d'appliquer les fréquences théoriques de votre choix. Évidemment, il faut justifier ce choix.

Il est évident que la formulation d'une hypothèse variera selon la façon de déterminer les fréquences théoriques : **et que dire des résultats !** Le tableau qui suit montre deux façons différentes de définir les fréquences théoriques qui serviront de critère de comparaison aux fréquences observées.

Dans le premier cas, on affirme que les fréquences observées devraient se répartir également entre les 3 catégories. On constate que les fréquences observées ne se répartissent pas également entre les catégories. Le chi carré pourrait être significatif. Dans le deuxième cas, le profil des fréquences observées suit d'assez près celui établi par les fréquences théoriques. Le chi carré pourrait ne pas être significatif. *Pourtant, il s'agit toujours de la même distribution de fréquences observées!*

F. observées	Fréq. théoriques 1	Fréq. théoriques 2
80	100	70
130	100	160
90	100	70

F. Une distinction nécessaire

Nous venons de voir que le chi carré s'applique dans deux situations différentes. Dans le premier cas, une distribution de fréquences observées est comparée à une distribution de fréquences théoriques. Le chi carré devient alors un test de validité de l'ajustement (*goodness-of-fit)*. La question posée est la suivante : *Les données obtenues diffèrent-elles ou non d'une distribution produite selon les lois de la probabilité ?*

Dans le deuxième cas, deux distributions de fréquences observées sont comparées entre elles. La question posée est la suivante : *Les données obtenues pour un groupe diffèrent-elles ou non des données obtenues pour un autre groupe ?* C'est dans ce cas qu'on voit apparaître le tableau de contingence (rangées x colonnes).

Test	Type	Données
Validité de l'ajustement	Unidimensionnel	théoriques ⇔ observées
Indépendance	Multidimensionnel	observées ⇔ observées

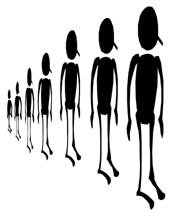
Les rangées du tableau de contingence peuvent être occupées par les catégories d'une variable aussi bien que par des groupes.

L'exemple portant sur le taux de grippe à la garderie établit une comparaison entre deux groupes (enfants de la garderie/enfants à la maison) quant à une variable (grippe/non grippe). **On compare alors le profil d'un groupe au profil d'un autre groupe quant à une variable.**

L'animatrice de la garderie aurait pu tout aussi bien séparer les enfants de sa garderie en groupes d'âge afin de savoir s'il existe une relation entre l'âge des enfants et le fait d'être affecté par la grippe ou non. Admettons que le nombre d'enfants à sa garderie soit, non pas 32, mais 183. Elle classe les enfants selon le niveau de grippe : 0=pas ou peu affectés; 1=modérément affectés; et 2=gravement affectés. Le tableau de contingence comprend alors 3 rangées et 3 colonnes (3 x 3).

La question posée n'est plus la même non plus : *Le nombre d'enfants affectés par divers niveaux de gravité de la grippe est-il le même à divers niveaux d'âge ?* **Ou bien** *: Existe-t-il une relation entre l'âge des enfants et leur état de santé ?*

Groupes d’âge	État de l’enfant		
	Peu ou pas affectés	Modérément affectés	Gravement affectés
1 à 2 ans			
3 à 4 ans			
5 à 6 ans			



Les enfants sont, en fait, regroupés selon la variable de l’âge. Vous remarquez que cette variable révèle une progression. Ils sont aussi classés selon une seconde variable, soit la gravité de la grippe, qui révèle, elle aussi, une progression. Ces deux variables discrètes peuvent être considérées au moins comme des variables ordinales, sinon à intervalles.

Par conséquent, **le test de chi carré s’applique à des variables dichotomiques et discrètes *ou à une combinaison des deux***. Cependant, selon la nature des variables étudiées, l’interprétation des résultats variera.

Le test de chi carré ne mesure pas la force d’association entre deux variables : cette fonction appartient à des tests tels que la corrélation. Le chi carré est plutôt un test d’indépendance qui indique si une relation est présente ou non entre deux variables : *Les enfants de 1 à 2 ans se distribuent-ils selon leur état de santé de la même façon que les enfants des autres groupes d’âge ?* Ou encore. *Le profil de l’état de la grippe est-il le même à chacun des groupes d’âge* ?

Bien que ces questions semblent traduire un degré d’association entre variables, on doit se rappeler que la méthode de calcul du chi carré n’est pas conçue pour mesurer la force ou la direction d’une relation entre deux variables. Par contre,

le chi carré mesure s’il existe une relation ou non entre les deux variables.

Lorsque le chi carré n’est pas significatif, les données d’une rangée du tableau de contingence sont équivalentes aux données des autres rangées. Donc, les variables du tableau n’interagissent pas entre elles et on peut les considérer comme indépendantes l’une de l’autre.

Lorsque le chi carré est significatif, les données d’une rangée varient de celles d’une autre rangée. En fait, les données ne sont pas réparties également entre les colonnes, mais dépendent plutôt de leur rangée d’appartenance. Il y a donc interdépendance entre les rangées et les colonnes, entre les variables. Il faudra d’autres tests pour déterminer la force de cette interdépendance. Dans les cas où il y a plus de deux rangées ou plus de deux colonnes, des analyses plus fines (a posteriori) aideront à préciser où l’interdépendance se manifeste.

	Aucune relation	
	X	**Y**
A	50	50
B	50	50

	Relation	
	X	**Y**
A	10	90
B	90	10

	Aucune relation	
	X	**Y**
A	10	90
B	10	90

Les tableaux précédents montrent des situations fictives où on constate une relation ou une absence de relation entre deux variables. Dans le dernier tableau en particulier, on voit que, même si les fréquences associées à la variable Y sont démesurément plus élevées que pour la variable X, le chi carré n'est pas significatif parce que le profil de la variable A ne diffère pas de celui de la variable B. La valeur du chi carré serait la même que pour le premier tableau à gauche.

G. Comment calcule-t-on les degrés de liberté du chi carré ?

Question. Comment fait-on pour calculer le nombre de degrés de liberté ?

Réponse. Nombre de catégories moins 1 (k - 1 = dl)

Pour un tableau de contingence (3 x 4), le nombre de degrés de liberté est donc de 6 (3 -1 x 4 -1). Pour un tableau de contingence (2 x 2), le nombre de degrés de liberté est de 1 (2 -1 x 2 -1).

H. Comment formule-t-on une hypothèse concernant le chi carré ?

Le chi carré n'est pas conçu pour vérifier des hypothèses directionnelles (bilatéralles). Il ne permet que d'établir si une fréquence observée est égale ou non à une autre fréquence théorique ou observée (unilatérale).

Voici comment peuvent être formulées l'hypothèse nulle et l'hypothèse alternative :

H_0 : Pour chacune des catégories, les fréquences observées sont équivalentes à celles d'une distribution théorique ou d'une autre distribution observée.

H_1 : La distribution des fréquences observées parmi les catégories diffère de la distribution des fréquences prévues ou des fréquences observées d'une autre variable.

En se référant à la recherche concernant les arrêts de travail à chaque jour de la semaine des deux groupes de travailleurs d'usine, voici comment les hypothèses pourraient être formulées :

H_0 : Pour chaque jour de la semaine, le nombre d'arrêts de travail des travailleurs de jour sera équivalent à celui des travailleurs de nuit à l'usine ABC.

H_1 : Le nombre d'arrêts de travail des travailleurs de jour à l'usine ABC diffère significativement de celui des travailleurs de nuit selon le jour de la semaine.

I. Quelles sont les exigences de base du chi carré ?

Tout au long du présent document certaines conditions d'application ont été abordées. En s'inspirant de Shavelson (1996), nous allons énumérer les 5 conditions d'application qui doivent être respectées pour que les résultats du chi carré soient valides. Les statisticiens en ajoutent une autre. Elle porte sur le nombre de fréquences minimales dans les catégories. Nous ajouterons donc une 6e condition d'application.

	Conditions d'application
1	Chaque donnée doit être insérée dans une seule catégorie. Les catégories doivent donc être mutuellement **exclusives**.
2	Les données doivent être **indépendantes** l'une de l'autre.
3	Les données doivent être des **fréquences**.
4	La fréquence **théorique** ou prévue de chaque catégorie ne doit pas être inférieure à 5 pour un degré de liberté de 2 ou plus et non inférieure à 10 si le degré de liberté est de 1 (dl = 1).
5	Le chi carré ne comportant qu'un seul degré de liberté (dl = 1) doit être corrigé au regard de la **continuité** pour accéder à la table des valeurs critiques du chi carré.
6	Dans un tableau de contingence, pas plus de **20 %** des cellules du tableau doivent contenir moins de 5 fréquences. **Évitez d'appliquer le chi carré si une cellule ne contient aucune fréquence.**

J. Comment interprète-t-on les résultats du chi carré ?

Réponse : Cela dépend. Mais de quoi, au juste ? Nous avons vu que l'hypothèse de recherche varie selon que les données observées sont comparées à une distribution théorique **ou** à une distribution d'autres données observées.

Le premier exemple qui suit compare des données observées à une distribution théorique.

Données	Catégories		
	A	B	C
Observées	9	36	30
Théoriques	25	25	25

Interprétation : *Il existe une différence significative entre les catégories de réponses quant à leur nombre respectif de cas*. Ou bien simplement. *Le nombre de cas varie selon la catégorie*.

Le deuxième exemple porte sur la comparaison de deux distributions de données observées.

Groupes	**Catégories**		
	A	B	C
Groupe 1	9	36	30
Groupe 2	14	12	26

Interprétation : *Le nombre de cas dans chacune des catégories diffère selon le groupe d'appartenance*. Ou bien. *Le groupe 1 diffère du groupe 2 quant à leur nombre dans chacune des catégories*.

Variable A	**Variable B**		
	1	2	3
1	56	23	11
2	32	16	42

Interprétation : *Il existe une relation entre les variables A et B*. Ou bien. *Le nombre de cas dans chacune des catégories de la variable B varie en fonction des catégories de la variable A*.

K. Un exemple complet de chi carré

Un sondage a été réalisé auprès des adultes de 21 ans ou plus concernant la venue d'un casino dans la région. Plusieurs questions de recherche ont été formulées, entre autres la suivante :

Question de recherche

Les femmes souhaitent-elles autant que les hommes
la venue d'un casino dans la région ?

Hypothèse de recherche

Les femmes souhaitent autant que les hommes la venue d'un casino.

Données de recherche

Sexe	Fréquences observées / *théoriques*		Total
	Pour le casino	Contre le casino	
Femmes	197/*164*	131/*164*	328
Hommes	120/*169*	218/*169*	338
Total	317	349	666

Résultat du chi carré

La valeur du chi carré s'élève à 21,36. Le nombre de degrés de liberté est de 1. On constate en consultant la table du chi carré qu'il fallait atteindre au moins la valeur de 10,83 pour que le chi carré soit significatif à 0,001. De toute évidence, la valeur obtenue dépasse largement cette limite.

Interprétation du chi carré

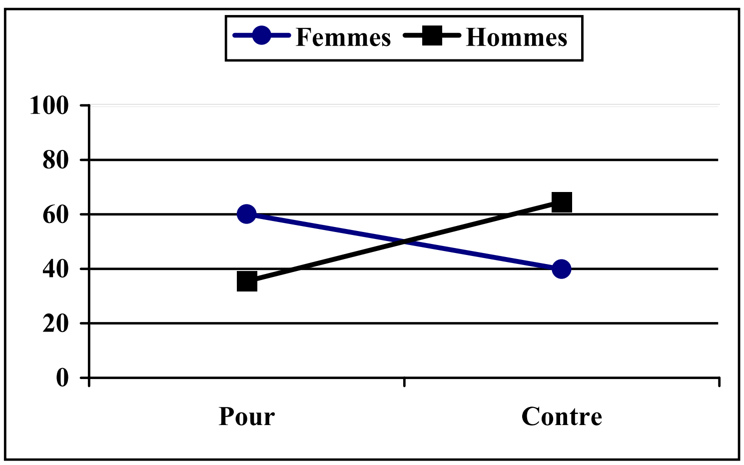
Voici diverses façons d'interpréter les résultats.

Les femmes sont significativement plus nombreuses que les hommes à souhaiter la venue d'un casino dans la région.

Les femmes, plus que les hommes, sont en faveur de la venue d'un casino dans la région.

Les hommes, plus que les femmes, s'opposent à la venue d'un casino dans la région.

Figure 1. Le pourcentage de femmes et d'hommes qui se disent pour ou contre la venue d'un casino dans la région.



L. Je fais mes exercices

1. Le chi carré est un test non paramétrique.

   1. Vrai 2. Faux

2. Le chi carré permet-il de préciser des fréquences théoriques autres que celles données par la loi des probabilités ?

   1. Oui 2. Non

3. Peut-on appliquer le test du chi carré à une variable nominale ?

   1. Oui 2. Non

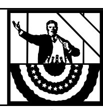
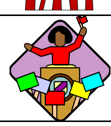
4. Pour un tableau de contingence 2 x 2, le nombre de degrés de liberté est :

   1. 2dl 2. 1dl 3. 4dl

5. Est-il possible de recoder une variable continue pour en faire une variable discrète (ordinale ou nominale) afin d'appliquer le test du chi carré ?

   1. Oui 2. Non

6. Appliquez le test du chi carré aux données suivantes tirées de la page 5 du présent document.

Candidats	Fréquences théoriques	Fréquences observées	Différence
	f_t	f_o	$(f_t - f_o)$
	110	62	-48
	110	77	-33
	110	191	81

7. À partir des données du tableau suivant, interprétez les résultats. Les garçons et les filles d'une école secondaire ont été comparés quant au port du jeans à une journée particulière de la semaine.

Sexe	Port du jeans	
	Oui	Non
Garçons	142	23
Filles	51	103

M. Sources

Corner, Rebecca C. et Thomas R. Corner. (1994). Chi-square analysis. *Science Teacher v61 n4 p44-47*. ERIC # : EJ484210.

Cox, Myron K. et Coretta H. Key. (1993). Post-hoc pair-wise comparisons for the chi-square test of homogeneity of proportions. *Educational and Psychological Measurement v53 n4 p951-962. ERIC #* : EJ485787.

Dayhaw, Laurence T. (1963). *Manuel de statistique. 2e édition*. Éditions de l'Université d'Ottawa, Canada.

Delucchi, Kevin, L. (1981). The use and misuse of chi-square: Lewis and Burke revisited. *Paper presented at the annual meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981)*. ERIC # : ED204399.

Hughes, Ann et Dennis Grawoig. (1971). *Statistics : A Foundation for Analysis.* Addison-Wesley Publishing Company. Don Mills, Ontario.

Klugh, Henry E. (1974). *Statistics: The Essentials for Research. 2nd Edition.* John Wiley and Sons, Inc. Toronto.

Shavelson, Richard J. (1996). *Statistical reasoning for the behavioural sciences*. 3e edition. Allyn and Bacon, Toronto.

Siegel, S. (1956). *Nonparametic statistics for the behavioural sciences*. New York, McGraw-Hill.

Yang, Shouu-Chyuan (1985). The single sample chi-square test: lesson plan. *Hawai University, Manoa. Western Curriculum Coordination Center.* ERIC # : ED263967.