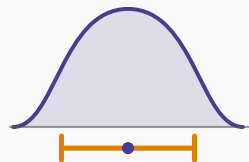


Statistique inférentielle

Chapitre 19 : Les tests du khi-deux



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

Le principe commun

Le test d'adéquation

Le test d'indépendance

La carte des tests

Ce qu'il faut retenir

Le principe commun

Le dernier type de données

Salaire, ancienneté : des grandeurs, sur lesquelles on calcule des moyennes.

Diplôme, région, sexe : des **catégories**, sur lesquelles on ne peut que compter. Aucun test précédent ne s'applique.

La statistique du khi-deux

$$\chi^2 = \sum_{\text{cases}} \frac{(\text{observé} - \text{attendu})^2}{\text{attendu}}$$

Deux tests, une seule formule : ce qui change est le calcul de l'**attendu**.

Un écart n'a pas le même poids partout

Dix de plus que prévu sur un effectif attendu de 20 : considérable.

Dix de plus sur un effectif attendu de 10 000 : négligeable.

La division ramène tous les écarts à une échelle commune. Sans elle, les grosses cases écraseraient tout.

Trois conséquences immédiates

1. La statistique est **positive** — c'est une somme de carrés.
2. Le test est **toujours unilatéral à droite** : seuls les grands écarts contredisent H_0 .
3. Elle suit une loi du **khi-deux**, dont le degré de liberté dépend du test.

Le test d'adéquation

Une distribution suit-elle une loi donnée

H_0 : la distribution observée est conforme à la loi théorique.

Effectif attendu de la modalité i : $n \times p_i$, où p_i vient de la loi supposée.

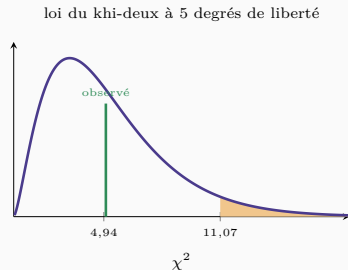
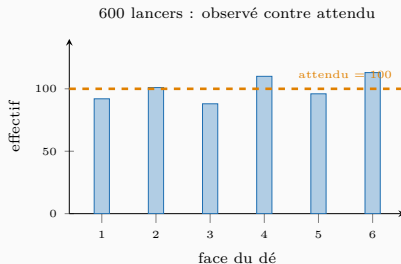
Degrés de liberté : $k - 1$, où k est le nombre de modalités.

D'où vient le -1

La règle du chapitre 4 : un degré de liberté perdu par contrainte imposée aux données.

Ici une seule : la somme des effectifs attendus doit évaluer n . Si la loi théorique comportait des paramètres estimés sur l'échantillon, il faudrait en retrancher autant.

Le dé du cours de probabilités



$$\chi^2 = \sum_i \frac{(\text{observé}_i - \text{attendu}_i)^2}{\text{attendu}_i} = 4,94 \quad \text{seuil à 5 \% : } 11,07 \quad \text{on ne rejette pas.}$$

Les écarts sont mis au carré, puis rapportés à l'effectif attendu. Diviser par l'attendu est essentiel : un écart de 10 sur un effectif de 100 n'a pas le même poids qu'un écart de 10 sur un effectif de 10 000.

Le test est **toujours unilatéral à droite** : seuls les grands écarts contredisent H_0 . Un χ^2 anormalement *petit* ne prouve pas l'adéquation — il signale plutôt des données trop belles pour être vraies.

Six cents lancers

H_0 : le dé est équilibré, donc $p_i = 1/6$ et l'effectif attendu vaut 100 par face.

$$\chi^2 = 4,94 \quad \text{ddl} = 5 \quad \text{seuil} = 11,07$$

$4,94 < 11,07$: on ne rejette pas. Rien ne permet d'affirmer que le dé est truqué.

Un usage professionnel du même test

Vérifier qu'un nombre de pannes suit une loi de Poisson, qu'une file d'attente est bien exponentielle, qu'une série de rendements est normale.

Chaque fois qu'un modèle **suppose** une loi, ce test est le moyen de contrôler la supposition — au lieu de la laisser implicite.

Le test d'indépendance

Deux caractères sont-ils liés

H_0 : les deux variables sont indépendantes.

Effectif attendu : $\frac{n_{i\bullet} \times n_{\bullet j}}{n}$.

Degrés de liberté : $(l - 1)(c - 1)$, pour l lignes et c colonnes.

L'attendu vient tout droit des probabilités

Sous indépendance, $p_{ij} = p_{i\bullet} \times p_{\bullet j}$ – c'est la définition du chapitre 12 des probabilités.

En multipliant par n , on obtient exactement la formule ci-dessus. **Le test ne fait que confronter cette prédiction aux effectifs observés.**

effectifs observés, et *entre parenthèses* les effectifs attendus sous indépendance

	Femmes	Hommes	Total
Bac	94 (94,2)	125 (124,8)	219
Bac+2	73 (74,8)	101 (99,2)	174
Bac+5	48 (46,0)	59 (61,0)	107
Total	215	285	500

L'effectif attendu se calcule par les marges

$$n_{i\bullet} \times n_{\bullet j} / n \quad \text{ex.} \quad \frac{219 \times 215}{500} = 94,2$$

C'est exactement la règle $p_{ij} = p_{i\bullet} p_{\bullet j}$ du chapitre 12 des probabilités, écrite en effectifs.

Le résultat

$$\chi^2 = 0,229 \quad \text{ddl} = (3 - 1)(2 - 1) = 2 \quad \text{seuil } 5,99$$

$p = 0,89$: les observés collent aux attendus de façon presque parfaite. On ne rejette pas l'indépendance.

Ce qu'il faut en conclure

Le diplôme et le sexe se distribuent indépendamment dans cette entreprise — ce qui rend d'autant plus notable que les salaires, eux, dépendent fortement du diplôme.

Le calcul

$$\chi^2 = 0,229 \quad \text{ddl} = (3 - 1)(2 - 1) = 2 \quad \text{seuil} = 5,99$$

$p = 0,89$: les effectifs observés collent aux attendus de façon remarquable. On ne rejette pas l'indépendance.

Ce que cela apprend, et ce que cela n'apprend pas

Dans cette entreprise, diplôme et sexe se distribuent indépendamment : il n'y a pas de sur-représentation d'un sexe dans un niveau de formation.

Cela ne dit rien des salaires — mais le rapproche du chapitre 17 est instructif : ni le sexe seul, ni sa liaison au diplôme ne structurent la rémunération. **Le diplôme, lui, la structure massivement.**

La condition de validité

Tout effectif **attendu** doit être supérieur à 5.

Sinon, on regroupe des modalités voisines — ou l'on emploie un test exact. La condition porte sur l'attendu, jamais sur l'observé.

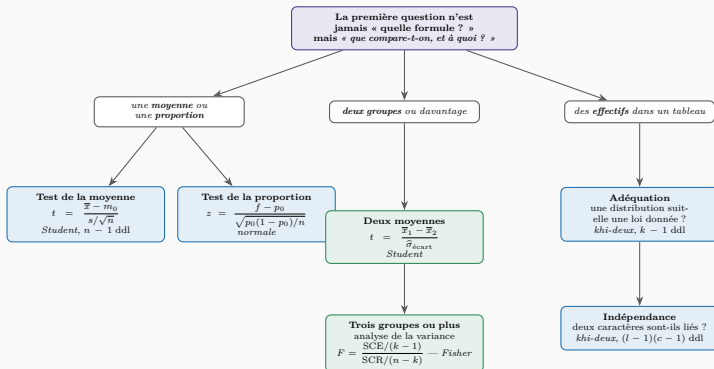
Deux limites à connaître

Le khi-deux détecte une **liaison**, pas une **causalité** : c'est la mise en garde du cours de descriptive, inchangée.

Et il est **sensible** à n : sur cent mille observations, la moindre irrégularité devient significative. Il faut alors accompagner le test d'un indicateur d'intensité de la liaison.

La carte des tests

Les six tests du cours, sur une seule page



Les cinq étapes sont les mêmes pour les six tests. Seule change la statistique — et donc la loi qu'elle suit.

1. écrire H_0 et H_1 2. choisir α et la forme (uni ou bilatérale) 3. calculer la statistique 4. lire le seuil dans la table 5. conclure en une phrase de gestion.

Devant un énoncé inconnu, ne cherchez pas la formule : identifiez d'abord la nature de ce qui est comparé. La formule suit toujours, la carte s'en charge.

Trois questions, dans cet ordre

1. **Que compare-t-on ?** Un paramètre à une valeur, deux groupes, plusieurs groupes, ou des effectifs.
2. **De quelle nature sont les données ?** Quantitatives ou catégorielles.
3. **Les observations vont-elles deux par deux ?** Si oui, il faut appairer.

Le réflexe à prendre pour l'examen

Devant un énoncé inconnu, ne cherchez pas la formule : identifiez d'abord **la nature de ce qui est comparé**.

Les cinq étapes sont ensuite les mêmes pour les six tests. La formule suit toujours ; c'est l'identification qui se rate.

Ce qu'il faut retenir

Six points

1. Le khi-deux traite les **effectifs de catégories**, là où aucun autre test ne s'applique.
2. $\chi^2 = \sum (\text{observé} - \text{attendu})^2 / \text{attendu}$; diviser par l'attendu met les écarts à la même échelle.
3. Le test est **toujours unilatéral à droite**.
4. **Adéquation** : attendu = $n p_i$, ddl = $k - 1$. Sur le dé : $\chi^2 = 4,94$, on ne rejette pas.
5. **Indépendance** : attendu = $n_{i\bullet} n_{\bullet j} / n$, ddl = $(l - 1)(c - 1)$. Sur nos données : $\chi^2 = 0,229$, on ne rejette pas.
6. Condition : tout effectif **attendu** supérieur à 5.

La suite

Le cours a traité une moyenne, deux groupes, plusieurs groupes, un tableau. Il reste la situation la plus riche : **deux variables quantitatives**.

La statistique descriptive s'y arrêta à une droite d'ajustement. La dernière partie en fait un modèle, et ouvre l'économétrie.