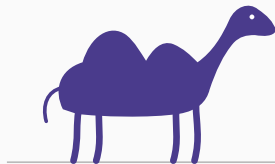


Introduction à l'analyse des données

Chapitre 13 : Du khi-deux à l'analyse des correspondances



Jaouad Madkour

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

Un autre type de tableau

L'indépendance, et l'écart à l'indépendance

L'inertie d'un tableau croisé

Les profils

Ce qu'il faut retenir

Un autre type de tableau

Le tableau de contingence

La situation

Deux variables **qualitatives** croisées : chaque case porte un **effectif**, non une mesure.

L'ACP ne s'applique pas : ni moyenne, ni variance, ni corrélation sur des catégories.

Le tableau du chapitre : type de client et canal d'achat

	Magasin	Web	Téléphone	Total
Jeunes	30	60	10	100
Familles	80	40	30	150
Seniors	90	20	40	150
Total	200	120	80	400

Le tableau de contingence, en image

observés	Prix	Choix	Proximité	Service	total
Étudiant	14	2	5	1	22
Employé	45	20	26	10	101
Cadre	8	40	8	16	72
Indépendant	12	19	12	32	75
Retraité	5	1	17	7	30
total	84	82	68	66	300

théoriques	Prix	Choix	Proximité	Service
Étudiant	6,16	6,01	4,99	4,84
Employé	28,28	27,61	22,89	22,22
Cadre	20,16	19,68	16,32	15,84
Indépendant	21,00	20,50	17,00	16,50
Retraité	8,40	8,20	6,80	6,60

Le khi-deux du chapitre 19 de l'inférence.

$$\chi^2 = \sum \frac{(n_{ij} - t_{ij})^2}{t_{ij}} = 110,41$$

Avec 12 degrés de liberté, le seuil à 5 % vaut 21,03.

L'indépendance est rejetée sans hésitation.

Et ensuite ? Le test dit qu'il y a une liaison. **Il ne dit pas laquelle.**

Cinq lignes, quatre colonnes, vingt écarts à commenter — l'AFC va les mettre sur un plan.

L'indépendance, et l'écart à
l'indépendance

Sous indépendance

Si les deux variables étaient sans lien, chaque case vaudrait

$$\text{attendu} = \frac{\text{total ligne} \times \text{total colonne}}{\text{total général}}$$

Exemple

Effectif observé, et attendu entre parenthèses :

	Magasin	Web	Téléphone
Jeunes	30 (50)	60 (30)	10 (20)
Familles	80 (75)	40 (45)	30 (30)
Seniors	90 (75)	20 (45)	40 (30)

Les jeunes achètent **deux fois plus** sur le web qu'attendu, les seniors **deux fois moins** : c'est là qu'est l'information.

La mesure de l'écart

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

Le calcul

$$\chi^2 = \frac{400}{50} + \frac{900}{30} + \frac{100}{20} + \frac{25}{75} + \frac{25}{45} + 0 + \frac{225}{75} + \frac{625}{45} + \frac{100}{30}$$

$$\chi^2 \approx \mathbf{64,1}$$

pour $(3 - 1)(3 - 1) = 4$ degrés de liberté.

Application économique

Le chapitre 19 de *Statistique inférentielle* comparait ce 64,1 à un seuil, et concluait au **rejet de l'indépendance**.

Conclusion en un mot : *il y a un lien*. Elle ne dit **pas lequel** — ni quelles catégories s'attirent, ni selon quelles dimensions.

L'analyse des correspondances reprend exactement le même nombre et le décompose.

L'inertie d'un tableau croisé

Du khi-deux à l'inertie

Le phi-deux

$$\Phi^2 = \frac{\chi^2}{n}$$

C'est le khi-deux **rapporté à l'effectif total** — et c'est l'**inertie totale** du tableau.

Sur notre tableau

$$\Phi^2 = \frac{64,1}{400} = \mathbf{0,160}$$

Le pont avec tout ce qui précède

Le chapitre 7 avait défini l'inertie comme la dispersion d'un nuage autour de son centre.

Ici, le « centre » est la situation d'**indépendance**, et l'inertie mesure de combien le tableau s'en écarte.

Un tableau parfaitement indépendant aurait une inertie **nulle** : rien à représenter.

La démarche est donc identique à celle de l'ACP — décomposer une inertie totale sur des axes hiérarchisés — appliquée à un écart à l'indépendance plutôt qu'à une dispersion de mesures.

La décomposition

Les profils

Comparer des lignes de tailles différentes

Profil-ligne

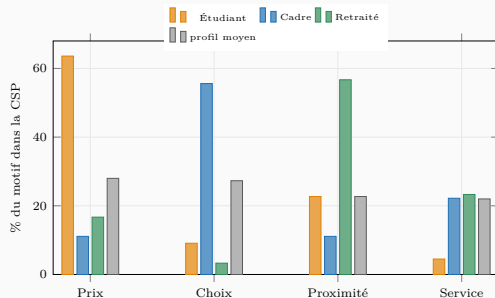
Le **profil** d'une ligne est sa répartition en pourcentages :

$$\text{profil}_j = \frac{n_{ij}}{\text{total de la ligne } i}$$

Les trois profils-lignes

	Magasin	Web	Téléphone
Jeunes	30 %	60 %	10 %
Familles	53 %	27 %	20 %
Seniors	60 %	13 %	27 %
Ensemble	50 %	30 %	20 %

Les trois profils, en image



L'AFC compare des profils, pas des effectifs.

Il y a 22 étudiants et 101 employés : les comparer en effectifs n'aurait aucun sens.

En **pourcentages de ligne**, la comparaison redevient possible — et l'écart au **profil moyen** devient l'objet de l'analyse.

Un profil ligne égal au profil moyen n'apporte rien : il ira au centre du plan factoriel. C'est exactement ce que dit l'indépendance.

Intuition

Les jeunes sont 100 et les seniors 150 : comparer les effectifs bruts confondrait la **taille** du groupe et son **comportement**.

Les profils suppriment cet effet : les jeunes achètent sur le web dans 60 % des cas contre 30 % en moyenne, les seniors dans 13 %.

Le nuage de l'AFC est un nuage de **profils**, non d'effectifs.

Définition

On peut construire le nuage des **profils-lignes** ou celui des **profils-colonnes**.

Les deux donnent **les mêmes valeurs propres** et deux représentations qui se superposent. C'est la **dualité** du chapitre 6, portée à sa forme la plus complète — et c'est ce qui permettra, au chapitre 14, de lire lignes et colonnes sur une seule figure.

Ce qu'il faut retenir

Cinq points

1. Un tableau de **contingence** croise deux variables qualitatives; l'ACP ne s'y applique pas.
2. Le **khi-deux** mesure l'écart à l'indépendance : ici $\chi^2 \approx 64,1$ pour 4 degrés de liberté.
3. L'inertie totale est $\Phi^2 = \chi^2/n = 64,1/400 = \mathbf{0,160}$.
4. L'AFC **décompose** cette inertie sur des axes, là où le test se contentait de conclure.
5. Le nuage est fait de **profils**, ce qui neutralise les différences de taille.

La suite

L'inertie est décomposée. Reste à lire la figure — et les règles de lecture ne sont pas tout à fait celles de l'ACP.