

Plaquette 5 - Problèmes de synthèse sur des données réelles

Tableaux croisés — Seconde

Corrections détaillées

Boîte à outils

Analyser le lien entre deux variables qualitatives.

- Pour savoir si B dépend de A , on compare $f_A(B)$ et $f_{\bar{A}}(B)$ (ou $f(B)$).
- Si $f_{A_1}(B) = f_{A_2}(B) = f(B)$ pour tous les groupes, les données ne montrent **aucun lien**.
- Rapport des fréquences conditionnelles (« risque relatif ») :

$$\frac{f_A(B)}{f_{\bar{A}}(B)}$$

- **Paradoxe de Simpson** : un groupe peut être meilleur dans chaque sous-groupe mais moins bon au total, quand les groupes ne se répartissent pas de la même façon entre les sous-groupes. On compare alors à **sous-groupe égal**.
- Effectifs bruts : « la plupart des B sont A » ne dit rien si A est majoritaire dans la population.
- **Lien n'est pas cause** : une troisième variable peut expliquer le lien observé.

Correction de l'exercice 1 - Admissions et paradoxe de Simpson

a) Taux d'admission, $f = \frac{\text{admis}}{\text{candidats}}$, dans chaque filière :

- filière A : hommes $\frac{360}{600} = 60\%$, femmes $\frac{70}{100} = 70\%$;
- filière B : hommes $\frac{10}{100} = 10\%$, femmes $\frac{90}{600} = 15\%$.

Dans **chaque** filière, les femmes sont mieux admises.

b) Globalement :

$$f_{\text{hommes}} = \frac{370}{700} \approx 52,9\% \quad f_{\text{femmes}} = \frac{160}{700} \approx 22,9\%$$

c) Les femmes ont surtout candidaté dans la filière B, très sélective, et les hommes dans la filière A. Le total mélange des filières de difficultés différentes : c'est le **paradoxe de Simpson**. La comparaison juste se fait filière par filière.

Réponse. a) A : 60 % / 70 % ; B : 10 % / 15 % · b) $\approx 52,9\%$ / $\approx 22,9\%$ · c) répartition différente entre filières.

Correction de l'exercice 2 - Deux traitements des calculs rénaux

a) Taux de succès $f = \frac{\text{succès}}{\text{patients}}$:

Calculs	Traitement A	Traitement B
Petits	$\approx 93,1\%$	$\approx 86,7\%$
Gros	$\approx 73,0\%$	$\approx 68,8\%$
Total	78 %	$\approx 82,6\%$

b) Le traitement A est meilleur **pour les petits calculs et pour les gros**, mais moins bon au total. Explication : A a été donné surtout aux patients à gros calculs (263 sur 350), les plus difficiles à soigner, alors que B a surtout traité des petits calculs. Le total mélange des patients de gravités différentes. À gravité égale, le traitement A est le meilleur.

Réponse. A est meilleur dans chaque groupe (93,1 % et 73 %) ; le total (78 % contre 82,6 %) est trompeur.

Correction de l'exercice 3 - La ceinture de sécurité

a) Dans chaque groupe :

$$f_{\text{attachés}}(\text{décès}) = \frac{18}{1\,800} = 1\% \quad f_{\text{non attachés}}(\text{décès}) = \frac{20}{200} = 10\%$$

Le taux de décès est dix fois plus élevé chez les occupants non attachés.

b) Parmi les 38 décès : $f_{\text{décès}}(\text{attaché}) = \frac{18}{38} \approx 47,4\%$.

Cela ne remet pas en cause la ceinture : presque tout le monde est attaché (90 % des occupants), il est donc normal que les attachés soient nombreux parmi les victimes. La comparaison pertinente est celle de la question a).

Réponse. a) 1 % contre 10 % · b) environ 47,4 % ; non.

Correction de l'exercice 4 - Sport et tabac

a) On compare des fréquences conditionnelles :

$$f_{\text{sportifs}}(\text{fumeur}) = \frac{30}{200} = 15\% \quad f_{\text{non sportifs}}(\text{fumeur}) = \frac{90}{300} = 30\%$$

Les non-sportifs fument deux fois plus souvent (rapport $\frac{30}{15} = 2$).

b) Non. Le tableau montre un **lien** entre les deux variables, pas une **cause**. Il est possible que fumer décourage le sport, ou qu'une troisième variable (âge, santé, mode de vie) agisse sur les deux. Pour établir une cause, il faudrait une expérience contrôlée.

Réponse. a) 15 % contre 30 % · b) Non : lien n'est pas cause.

Correction de l'exercice 5 - Aucun lien

a) $f_{\text{filles}}(\text{lunettes}) = \frac{30}{120} = 25\%$, $f_{\text{garçons}}(\text{lunettes}) = \frac{20}{80} = 25\%$, et $f(\text{lunettes}) = \frac{50}{200} = 25\%$.

b) Parmi les 50 porteurs de lunettes :

$$f_{\text{lunettes}}(\text{filles}) = \frac{30}{50} = 60\% \quad f(\text{filles}) = \frac{120}{200} = 60\%$$

Le port de lunettes est aussi fréquent chez les filles que chez les garçons, et les filles sont aussi représentées parmi les porteurs de lunettes que dans la classe. Ces données ne montrent **aucun lien** entre le sexe et le port de lunettes. (On parlera en terminale d'indépendance.)

Réponse. a) 25 %, 25 %, 25 % · b) 60 % et 60 % : aucun lien.

Correction de l'exercice 6 - Dix ans de pratique sportive

a) Taux de pratique $f = \frac{\text{sportifs}}{\text{effectif}}$:

Année	Jeunes	Adultes	Total
2015	55 %	30 %	40 %
2025	60 %	≈ 38,6 %	45 %

b) Écarts en points :

$$\text{jeunes: } 60 - 55 = +5 \quad \text{adultes: } 38,6 - 30 \approx +8,6 \quad \text{total: } +5$$

La pratique progresse dans les deux groupes, surtout chez les adultes. Le total progresse moins que chez les adultes car la part des adultes (qui pratiquent moins) dans l'échantillon a augmenté : 70 % en 2025 contre 60 % en 2015.

Réponse. Jeunes : 55 → 60 % ; adultes : 30 → 38,6 % ; total : 40 → 45 %.

Correction de l'exercice 7 - Vote et type de commune

a) Score de A dans chaque ligne, $f_{\text{commune}}(A) = \frac{\text{voix A}}{\text{électeurs}}$:

— urbaine : $\frac{330}{600} = 55\%$; périurbaine : $\frac{120}{300} = 40\%$; rurale : $\frac{30}{100} = 30\%$;

— total : $\frac{480}{1000} = 48\%$.

A gagne en ville mais perd l'élection : son avance urbaine ne compense pas son retard ailleurs.

b) Référence : les 480 électeurs de A.

$$f_A(\text{urbaine}) = \frac{330}{480} = 0,6875$$

Près de 69 % des électeurs de A sont urbains, alors que les urbains forment 60 % de l'échantillon.

Réponse. a) 55 %, 40 %, 30 % ; total 48 % · b) 68,75 %.

Correction de l'exercice 8 - Travail de nuit et sommeil

a) Fréquences conditionnelles :

$$f_{\text{nuit}}(\text{troubles}) = \frac{60}{150} = 40\% \quad f_{\text{jour}}(\text{troubles}) = \frac{170}{850} = 20\%$$

Rapport :

$$\frac{0,40}{0,20} = 2$$

Les troubles sont deux fois plus fréquents chez les travailleurs de nuit.

b) Référence : les 230 salariés qui souffrent de troubles : $\frac{60}{230} \approx 26,1\%$.

Les travailleurs de nuit ne sont qu'un quart des personnes atteintes, parce qu'ils sont peu nombreux (15 % des salariés) ; cela n'empêche pas le travail de nuit de doubler la fréquence des troubles.

Réponse. a) 40 % contre 20 %, rapport 2 · b) environ 26,1 %.

Correction de l'exercice 9 - Accidents près de chez soi

a) Référence : les 50 incidents.

$$f_{\text{incidents}}(\text{courts}) = \frac{40}{50} = 80\%$$

b) On compare le risque **dans chaque type de trajet** :

$$f_{\text{courts}}(\text{incident}) = \frac{40}{800} = 5\% \quad f_{\text{longs}}(\text{incident}) = \frac{10}{200} = 5\%$$

Le risque par trajet est le même. Si 80 % des incidents ont lieu sur des trajets courts, c'est simplement que 80 % des trajets sont courts. La phrase « la plupart des accidents ont lieu près de chez soi » ne prouve donc pas que conduire près de chez soi est plus dangereux.

Réponse. a) 80 % · b) Non : 5 % dans les deux cas.

Correction de l'exercice 10 - Les « ne sait pas »

a) Deux populations de référence possibles :

$$f(\text{oui}) = \frac{420}{1\,000} = 42\% \quad f_{\text{avis exprimé}}(\text{oui}) = \frac{420}{800} = 52,5\%$$

b) Les deux disent vrai, mais pas sur la même population : 52,5 % des personnes qui ont un avis sont favorables ; 42 % des personnes interrogées le sont. Le partisan choisit la référence qui l'avantage ; une présentation honnête donne les trois proportions (42 % oui, 38 % non, 20 % sans avis). De plus, il s'agit d'un échantillon, pas de tous les Français.

Réponse. a) 42 % ; 52,5 % · b) chacun a raison avec sa population de référence.

Correction de l'exercice 11 - Engrais et type de sol

a) Taux $f = \frac{\text{bons rendements}}{\text{parcelles}}$:

— sol argileux : avec engrais $\frac{30}{40} = 75\%$, sans $\frac{12}{20} = 60\%$;

— sol sableux : avec $\frac{8}{20} = 40\%$, sans $\frac{12}{40} = 30\%$;

— total : avec $\frac{38}{60} \approx 63,3\%$, sans $\frac{24}{60} = 40\%$.

b) L'engrais améliore le rendement sur chaque sol : +15 points sur sol argileux, +10 points sur sol sableux. L'écart total (environ +23 points) est plus grand que ces deux écarts, car l'engrais a été utilisé surtout sur le sol argileux, le plus fertile : c'est l'écart à **sol égal** qui mesure l'effet de l'engrais.

Réponse. b) Oui : +15 points (argileux), +10 points (sableux).

Correction de l'exercice 12 - Heures de conduite et réussite

a) Taux de réussite $f = \frac{\text{reçus}}{\text{candidats}}$:

$$\frac{105}{150} = 70\% \quad \frac{150}{250} = 60\% \quad \frac{45}{100} = 45\%$$

Le taux de réussite baisse quand le nombre d'heures augmente. Taux global :

$$\frac{300}{500} = 60\%.$$

b) Non. Les candidats qui prennent beaucoup d'heures sont souvent ceux qui ont le plus de difficultés au départ : c'est leur niveau initial, et non les heures de conduite, qui explique leur taux de réussite plus faible. Le tableau montre un lien, pas une cause ; pour conclure, il faudrait comparer des candidats de même niveau de départ.

Réponse. a) 70 %, 60 %, 45 % · b) Non : lien n'est pas cause.

Correction de l'exercice 13 - Atelier et bureau

a) Marges : atelier $0,4 \times 250 = 100$, bureau 150 ; femmes $0,3 \times 250 = 75$, hommes 175. Femmes en atelier : $0,2 \times 75 = 15$. Par différences : femmes en bureau 60, hommes en atelier 85, hommes en bureau 90.

	Femmes	Hommes	Total
Atelier	15	85	100
Bureau	60	90	150
Total	75	175	250

b) Dans chaque ligne :

$$f_{\text{atelier}}(\text{femmes}) = \frac{15}{100} = 15 \% \quad f_{\text{bureau}}(\text{femmes}) = \frac{60}{150} = 40 \%$$

Les femmes sont nettement moins présentes en atelier (15 %) qu'en bureau (40 %), pour une moyenne de 30 %.

Réponse. b) 15 % en atelier contre 40 % en bureau.

Correction de l'exercice 14 - Langues vivantes de deux classes

a) Dans chaque colonne, on divise par l'effectif de la classe :

LVB	Classe A	Classe B
Espagnol	60 %	$\approx 66,7 \%$
Allemand	30 %	$\approx 16,7 \%$
Italien	10 %	$\approx 16,7 \%$

L'allemand est presque deux fois plus choisi en classe A ; l'espagnol domine dans les deux classes.

b) Référence : les 15 germanistes.

$$f_{\text{allemand}}(\text{classe A}) = \frac{9}{15} = 60 \%$$

alors que la classe A ne représente que $\frac{30}{66} \approx 45 \%$ des élèves.

Réponse. b) 60 % des germanistes sont en classe A.

Correction de l'exercice 15 - Voitures électriques

a) Avec $n = f \times n_{\text{client}}$:

- particuliers : $0,15 \times 1\,200 = 180$; $0,35 \times 1\,200 = 420$; reste 600 ;
- entreprises : $0,30 \times 800 = 240$; $0,40 \times 800 = 320$; reste 240.

	Électrique	Hybride	Thermique	Total
Particuliers	180	420	600	1 200
Entreprises	240	320	240	800
Total	420	740	840	2 000

b) $\frac{420}{2\,000} = 21 \%$.

c) Référence : les 420 électriques.

$$f_{\text{élec}}(\text{entreprises}) = \frac{240}{420} \approx 0,571$$

Les entreprises, qui font 40 % des ventes, achètent environ 57 % des électriques.

Réponse. b) 21 % · c) environ 57,1 %.