

Approfondissement

Tableaux croisés et fréquences conditionnelles

Pour aller plus loin : paradoxe de Simpson, effectifs attendus, sur-représentation, tableau reconstitué, lois de De Morgan

5 exercices · niveau ●●● · durée conseillée : 2 h 30 · calculatrice autorisée

Exercice 1 Un paradoxe célèbre (hors programme — pour aller plus loin) ●●●

Une université propose deux filières, A et B. Les tableaux donnent, pour chaque filière, le nombre de candidats et le nombre d'admis.

Filière A	Admis	Refusés	Total
Hommes	240	160	400
Femmes	70	30	100

Filière B	Admis	Refusés	Total
Hommes	20	80	100
Femmes	120	280	400

- 1.1 Dans chaque filière, calculer la fréquence des admis parmi les hommes, puis parmi les femmes. Qui est le mieux admis dans la filière A ? dans la filière B ?
- 1.2 Construire le tableau croisé de l'ensemble des 1 000 candidatures (hommes / femmes en lignes, admis / refusés en colonnes), puis calculer la fréquence des admis parmi les hommes et parmi les femmes sur cet ensemble.
- 1.3 Ce résultat contredit-il celui de la question 1.1 ? Expliquer le phénomène en comparant la fréquence d'admission de chaque filière et la répartition des candidatures.
- 1.4 Peut-on additionner deux fréquences conditionnelles calculées sur des groupes d'effectifs différents ? Illustrer la réponse à l'aide de cet exemple.

Exercice 2 Effectifs attendus s'il n'y a aucun lien ●●●

On a interrogé 1 000 personnes sur leur temps quotidien passé devant un écran et sur le port de lunettes.

Effectifs observés	Porte des lunettes	Sans lunettes	Total
Plus de 4 h par jour	245	355	600
4 h par jour ou moins	105	295	400
Total	350	650	1 000

On dit que les deux caractères sont *sans lien* lorsque, dans chaque case, l'effectif est égal à $\frac{n(\text{ligne}) \times n(\text{colonne})}{N}$: on l'appelle l'**effectif attendu**.

- 2.1 Justifier cette formule : montrer que si la fréquence des porteurs de lunettes est la même dans chaque ligne que dans la population entière, alors l'effectif de la case vaut bien $\frac{n(\text{ligne}) \times n(\text{colonne})}{N}$.
- 2.2 Calculer les quatre effectifs attendus, puis les écarts avec les effectifs observés. Que remarque-t-on sur ces quatre écarts ?
- 2.3 Calculer la fréquence des porteurs de lunettes dans chacune des deux lignes, ainsi que la fréquence marginale. Conclure sur l'existence d'un lien et sur son sens.
- 2.4 Ce tableau prouve-t-il que le temps d'écran abîme la vue ? Proposer une autre explication de l'écart observé.

Exercice 3 Sur-représentation



Une enquête porte sur 4 000 actifs d'une région, classés selon leur catégorie professionnelle et selon la possession d'un diplôme de l'enseignement supérieur.

Effectifs	Diplômé	Non diplômé	Total
Cadres	640	160	800
Employés	500	1 500	2 000
Ouvriers	60	1 140	1 200
Total	1 200	2 800	4 000

Pour deux groupes A et B , on appelle **indice de sur-représentation** de A dans B le quotient $I = \frac{f(A \text{ parmi les } B)}{f(A)}$, où $f(A)$ est la fréquence marginale de A .

- 3.1 Calculer la fréquence des diplômés dans chacune des trois catégories, puis leur fréquence marginale.
- 3.2 Calculer l'indice de sur-représentation des diplômés dans chaque catégorie. Interpréter les valeurs obtenues selon qu'elles sont supérieures ou inférieures à 1.
- 3.3 Démontrer que, pour tous groupes A et B d'effectifs non nuls, $\frac{f(A \text{ parmi les } B)}{f(A)} = \frac{f(B \text{ parmi les } A)}{f(B)}$. (On exprimera chaque membre en fonction de $n(A \text{ et } B)$, $n(A)$, $n(B)$ et N .)
- 3.4 Vérifier ce résultat sur les cadres, en calculant la fréquence des cadres parmi les diplômés et la fréquence marginale des cadres.
- 3.5 Que vaut cet indice lorsque les deux caractères sont sans lien, au sens de l'exercice 2 ?

Exercice 4 Reconstituer un tableau à partir de trois fréquences



Dans un lycée de 1 200 élèves, 55 % des élèves sont des filles. Parmi les filles, 30 % sont internes ; par ailleurs, 66 % des internes sont des filles.

- 4.1 Déterminer le nombre de filles et le nombre de garçons, puis le nombre de filles internes.
- 4.2 En déduire l'effectif total des internes, puis construire le tableau croisé complet.
- 4.3 Calculer la fréquence des internes parmi les garçons et la comparer à celle des internes parmi les filles, puis à la fréquence marginale des internes.
- 4.4 Si l'on ne connaissait pas l'effectif total 1 200, pourrait-on encore déterminer toutes les fréquences conditionnelles ? Et tous les effectifs ? Justifier.

Exercice 5 « Et », « ou », « ni » : les lois de De Morgan



On reprend l'exercice 2. Deux listes de booléens de longueur 1 000 sont construites : `ecran[i]` vaut True si la personne i passe plus de 4 h par jour devant un écran, `lun[i]` si elle porte des lunettes. On note A le groupe « plus de 4 h d'écran », B celui des porteurs de lunettes, $N = 1 000$.

- 5.1 Que renvoie `compte(ecran, lun, "et")` ? Et `compte(ecran, lun, "ou")` ?
- 5.2 Compléter la ligne en pointillés pour que le mode "ni" compte les personnes qui ne sont dans aucun des deux groupes, et donner la valeur renvoyée.
- 5.3 Démontrer sur un tableau croisé que $n(\text{ni } A \text{ ni } B) = N - n(A \text{ ou } B)$, puis que « non (A ou B) » et « (non A) et (non B) » sont le même groupe.
- 5.4 Traduire de même le groupe « non (A et B) » à l'aide de « non A » et « non B », et vérifier la relation sur les effectifs du tableau de l'exercice 2.

```

1 def compte(a, b, mode):
2     n = 0
3     for i in range(len(a)):
4         if mode == "et" and a[i] and b[i]:
5             n = n + 1
6         if mode == "ou" and (a[i] or b[i]):
7             n = n + 1
8         if mode == "ni" and .....:
9             n = n + 1
10    return n

```

Programme 1 — Comptage selon trois modes