

Tableaux croisés

Deux variables qualitatives

Effectifs ■ Fréquences ■ Conditionnement

		Variable B	
Variable A		42	58
		18	32
		parmi ?	quelle proportion ?

Seconde générale ■ Mathématiques ■ Programme officiel

 Cours complet

 15 exercices

 Problème défi

 Corrigés

Table des matières

1	 Pourquoi croiser deux variables ?	3
2	 Variables qualitatives et modalités	3
2.1	Nominale ou ordinale	3
2.2	Un individu fournit un couple	4
3	 Construire et lire un tableau croisé	4
3.1	Cases intérieures et marges	4
3.2	Le langage des ensembles	5
4	 Fréquences globales et marginales	5
4.1	Fréquence d'une case	5
4.2	Fréquence marginale	6
5	 Fréquences conditionnelles	6
5.1	Le dénominateur est le groupe après « parmi »	6
5.2	L'ordre change la question	7
5.3	Tableaux de fréquences conditionnelles	7
5.4	Reconstruire une fréquence globale	8
6	 Compléter un tableau croisé	8
6.1	À partir des marges	8
6.2	À partir d'une fréquence conditionnelle	9
7	 Représenter et comparer les groupes	9
7.1	Barres empilées à 100 %	9
7.2	Choisir la bonne représentation	10
8	 Comparer et interpréter sans se tromper	10
8.1	Écart en points de pourcentage	10
8.2	Association et causalité	10
9	 Filtrer des données avec la logique et Python	11
9.1	ET, OU, NON	11
9.2	Construire le tableau à partir de listes	11
10	 Du tableau statistique au modèle probabiliste	12
11	 Pour aller plus loin - Accessible à tous	12
11.1	Indépendance observée	12
11.2	Le test du produit en croix pour un tableau 2×2	13
11.3	Le paradoxe de Simpson	13
11.4	Standardiser pour comparer avec les mêmes poids	14
12	 Boîte à outils et erreurs à éviter	15
13	 Exercices progressifs	16
14	 Problème défi : évaluer un programme de tutorat	19
15	 Corrigés détaillés	21
16	 Corrigé du problème défi	27
17	 Fiche-mémoire	30

1 ? Pourquoi croiser deux variables ?

Une statistique à une variable répond à une question comme : « quelle proportion d'élèves arrive à l'heure ? ». Croiser deux variables permet d'aller plus loin : cette proportion est-elle la même selon le mode de transport ? Le tableau croisé sert à organiser les couples d'informations observés sur les mêmes individus.

Intuition | Un tableau qui répond à plusieurs questions

Chaque case intérieure décrit une intersection : des individus vérifient simultanément deux critères. Les marges décrivent les groupes entiers. Changer de dénominateur change la question posée.

Exemple | Une enquête de ponctualité

On interroge 160 élèves. Pour chacun, on relève son mode de transport et s'il est arrivé à l'heure. Le croisement permettra de comparer les taux de ponctualité entre les modes de transport.

Attention | Observer n'est pas expliquer

Si deux variables semblent liées, le tableau établit une **association**, pas nécessairement une cause. La distance au lycée, l'horaire ou d'autres facteurs peuvent influencer simultanément le transport et la ponctualité.

2 ➡ Variables qualitatives et modalités

2.1 Nominale ou ordinale

Définition | Variable qualitative

Une variable est **qualitative** lorsque ses valeurs sont des catégories appelées **modalités**.

- Elle est **nominale** si les catégories n'ont pas d'ordre naturel : couleur, moyen de transport.
- Elle est **ordinale** si les catégories peuvent être ordonnées : faible, moyen, élevé.

Exemple | Reconnaître la nature

Le mode de transport est qualitatif nominal. Un degré de satisfaction « insatisfait, neutre, satisfait » est qualitatif ordinal. L'âge exact est quantitatif ; une tranche d'âge est qualitative ordinale.

Attention | Les nombres peuvent être des étiquettes

Un numéro de département ou un code postal n'est pas une grandeur mesurable : calculer sa moyenne n'a pas de sens. Ce sont des catégories, même si elles sont écrites avec des chiffres.

2.2 Un individu fournit un couple

Si l'on observe deux variables X et Y , chaque individu fournit un couple

$$(X(\text{individu}), Y(\text{individu})).$$

Par exemple, un élève peut fournir le couple (transports en commun, à l'heure).

Méthode | Avant de construire le tableau

1. identifie la population et l'effectif total N ;
2. nomme les deux variables ;
3. liste leurs modalités sans chevauchement ;
4. vérifie que chaque individu appartient à une seule modalité de chaque variable.

3 Construire et lire un tableau croisé

3.1 Cases intérieures et marges

Définition | Tableau croisé d'effectifs

Un tableau croisé répartit les individus selon deux variables qualitatives.

- Une **case intérieure** contient l'effectif d'une combinaison de deux modalités.
- Les **effectifs marginaux** sont les totaux des lignes et des colonnes.
- La case en bas à droite contient l'effectif total N .

Exemple | Transport et ponctualité

	Actif	Transports	Motorisé	Total
À l'heure	42	58	24	124
En retard	8	12	16	36
Total	50	70	40	160

« 42 » signifie : 42 élèves utilisent un transport actif **et** arrivent à l'heure.

Propriété | Retrouver les marges

Chaque total de ligne est la somme des cases de sa ligne ; chaque total de colonne est la somme des cases de sa colonne. Par exemple,

$$42 + 58 + 24 = 124, \quad 42 + 8 = 50.$$

 Démonstration | Pourquoi les deux sommes des marges valent-elles N ?

Chaque individu est compté une fois dans exactement une modalité de la première variable. La somme des totaux de lignes compte donc tous les individus une fois. Le même raisonnement s'applique aux colonnes :

$$\sum_i n_{i\bullet} = N \quad \text{et} \quad \sum_j n_{\bullet j} = N.$$

3.2 Le langage des ensembles

Notons A l'ensemble des élèves utilisant un transport actif et H l'ensemble des élèves arrivés à l'heure. Alors

$$\text{Card}(A) = 50, \quad \text{Card}(H) = 124, \quad \text{Card}(A \cap H) = 42.$$

 Intuition | Le mot « et »

Une case intérieure correspond toujours à une intersection. « Actif et à l'heure » signifie $A \cap H$. Le mot « ou » correspondrait à une réunion, qui n'est généralement pas une seule case du tableau.

 Propriété | Cardinal d'une réunion

Pour deux ensembles A et B ,

$$\text{Card}(A \cup B) = \text{Card}(A) + \text{Card}(B) - \text{Card}(A \cap B).$$

 Démonstration | Pourquoi soustraire l'intersection ?

En additionnant $\text{Card}(A)$ et $\text{Card}(B)$, les individus de $A \cap B$ sont comptés deux fois. On retire donc une fois leur effectif pour que chaque individu de $A \cup B$ soit compté exactement une fois.

4 % Fréquences globales et marginales

4.1 Fréquence d'une case

 Définition | Fréquence globale

La fréquence globale d'une case est son effectif divisé par l'effectif total :

$$f(A \cap B) = \frac{\text{Card}(A \cap B)}{N}.$$

Dans l'exemple,

$$f(A \cap H) = \frac{42}{160} = 0,2625 = \boxed{26,25 \%}.$$

Cela signifie que 26,25 % de tous les élèves sont à la fois utilisateurs d'un transport actif et ponctuels.

4.2 Fréquence marginale

Définition | Fréquence marginale

La fréquence marginale d'une modalité est son effectif marginal divisé par N :

$$f(A) = \frac{\text{Card}(A)}{N}.$$

Ainsi,

$$f(A) = \frac{50}{160} = 31,25 \%, \quad f(H) = \frac{124}{160} = 77,5 \%.$$

Démonstration | Une marge est la somme de ses cases

Les catégories de ponctualité « à l'heure » et « en retard » partagent l'ensemble A sans chevauchement. Donc

$$f(A) = f(A \cap H) + f(A \cap \overline{H}) = \frac{42}{160} + \frac{8}{160} = \frac{50}{160}.$$

Méthode | Contrôles indispensables

- La somme de toutes les fréquences intérieures vaut 1.
- La somme des fréquences marginales d'une même variable vaut 1.
- Une fréquence doit toujours être comprise entre 0 et 1.

5 Fréquences conditionnelles

5.1 Le dénominateur est le groupe après « parmi »

Définition | Fréquence de B parmi A

Si A n'est pas vide, la fréquence de B **parmi** les individus de A est

$$f_A(B) = \frac{\text{Card}(A \cap B)}{\text{Card}(A)}.$$

Le groupe A est appelé groupe de référence ou groupe conditionnant.

Exemple | Ponctualité parmi les transports actifs

Parmi les 50 élèves utilisant un transport actif, 42 arrivent à l'heure :

$$f_A(H) = \frac{42}{50} = 0,84 = \boxed{84 \%}.$$

Le dénominateur est 50, et non 160, car la question commence par « parmi les utilisateurs d'un transport actif ».

Intuition | Zoomer sur une sous-population

Calculer $f_A(B)$ revient à oublier momentanément tous les individus qui ne sont pas dans A . Le groupe A devient alors le nouveau « tout », donc son effectif devient le dénominateur.

5.2 L'ordre change la question

Attention | $f_A(B)$ et $f_B(A)$ sont différents

$$f_A(B) = \frac{\text{Card}(A \cap B)}{\text{Card}(A)}, \quad f_B(A) = \frac{\text{Card}(A \cap B)}{\text{Card}(B)}.$$

Le numérateur est le même, mais les dénominateurs sont généralement différents.

Exemple | Deux questions, deux réponses

$$f_A(H) = \frac{42}{50} = 84\%, \quad f_H(A) = \frac{42}{124} \approx 33,9\%.$$

84 % des utilisateurs d'un transport actif sont ponctuels, mais seulement 33,9 % des élèves ponctuels utilisent un transport actif. Ces phrases ne se contredisent pas.

Méthode | Traduire une phrase

1. repère le groupe introduit par « parmi », « chez » ou « sachant que » ;
2. place son effectif au dénominateur ;
3. place l'intersection des deux critères au numérateur ;
4. reformule la réponse avec le même ordre que la question.

5.3 Tableaux de fréquences conditionnelles

Si l'on conditionne par le mode de transport, chaque colonne devient un groupe de référence et totalise 100 %.

	Actif	Transports	Motorisé
À l'heure	$42/50 = 84\%$	$58/70 \approx 82,9\%$	$24/40 = 60\%$
En retard	$8/50 = 16\%$	$12/70 \approx 17,1\%$	$16/40 = 40\%$
Total	100 %	100 %	100 %

Attention | Ne pas mélanger les dénominateurs

Dans un même tableau conditionnel, les pourcentages d'une colonne doivent tous utiliser le même total de colonne. Additionner des pourcentages calculés sur des groupes de référence différents n'a pas de sens.

5.4 Reconstruire une fréquence globale

★ Théorème | Moyenne pondérée des fréquences conditionnelles

Si $A_1, \dots, A_k$ forment une partition de la population, alors

$$f(B) = \sum_{i=1}^k f(A_i) f_{A_i}(B).$$

Démonstration | Décomposer la population

Comme les groupes A_i sont disjoints et couvrent toute la population,

$$\text{Card}(B) = \sum_i \text{Card}(A_i \cap B).$$

Or

$$\text{Card}(A_i \cap B) = \text{Card}(A_i) f_{A_i}(B) = N f(A_i) f_{A_i}(B).$$

En divisant la première égalité par N , on obtient la formule.

Exemple | Retrouver la ponctualité globale

$$f(H) = \frac{50}{160} \times 0,84 + \frac{70}{160} \times \frac{58}{70} + \frac{40}{160} \times 0,60 = \frac{124}{160} = 77,5\%.$$

La fréquence globale est une moyenne pondérée : les groupes les plus nombreux pèsent davantage.

6 Compléter un tableau croisé

6.1 À partir des marges

Méthode | Raisonner ligne par ligne

Pour retrouver une case manquante, soustrais les cases connues au total de sa ligne ou de sa colonne. Termine par un contrôle sur l'effectif total.

Exemple | Une case manquante

	B	$\bar{B}$	Total
A	36	?	60
$\bar{A}$	24	16	40
Total	60	40	100

La case manquante vaut $60 - 36 = 24$. On vérifie aussi $24 + 16 = 40$ dans la deuxième colonne.

6.2 À partir d'une fréquence conditionnelle

✓ Propriété | Revenir à un effectif

Si $f_A(B) = p$, alors

$$\text{Card}(A \cap B) = p \times \text{Card}(A).$$

💡 Exemple | Utiliser 35 %

Parmi 240 personnes abonnées, 35 % choisissent l'option premium. L'effectif correspondant est

$$0,35 \times 240 = \boxed{84}.$$

Le pourcentage s'applique à 240, car les abonnés constituent le groupe de référence.

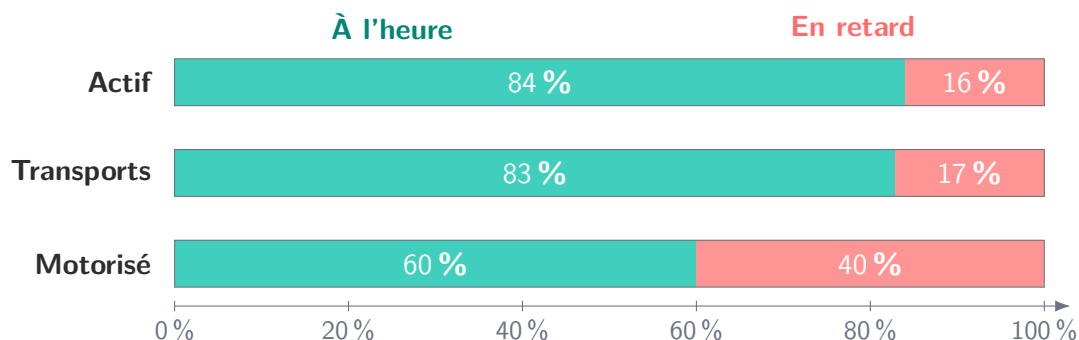
⚠ Attention | Un contrôle d'intégralité

Un effectif est un entier. Si un calcul donne 26,4 personnes, les données sont arrondies, incohérentes ou incomplètes. Il faut alors le signaler plutôt que masquer le problème.

7 📊 Représenter et comparer les groupes

7.1 Barres empilées à 100 %

Pour comparer des groupes de tailles différentes, on représente leurs fréquences conditionnelles sur des barres de même longueur.



🧠 Intuition | Comparer des proportions, pas des hauteurs brutes

La largeur verte répond directement à la question « quelle part du groupe est ponctuelle ? ». Des barres d'effectifs pourraient surtout refléter que certains groupes sont plus nombreux.

7.2 Choisir la bonne représentation

Méthode | Selon la question

- Pour comparer les **tailles** des groupes : effectifs ou fréquences globales.
- Pour comparer une modalité à **l'intérieur** des groupes : fréquences conditionnelles.
- Pour montrer les deux variables ensemble : tableau croisé, barres groupées ou empilées.

8 Comparer et interpréter sans se tromper

8.1 Écart en points de pourcentage

Dans l'exemple, l'écart de ponctualité entre les transports actifs et motorisés vaut

$$84 \% - 60 \% = \boxed{24 \text{ points de pourcentage}}.$$

Dire « 24 % de plus » serait ambigu : un écart de proportions se formule de préférence en points.

8.2 Association et causalité

Définition | Association observée

Deux variables sont **associées dans les données** lorsque les fréquences conditionnelles changent sensiblement d'un groupe à l'autre.

Attention | Une association ne prouve pas une cause

Le tableau ne permet pas d'affirmer que choisir un transport actif *cause* la ponctualité. Il manque des informations sur la distance, les horaires, la météo et la manière dont les données ont été recueillies.

Méthode | Une conclusion statistique honnête

1. nomme les groupes comparés ;
2. donne les deux fréquences conditionnelles ;
3. calcule leur écart ;
4. écris « dans cet échantillon » ou « dans ces données » ;
5. évite toute causalité non démontrée.

9 Filtrer des données avec la logique et Python

9.1 ET, OU, NON

✓ Propriété | Traduction ensembliste

$$\text{ET} \longleftrightarrow \cap, \quad \text{OU} \longleftrightarrow \cup, \quad \text{NON } A \longleftrightarrow \bar{A}.$$

Le « ou » mathématique est inclusif : $A \cup B$ contient aussi les individus qui vérifient les deux critères.

💡 Exemple | Filtrer une base

« Élèves en transports en commun ET en retard » sélectionne la case d'effectif 12. « Transport actif OU à l'heure » sélectionne $A \cup H$, d'effectif

$$50 + 124 - 42 = \boxed{132}.$$

9.2 Construire le tableau à partir de listes

```
1 transports = ["actif", "commun", "actif", "motorise"]
2 ponctuels = [True, True, False, False]
3
4 effectif = 0
5 for t, h in zip(transports, ponctuels):
6     if t == "actif" and h:
7         effectif = effectif + 1
8
9 print(effectif)
```

✂ Méthode | Lire le programme

La boucle examine chaque individu. La condition avec `and` traduit une intersection. Pour construire tout le tableau, on initialise un compteur par combinaison de modalités puis on incrémente exactement une case par individu.

⚠ Attention | Données manquantes

Une valeur vide ne doit pas être confondue avec une modalité réelle. Avant de compter, vérifie comment les absences de réponse sont codées et décide explicitement si tu les exclus ou si tu crées une modalité « non renseigné ».

10 Du tableau statistique au modèle probabiliste

Intuition | Un pont, pas encore tout le chapitre

Si l'on choisit au hasard un individu parmi les 160 élèves, chaque fréquence devient la probabilité d'un événement dans ce modèle d'équiprobabilité. La prochaine fiche étudiera systématiquement les probabilités conditionnelles et les arbres.

Avec $A =$ « transport actif » et $H =$ « à l'heure »,

$$P(A) = \frac{50}{160}, \quad P(A \cap H) = \frac{42}{160}, \quad P_A(H) = \frac{42}{50}.$$

La même fraction apparaît, mais son statut change : dans l'enquête, elle décrit une fréquence observée ; dans le modèle, elle devient une probabilité.

Attention | Le modèle doit être annoncé

Une probabilité n'est pas automatiquement une fréquence passée. On choisit un modèle en explicitant ses hypothèses, notamment la manière dont l'individu est sélectionné.

11 Pour aller plus loin - Accessible à tous

Intuition | Statut de cette partie

Le socle officiel est entièrement présenté dans les sections précédentes. Les idées suivantes permettent d'analyser des données comme dans des classes très exigeantes, mais chaque étape est démontrée sans prérequis caché.

11.1 Indépendance observée

Définition | Critère descriptif

Dans un tableau, on dit que les variables paraissent **indépendantes** lorsque la distribution d'une variable est la même dans chaque groupe de l'autre variable. Par exemple,

$$f_A(B) = f(B).$$

Il s'agit ici d'un constat sur les données, pas d'une vérité universelle sur la population.

 Exemple | Indépendance parfaite dans un tableau

	B	$\bar{B}$	Total
A	30	20	50
$\bar{A}$	30	20	50
Total	60	40	100

Dans chaque ligne, la fréquence de B vaut 60 %, comme dans toute la population.

11.2 Le test du produit en croix pour un tableau 2×2

Considérons le tableau

	B	$\bar{B}$
A	a	b
$\bar{A}$	c	d

 Théorème | Critère $ad = bc$

Si les quatre effectifs sont positifs, les deux lignes ont la même fréquence de B si et seulement si

$$\boxed{ad = bc}.$$

 Démonstration | Un simple produit en croix

Les fréquences de B dans les deux lignes sont $a/(a+b)$ et $c/(c+d)$. Alors

$$\begin{aligned} \frac{a}{a+b} &= \frac{c}{c+d} \text{ si et seulement si } a(c+d) = c(a+b) \\ &\text{si et seulement si } ac + ad = ac + bc \\ &\text{si et seulement si } \boxed{ad = bc}. \end{aligned}$$

11.3 Le paradoxe de Simpson

 Exemple | Un renversement surprenant

On compare la réussite avec ou sans tutorat dans deux niveaux initiaux.

	Avec tutorat	Sans tutorat
Niveau standard	18/20 = 90 %	80/100 = 80 %
Niveau fragile	60/100 = 60 %	5/10 = 50 %
Ensemble	78/120 = 65 %	85/110 $\approx$ 77,3 %

Le tutorat est associé à un meilleur taux dans chaque niveau, mais à un taux global plus faible. Ce renversement s'explique par la composition des groupes : le groupe tutoré contient beaucoup plus d'élèves initialement fragiles.

 Intuition | Pourquoi le global peut-il tromper ?

Chaque taux global est une moyenne pondérée des taux par niveau. Les poids ne sont pas les mêmes dans les deux groupes. Comparer les seuls taux globaux mélange donc l'effet du niveau initial et l'association avec le tutorat.

11.4 Standardiser pour comparer avec les mêmes poids

Si l'on donne artificiellement le même poids $1/2$ aux deux niveaux, les taux standardisés deviennent

$$\text{tutorat : } \frac{1}{2} \times 90 \% + \frac{1}{2} \times 60 \% = 75 \%,$$

$$\text{sans tutorat : } \frac{1}{2} \times 80 \% + \frac{1}{2} \times 50 \% = 65 \%.$$

Cette comparaison descriptive neutralise la différence de composition. Elle ne prouve toujours pas une causalité, mais elle évite une partie du piège.

12 Boîte à outils et erreurs à éviter

Méthode | Case intérieure

Intersection de deux modalités.

Méthode | Fréquence globale

$$n(A \cap B) / N.$$

Méthode | Fréquence marginale

$$n(A) / N.$$

Méthode | Fréquence conditionnelle

$$f_A(B) = n(A \cap B) / n(A).$$

Méthode | Lire « parmi »

Le groupe après « parmi » fournit le dénominateur.

Méthode | Compléter

Case manquante = total - cases connues.

Méthode | Comparer

Utilise des fréquences conditionnelles et des points de pourcentage.

Attention | Réflexes

Vérifie les marges, les dénominateurs, la somme 100 % et évite toute causalité automatique.

Intuition | La question qui sauve

Avant chaque pourcentage, demande-toi : « parmi quel groupe ? ». Si tu connais ce groupe, tu connais le dénominateur.

13 Exercices progressifs

Intuition | Mode d'emploi

Pour chaque pourcentage, écris d'abord le groupe de référence. Une fraction correctement nommée vaut mieux qu'un calcul rapide dont le dénominateur reste mystérieux.

Exercice 1 : Reconnaître les variables

On interroge 80 foyers sur leur type principal de chauffage, leur degré de satisfaction (*faible, moyen, élevé*) et leur facture mensuelle d'énergie.

Identifier la population, un individu et les trois variables. Pour chacune, préciser si elle est qualitative nominale, qualitative ordinale ou quantitative.

Exercice 2 : Compléter avec les marges

Compléter le tableau et expliquer chaque soustraction.

	B	$\bar{B}$	Total
A	36	?	60
$\bar{A}$	24	16	40
Total	60	40	100

Exercice 3 : Fréquences globales et marginales

Reprendre le tableau « transport et ponctualité » du cours.

1. Calculer la fréquence des transports actifs.
2. Calculer la fréquence des élèves à l'heure.
3. Calculer la fréquence des élèves à la fois actifs et en retard.
4. Calculer la fréquence des transports motorisés.

Exercice 4 : Le bon dénominateur

Toujours avec le tableau du cours, calculer et interpréter :

$$f_A(H), \quad f_H(A), \quad f_M(H),$$

où A signifie « transport actif », M « transport motorisé » et H « à l'heure ».

Exercice 5 : Traduire en français

Dans le tableau du cours, expliquer précisément ce que représentent les fractions

$$\frac{12}{70}, \quad \frac{12}{36}, \quad \frac{24}{160}.$$

Pour chaque fraction, dire s'il s'agit d'une fréquence conditionnelle ou globale.

Exercice 6 ★★☆☆ : Un tableau conditionnel

Construire le tableau des fréquences de ponctualité conditionnelles au mode de transport. Chaque colonne devra totaliser 100 %. Arrondir au dixième de pourcentage et commenter la principale différence.

Exercice 7 ★★☆☆ : Retrouver les effectifs

Une plateforme compte 600 clients, dont 240 abonnés. Parmi les abonnés, 35 % choisissent l'option premium. Au total, 150 clients ont choisi cette option.

1. Construire le tableau croisant abonnement et option premium.
2. Calculer la fréquence du premium parmi les non-abonnés.
3. Comparer les deux groupes en points de pourcentage.

Exercice 8 ★★☆☆ : Points ou pourcentage relatif ?

Dans le groupe X, 72 % des personnes répondent « oui », contre 60 % dans le groupe Y.

1. Calculer l'écart en points de pourcentage.
2. Calculer l'augmentation relative de 60 % à 72 %.
3. Expliquer pourquoi les réponses 12 % et 20 % ne décrivent pas la même comparaison.

Exercice 9 ★★☆☆ : ET, OU, NON

Parmi 160 élèves, 50 utilisent un transport actif, 124 arrivent à l'heure et 42 vérifient les deux critères.

1. Combien utilisent un transport actif **ou** arrivent à l'heure ?
2. Combien ne vérifient aucun des deux critères ?
3. Expliquer pourquoi $50 + 124$ ne répond pas directement à la première question.

Exercice 10 ★★☆☆ : Passer de listes à un tableau

On dispose des listes suivantes, dans lesquelles les éléments de même rang concernent la même personne.

```
1 groupe = ["A", "B", "A", "A", "B", "B", "A", "B"]
2 choix  = ["oui", "non", "oui", "non",
3          "oui", "oui", "non", "non"]
```

1. Construire le tableau croisé d'effectifs.
2. Calculer les fréquences de « oui » dans les deux groupes.
3. Les variables semblent-elles associées dans ces huit données ?

Exercice 11 ★★☆☆ : Une fréquence conditionnelle impose une case

Une enquête porte sur 400 personnes. Le groupe A contient 150 personnes et

$$f_A(B) = 28 \%$$

Au total, 100 personnes vérifient B.

Construire entièrement le tableau croisant $A/\bar{A}$ et $B/\bar{B}$.

« Toujours demander : parmi quel groupe ? »

Exercice 12 ★★☆☆ : Lire les performances d'un test

Sur 1 000 personnes, 100 sont malades. Le tableau d'observation est :

	Test positif	Test négatif	Total
Malade	90	10	100
Non malade	45	855	900
Total	135	865	1 000

1. Parmi les malades, quelle est la fréquence des tests positifs ?
2. Parmi les non-malades, quelle est la fréquence des tests négatifs ?
3. Parmi les tests positifs, quelle est la fréquence des personnes malades ?
4. Pourquoi les réponses aux questions 1 et 3 sont-elles différentes ?

Exercice 13 ★★☆☆ : Indépendance dans un tableau

On considère le tableau

	B	$\bar{B}$	Total
A	30	20	50
$\bar{A}$	x	40	$x + 40$

Déterminer x pour que la fréquence de B soit la même dans les deux lignes. Vérifier ensuite le critère $ad = bc$.

Exercice 14 ★★☆☆ : Une moyenne pondérée de proportions

Un premier site compte 200 personnes et un taux de réponse de 75 %. Un second site compte 300 personnes et un taux de réponse de 60 %.

Calculer le taux global de réponse de deux façons : avec les effectifs, puis avec une moyenne pondérée. Expliquer pourquoi la moyenne simple de 75 % et 60 % ne convient pas.

Exercice 15 ★★☆☆ : Retrouver une fréquence conditionnelle

La population est partagée entre A et $\bar{A}$. On sait que

$$f(B) = 0,54, \quad f(A) = 0,40, \quad f_A(B) = 0,75.$$

1. Montrer que

$$f(B) = f(A)f_A(B) + (1 - f(A))f_{\bar{A}}(B).$$

2. En déduire $f_{\bar{A}}(B)$.
3. Interpréter la différence entre les deux fréquences conditionnelles.

14 🧠 Problème défi : évaluer un programme de tutorat

🧠 Intuition | Objectif

Ce problème montre pourquoi un taux global peut donner une conclusion opposée aux comparaisons détaillées. Toutes les questions sont guidées : les calculs simples construisent progressivement l'explication.

Un lycée étudie la réussite de 230 élèves à une évaluation. Certains ont suivi un tutorat. Les élèves sont également classés selon leur niveau initial : *standard* ou *fragile*.

Niveau standard	Réussite	Échec	Total
Avec tutorat	18	2	20
Sans tutorat	80	20	100
Total	98	22	120

Niveau fragile	Réussite	Échec	Total
Avec tutorat	60	40	100
Sans tutorat	5	5	10
Total	65	45	110

Partie A - Construire la vue d'ensemble

1. Identifier la population et les trois variables observées. Préciser leur nature.
2. Vérifier les totaux des deux tableaux.
3. Construire le tableau croisant seulement tutorat et réussite, sans distinguer le niveau initial.
4. Calculer les fréquences marginales du tutorat et de la réussite.
5. Calculer la fréquence de réussite parmi les élèves tutorés.
6. Calculer la fréquence de réussite parmi les élèves non tutorés.
7. Quelle conclusion naïve donnerait la seule comparaison globale ?

Partie B - Comparer à niveau initial fixé

8. Dans le niveau standard, calculer les taux de réussite avec et sans tutorat.
9. Dans le niveau fragile, calculer les taux de réussite avec et sans tutorat.
10. Comparer ces résultats avec la conclusion de la question 7.
11. Nommer le phénomène observé.

Partie C - Comprendre le renversement

12. Parmi les élèves tutorés, calculer la fréquence du niveau fragile.
13. Parmi les élèves non tutorés, calculer la fréquence du niveau fragile.
14. Expliquer en quoi la composition des groupes est très différente.
15. Retrouver le taux global de réussite des tutorés comme moyenne pondérée de 90 % et 60 %.
16. Faire le même travail pour les non-tutorés avec 80 % et 50 %.

« Toujours demander : parmi quel groupe ? »

Partie D - Comparer avec les mêmes poids

17. Calculer le taux standardisé des tutorés en donnant un poids 1/2 à chaque niveau.
18. Calculer de même le taux standardisé des non-tutorés.
19. Comparer ces deux taux en points de pourcentage.
20. Pourquoi cette standardisation est-elle plus équitable pour comparer les groupes ?

Partie E - Généraliser et garder un esprit critique

On note S le niveau standard, F le niveau fragile, T le tutorat et R la réussite.

21. Montrer que

$$f_T(R) = f_T(S)f_{T \cap S}(R) + f_T(F)f_{T \cap F}(R).$$

22. Expliquer comment des poids très différents peuvent renverser la comparaison globale.
23. Peut-on conclure que le tutorat cause une meilleure réussite ? Rédiger une réponse prudente et proposer une information supplémentaire utile.

⚠ Attention | Avant le corrigé

Dans chaque fraction, vérifie le groupe de référence. Une phrase comme « 60 % réussissent » est incomplète tant que l'on ne sait pas parmi quels élèves.

15 Corrigés détaillés

Corrigé 1

 *Démonstration / Nommer ce qui est observé*

La population est l'ensemble des 80 foyers interrogés ; un individu est l'un de ces foyers.

- Le type principal de chauffage est qualitatif nominal : gaz, électricité, bois, etc. ne sont pas naturellement ordonnés.
- Le degré de satisfaction est qualitatif ordinal : faible < moyen < élevé.
- La facture mensuelle est quantitative, en pratique continue si elle est exprimée en euros et centimes.

Corrigé 2

 *Démonstration / Soustraire au total*

Dans la ligne de A , la case manquante vaut

$$60 - 36 = \boxed{24}.$$

On contrôle avec la colonne de $\bar{B}$:

$$24 + 16 = 40.$$

Les deux marges donnent bien $60 + 40 = 100$.

	B	$\bar{B}$	Total
A	36	24	60
$\bar{A}$	24	16	40
Total	60	40	100

Corrigé 3

 *Démonstration / Le total $N = 160$ est le dénominateur*

Les quatre fréquences demandées sont globales ou marginales : le dénominateur est donc toujours 160.

$$f(A) = \frac{50}{160} = \boxed{31,25\%}, \quad f(H) = \frac{124}{160} = \boxed{77,5\%}.$$

$$f(A \cap \bar{H}) = \frac{8}{160} = \boxed{5\%}, \quad f(M) = \frac{40}{160} = \boxed{25\%}.$$

Corrigé 4

 *Démonstration | Lire le symbole de gauche à droite*

$$f_A(H) = \frac{\text{Card}(A \cap H)}{\text{Card}(A)} = \frac{42}{50} = \boxed{84 \%}.$$

Parmi les utilisateurs d'un transport actif, 84 % sont ponctuels.

$$f_H(A) = \frac{42}{124} \approx \boxed{33,9 \%}.$$

Parmi les élèves ponctuels, environ 33,9 % utilisent un transport actif.

$$f_M(H) = \frac{24}{40} = \boxed{60 \%}.$$

Parmi les utilisateurs d'un transport motorisé, 60 % sont ponctuels.

Corrigé 5

 *Démonstration | Le dénominateur révèle la question*

- $12/70 \approx 17,1 \%$: parmi les 70 élèves en transports en commun, 12 sont en retard. C'est une fréquence conditionnelle.
- $12/36 = 33,3 \%$: parmi les 36 élèves en retard, 12 utilisent les transports en commun. C'est l'autre conditionnement.
- $24/160 = 15 \%$: dans toute la population, 24 élèves sont motorisés et ponctuels. C'est une fréquence globale d'intersection.

Les deux premières fractions ont le même numérateur 12, mais elles répondent à deux questions différentes.

Corrigé 6

 *Démonstration | Chaque colonne totalise 100 %*

	Actif	Transports	Motorisé
À l'heure	84,0 %	82,9 %	60,0 %
En retard	16,0 %	17,1 %	40,0 %
Total	100 %	100 %	100 %

Dans ces données, le taux de ponctualité du groupe motorisé est inférieur de 24 points à celui du groupe actif et d'environ 22,9 points à celui des transports en commun. Cela décrit une association, pas une cause.

Corrigé 7

 *Démonstration / Du pourcentage à la case*

Parmi les 240 abonnés, l'effectif premium vaut

$$0,35 \times 240 = 84.$$

Comme il y a 150 clients premium au total, $150 - 84 = 66$ ne sont pas abonnés. On complète par soustraction.

	Premium	Standard	Total
Abonnés	84	156	240
Non-abonnés	66	294	360
Total	150	450	600

Parmi les non-abonnés,

$$\frac{66}{360} \approx 18,3\%$$

choisissent le premium. L'écart avec les abonnés vaut

$$35 - 18,3 = 16,7 \text{ points environ}.$$

Corrigé 8

 *Démonstration / Deux comparaisons correctes mais différentes*

L'écart absolu entre les proportions vaut

$$72\% - 60\% = 12 \text{ points de pourcentage}.$$

L'augmentation relative, en prenant 60 % comme valeur de départ, vaut

$$\frac{72 - 60}{60} = \frac{12}{60} = 0,20 = 20\%.$$

Les 12 points comparent directement deux proportions. Les 20 % indiquent que l'écart 12 représente un cinquième de la proportion de départ 60.

Corrigé 9

 *Démonstration / Éviter le double comptage*

$$\text{Card}(A \cup H) = \text{Card}(A) + \text{Card}(H) - \text{Card}(A \cap H).$$

Ainsi,

$$\text{Card}(A \cup H) = 50 + 124 - 42 = 132.$$

Il reste

$$160 - 132 = 28$$

élèves qui ne sont ni actifs ni ponctuels. Dans $50 + 124$, les 42 élèves de l'intersection sont comptés deux fois : on les retire une fois.

Corrigé 10

 *Démonstration / Compter les couples*

Les couples observés donnent :

	Oui	Non	Total
Groupe A	2	2	4
Groupe B	2	2	4
Total	4	4	8

Dans chaque groupe, la fréquence de « oui » vaut $2/4 = 50\%$. Les distributions conditionnelles sont identiques dans ces huit données : aucune association n'y apparaît. Avec un si petit effectif, cette conclusion reste très fragile.

Corrigé 11

 *Démonstration / Appliquer 28 % au bon groupe*

Le groupe A contient 150 personnes. Donc

$$\text{Card}(A \cap B) = 0,28 \times 150 = 42.$$

Il y a alors $150 - 42 = 108$ personnes de A dans $\bar{B}$. Comme B contient 100 personnes au total, $100 - 42 = 58$ sont dans $\bar{A} \cap B$. Enfin, $\bar{A}$ contient $400 - 150 = 250$ personnes, donc $250 - 58 = 192$ sont dans $\bar{A} \cap \bar{B}$.

	B	$\bar{B}$	Total
A	42	108	150
$\bar{A}$	58	192	250
Total	100	300	400

Corrigé 12

 *Démonstration / Ne jamais inverser le conditionnement*

Parmi les malades, la fréquence des tests positifs est

$$\frac{90}{100} = \boxed{90\%}.$$

Parmi les non-malades, la fréquence des tests négatifs est

$$\frac{855}{900} = \boxed{95\%}.$$

Mais parmi les tests positifs, la fréquence des personnes malades vaut

$$\frac{90}{135} = \frac{2}{3} \approx \boxed{66,7\%}.$$

La première fraction conditionne par « malade » ; la troisième par « test positif ». Le numérateur 90 est identique, mais les groupes de référence, 100 et 135, diffèrent. La fréquence de la maladie dans la population intervient donc dans la composition du groupe des tests positifs.

Corrigé 13

 *Démonstration / Égaliser les deux fréquences*

La fréquence de B dans la première ligne vaut

$$\frac{30}{30 + 20} = \frac{3}{5}.$$

On veut donc

$$\frac{x}{x + 40} = \frac{3}{5}.$$

Le produit en croix donne $5x = 3x + 120$, donc

$$\boxed{x = 60}.$$

Avec $a = 30$, $b = 20$, $c = 60$ et $d = 40$,

$$ad = 30 \times 40 = 1200 \quad \text{et} \quad bc = 20 \times 60 = 1200.$$

Le critère $ad = bc$ est vérifié.

Corrigé 14

 *Démonstration / Pondérer par les tailles*

Le premier site fournit $0,75 \times 200 = 150$ réponses ; le second en fournit $0,60 \times 300 = 180$. Le taux global est donc

$$\frac{150 + 180}{200 + 300} = \frac{330}{500} = \boxed{66\%}.$$

Avec une moyenne pondérée,

$$\frac{200}{500} \times 75 \% + \frac{300}{500} \times 60 \% = 30 \% + 36 \% = 66 \%.$$

La moyenne simple 67,5 % donnerait le même poids aux deux sites alors que le second est plus grand.

Corrigé 15



Démonstration / Décomposer selon A et $\bar{A}$

Les ensembles A et $\bar{A}$ forment une partition. La formule de moyenne pondérée donne

$$f(B) = f(A)f_A(B) + f(\bar{A})f_{\bar{A}}(B),$$

avec $f(\bar{A}) = 1 - f(A)$. En notant $x = f_{\bar{A}}(B)$,

$$0,54 = 0,40 \times 0,75 + 0,60x.$$

Ainsi,

$$0,54 = 0,30 + 0,60x, \quad x = \frac{0,24}{0,60} = \boxed{0,40}.$$

La fréquence de B vaut 75 % parmi A , contre 40 % parmi $\bar{A}$, soit un écart de 35 points. Dans ces données, les variables sont donc fortement associées.

16 Corrigé du problème défi

Partie A - Vue d'ensemble

 *Démonstration* / Questions 1 à 7

1. La population est constituée des 230 élèves. Les trois variables sont le suivi du tutorat (oui/non), le niveau initial (standard/fragile) et le résultat (réussite/échec). Elles sont toutes qualitatives ; le niveau initial peut être considéré comme ordinal.

2. Pour le niveau standard, $18 + 2 + 80 + 20 = 120$. Pour le niveau fragile, $60 + 40 + 5 + 5 = 110$. On retrouve $120 + 110 = 230$ élèves.

3. On additionne les deux niveaux case par case.

	Réussite	Échec	Total
Avec tutorat	78	42	120
Sans tutorat	85	25	110
Total	163	67	230

4. Les fréquences marginales principales sont

$$f(T) = \frac{120}{230} \approx \boxed{52,2\%}, \quad f(R) = \frac{163}{230} \approx \boxed{70,9\%}.$$

5. Parmi les élèves tutorés,

$$f_T(R) = \frac{78}{120} = \boxed{65\%}.$$

6. Parmi les élèves non tutorés,

$$f_{\bar{T}}(R) = \frac{85}{110} \approx \boxed{77,3\%}.$$

7. Une lecture globale naïve dirait que le groupe sans tutorat réussit davantage, avec environ 12,3 points de plus. Cette phrase décrit le tableau global mais ne tient pas compte du niveau initial.

Partie B - Comparaison à niveau fixé

 *Démonstration* / Questions 8 à 11

8. Dans le niveau standard,

$$f_{T \cap S}(R) = \frac{18}{20} = \boxed{90\%}, \quad f_{\bar{T} \cap S}(R) = \frac{80}{100} = \boxed{80\%}.$$

Le taux tutoré est supérieur de 10 points.

9. Dans le niveau fragile,

$$f_{T \cap F}(R) = \frac{60}{100} = \boxed{60\%}, \quad f_{\bar{T} \cap F}(R) = \frac{5}{10} = \boxed{50\%}.$$

Le taux tutoré est encore supérieur de 10 points.

10. Dans chaque niveau, le tutorat est associé à un taux supérieur ; pourtant, dans l'ensemble, son taux est inférieur. La comparaison globale est donc renversée.

11. Ce phénomène s'appelle le paradoxe de Simpson. Il ne s'agit pas d'une contradiction : les groupes globaux n'ont pas la même composition.

Partie C - Composition et pondération

 Démonstration / Questions 12 à 16

12. Parmi les tutorés,

$$f_T(F) = \frac{100}{120} \approx \boxed{83,3\%}.$$

13. Parmi les non-tutorés,

$$f_{\overline{T}}(F) = \frac{10}{110} \approx \boxed{9,1\%}.$$

14. Le groupe tutoré est composé en très grande majorité d'élèves fragiles, dont les taux de réussite sont plus faibles dans les deux situations. Le groupe non tutoré contient surtout des élèves de niveau standard. Cette différence de composition pèse fortement sur les taux globaux.

15. Pour les tutorés,

$$\frac{20}{120} \times 90\% + \frac{100}{120} \times 60\% = 15\% + 50\% = \boxed{65\%}.$$

16. Pour les non-tutorés,

$$\frac{100}{110} \times 80\% + \frac{10}{110} \times 50\% = \frac{80 + 5}{110} \approx \boxed{77,3\%}.$$

Les poids associés aux niveaux standard et fragile sont presque inversés entre les deux groupes.

Partie D - Standardisation

 Démonstration / Questions 17 à 20

17. Avec un poids 1/2 pour chaque niveau,

$$\frac{1}{2} \times 90\% + \frac{1}{2} \times 60\% = \boxed{75\%}.$$

18. Pour les non-tutorés,

$$\frac{1}{2} \times 80\% + \frac{1}{2} \times 50\% = \boxed{65\%}.$$

19. L'écart standardisé vaut

$$75\% - 65\% = \boxed{10 \text{ points de pourcentage}}.$$

20. Les deux taux utilisent maintenant exactement les mêmes poids pour les niveaux. La différence ne provient donc plus du fait que l'un des groupes contient davantage d'élèves fragiles. La comparaison est plus équitable sur ce facteur précis, mais d'autres différences peuvent encore exister.

Partie E - Généralisation et prudence *Démonstration / Questions 21 à 23*

21. Les tutorés sont partagés entre $T \cap S$ et $T \cap F$. La formule de décomposition donne

$$f_T(R) = f_T(S)f_{T \cap S}(R) + f_T(F)f_{T \cap F}(R).$$

Numériquement,

$$\frac{20}{120} \times \frac{18}{20} + \frac{100}{120} \times \frac{60}{100} = \frac{18 + 60}{120} = \frac{78}{120}.$$

22. Un taux global est une moyenne pondérée. Même si chacun des deux taux tutorés est supérieur au taux non tutoré correspondant, un poids très fort sur le niveau fragile peut faire baisser la moyenne tutorée sous l'autre moyenne globale.

23. On ne peut pas conclure à une causalité : les élèves n'ont peut-être pas été répartis au hasard, et le tutorat a pu être proposé prioritairement aux élèves les plus en difficulté. Il serait utile de disposer de scores initiaux plus précis, d'autres caractéristiques pertinentes ou, si le cadre le permet, d'une comparaison construite avec des groupes initialement comparables.

17 Fiche-mémoire

Un pourcentage n'a de sens qu'avec son groupe de référence

Définition | Tableau croisé

Une case intérieure représente une intersection. Les marges sont les totaux des lignes et colonnes.

Définition | Fréquence globale

$$f(A \cap B) = \frac{n(A \cap B)}{N}.$$

Le dénominateur est toute la population.

Définition | Fréquence marginale

$$f(A) = \frac{n(A)}{N}.$$

Elle se lit dans une marge du tableau.

Méthode | ET, OU, NON

$$\text{ET} \leftrightarrow \cap, \quad \text{OU} \leftrightarrow \cup, \quad \text{NON } A \leftrightarrow \bar{A}.$$

Définition | Fréquence conditionnelle

$$f_A(B) = \frac{n(A \cap B)}{n(A)}.$$

« Parmi A » signifie que A fournit le dénominateur.

Méthode | Décomposition

Si les A_i forment une partition,

$$f(B) = \sum_i f(A_i) f_{A_i}(B).$$

C'est une moyenne pondérée.

Méthode | Comparer

1. calcule les taux de chaque groupe ;
2. compare en points de pourcentage ;
3. précise « dans ces données » ;
4. ne conclus pas automatiquement à une cause.

Méthode | Indépendance observée

Dans un tableau 2×2 , mêmes distributions conditionnelles

$$\text{si et seulement si } ad = bc.$$

Attention | Les quatre pièges

Inverser $f_A(B)$ et $f_B(A)$; diviser par N quand la question dit « parmi » ; comparer des effectifs lorsque les groupes n'ont pas la même taille ; transformer une association en causalité.

Réflexe final : écris le groupe de référence avant chaque calcul. Le dénominateur devient alors presque toujours évident.