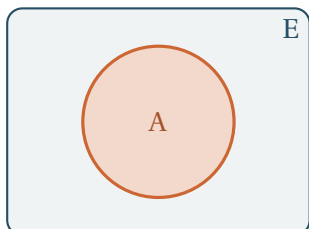


ANALYSE DE L'INFORMATION CHIFFRÉE

I Rappels : proportions et pourcentages

I.1 Une partie dans un ensemble

Définition 1 : Proportion



Soit E une **population** d'effectif n_E et A une **sous-population** de E, d'effectif n_A . La **proportion** de A dans E est :

$$p = \frac{n_A}{n_E} = \frac{\text{Nombre d'éléments de A}}{\text{Nombre d'éléments de E}}$$

Remarque I.1. Une proportion est **toujours** comprise entre 0 et 1.

Propriété 2 : Différentes écritures d'une même proportion

Une proportion peut s'écrire sous forme fractionnaire, décimale ou en pourcentage :

$$\underbrace{\frac{1}{4}}_{\text{fraction irréductible}} = \underbrace{0,25}_{\text{écriture décimale}} = \underbrace{\frac{25}{100}}_{\text{fraction sur 100}} = \underbrace{25\%}_{\text{pourcentage}}.$$

Méthode 3 : Calculer une partie connaissant le total

Pour calculer $p\%$ d'une quantité, on multiplie cette quantité par $\frac{p}{100}$.

Exemple 4 : Une TVA de 20 %

Sonia achète une voiture 2 635 €. Le montant de la TVA est égal à 20 % du prix :

$$\text{TVA} = 2\,635 \times \frac{20}{100} = 527. \quad \text{La TVA s'élève à 527 €}.$$

Exercice 5 : À partir de la classe

Déterminer la proportion de gauchers dans la classe. Donner le résultat sous forme d'une fraction, d'un nombre décimal puis d'un pourcentage.

La proportion de gauchers dans la classe est de : $\frac{\text{Nombre de gauchers}}{\text{Nombre de d'élèves total}}$.

II Croiser deux caractères

II.1 Tableau croisé d'effectifs

Définition 6

Un **tableau croisé**, ou tableau à double entrée, permet d'étudier simultanément deux caractères dans une population. Chaque case intérieure contient l'effectif correspondant à une combinaison des deux caractères. Les totaux des lignes et des colonnes sont appelés les **marges**.

Exemple 7 : Les employés d'une entreprise

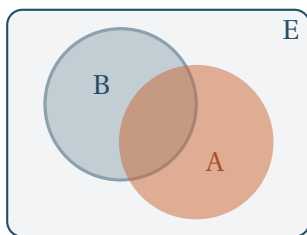
Une entreprise emploie 2 000 personnes. Parmi elles, 40 % sont cadres. On sait aussi que :

- 13,5 % des non-cadres ont entre 20 et 35 ans ;
- 88 cadres ont entre 20 et 35 ans ;
- 198 cadres ont plus de 50 ans ;
- parmi les personnes de plus de 50 ans, 24,75 % sont cadres.

Compléter le tableau.

	Entre 20 et 35 ans	Entre 35 et 50 ans	Plus de 50 ans	Total
Cadres	88	514	198	800
Non-cadres	162	436	602	1 200
Total	250	950	800	2 000

Remarque II.1. Il faut toujours vérifier les totaux sur les lignes et les colonnes.

II.2 Une proportion parmi un groupe**Définition 8 : Fréquence conditionnelle**

La **fréquence de A parmi B** est la proportion d'individus qui vérifient A lorsque l'on se limite au groupe B :

$$f_B(A) = \frac{\text{effectif vérifiant à la fois A et B}}{\text{effectif de B}}.$$

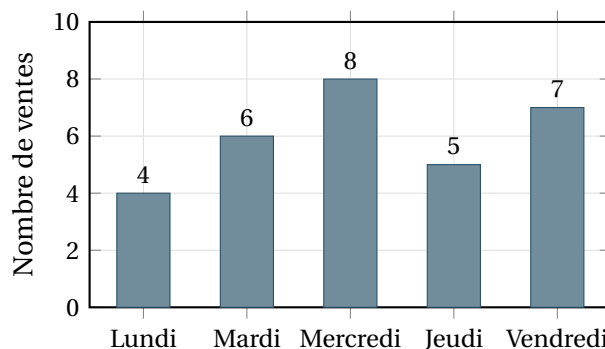
Remarque II.2. Dans « A **parmi** B », le mot **parmi** indique le groupe de référence. Le dénominateur est l'effectif de B.

Exemple 9 : Attention au groupe de référence

Dans le tableau précédent, la fréquence des cadres parmi les 20-35 ans est $\frac{88}{250} = 0,352 = 35,2\%$, alors que la fréquence des 20-35 ans parmi les cadres est $\frac{88}{800} = 0,11 = 11\%$.

III Choisir une représentation**III.1 Diagrammes en barres****Définition 10**

Un **diagramme en barres** représente les effectifs ou les fréquences des différentes catégories d'un caractère. La hauteur de chaque barre est proportionnelle à la valeur représentée.



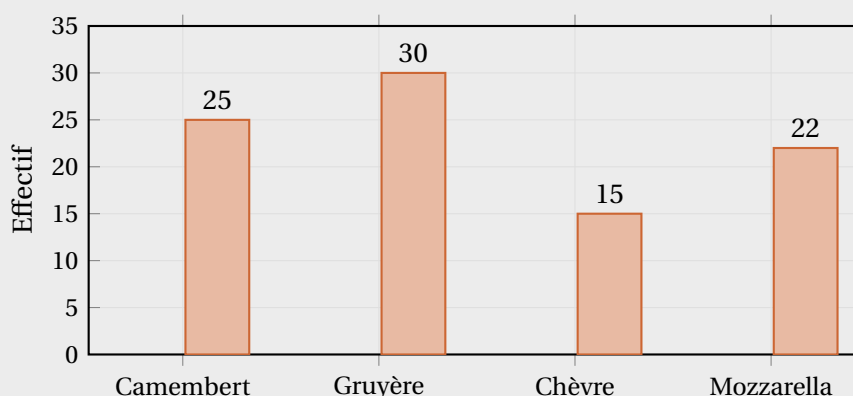
Remarque III.1. L'axe des abscisses n'est pas gradué : les jours de la semaine sont des catégories et « mardi est plus grand que lundi » n'a pas de sens. On conserve leur ordre chronologique pour faciliter la lecture.

Exercice 11 : Le fromage préféré des clients

Un supermarché interroge ses clients sur le fromage qu'ils consomment le plus souvent. Les résultats obtenus sont regroupés dans le tableau ci-contre.

Fromage	Camembert	Gruyère	Chèvre	Mozzarella
Effectif	25	30	15	22

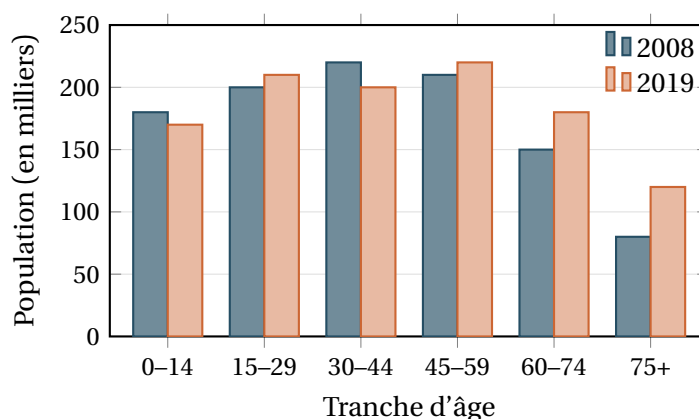
Compléter le repère ci-dessous en traçant le diagramme en barres correspondant.



Méthode 12 : Comparer deux séries avec un diagramme en barres

Pour comparer une même répartition à deux instants, on place, pour chaque catégorie, les deux barres côte à côte dans le même repère, avec la même échelle.

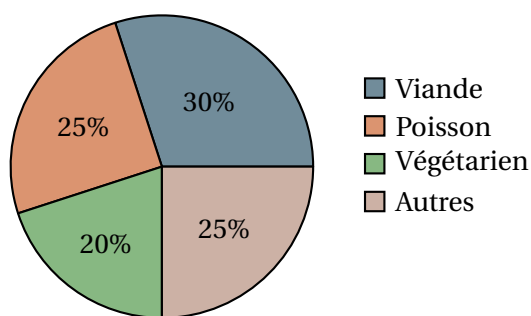
Le graphique ci-contre compare, à partir de données fictives, la population d'un département par tranche d'âge en 2008 et en 2019.



III.2 Diagramme circulaire

Définition 13

Un **diagramme circulaire** est un disque partagé en secteurs. L'angle de chaque secteur est proportionnel à l'effectif ou à la fréquence représentée. Le disque entier correspond à 100 %, soit 360°.



Pour la catégorie « Viande » :

$$30\% \times 360^\circ = \frac{30}{100} \times 360^\circ = 108^\circ.$$

Méthode 14

Pour passer d'un pourcentage à un angle, on calcule

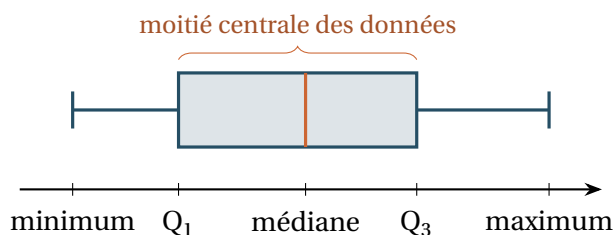
$$\frac{\text{pourcentage}}{100} \times 360^\circ.$$

III.3 Rappel : lire un diagramme en boîte

Définition 15

Un diagramme en boîte fait apparaître cinq valeurs : le minimum, le premier quartile Q_1 , la médiane, le troisième quartile Q_3 et le maximum.

Au moins la moitié des données se trouve entre Q_1 et Q_3 .



Remarque III.2. La « boîte à moustaches » résume seulement quelques indicateurs. Ce diagramme est utilisé principalement pour **comparer** un même caractère dans des populations différentes.

IV Étudier deux caractères quantitatifs

IV.1 Nuage de points

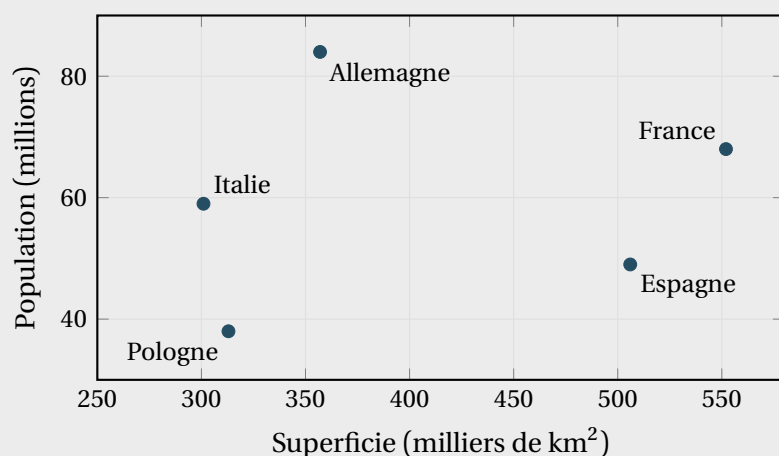
Définition 16

On dispose de n observations constituées de couples $(x_i ; y_i)$. Le **nuage de points** associé est l'ensemble des points $M_i(x_i ; y_i)$ placés dans un repère.

Exemple 17 : Superficie et population de quelques pays européens

Pour chaque pays, on associe sa superficie x , en milliers de km^2 , et sa population y , en millions d'habitants.

Pays	Allemagne	France	Italie	Espagne	Pologne
x	357	552	301	506	313
y	84	68	59	49	38



Le point qui représente la France a pour coordonnées

(552; 68).

Il associe la superficie et la population de la France.

IV.2 Point moyen

Définition 18

Le **point moyen** du nuage est le point $G(\bar{x}; \bar{y})$ dont les coordonnées sont les moyennes des deux séries :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad \text{et} \quad \bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n}.$$

Exemple 19 : Point moyen du nuage précédent

Pour les cinq pays :

$$\bar{x} = \frac{357 + 552 + 301 + 506 + 313}{5} = 405,8 \quad \text{et} \quad \bar{y} = \frac{84 + 68 + 59 + 49 + 38}{5} = 59,6.$$

Le point moyen est donc $G(405,8; 59,6)$.

IV.3 Ajuster un nuage par une droite

Définition 20

Lorsque le nuage présente une tendance approximativement rectiligne, on peut l'approcher par une droite d'équation $y = ax + b$. Cette droite est appelée **droite d'ajustement affine**.

Exemple 21 : Évolution de la population par âge

Dans le jeu de données fictif suivant, on associe, pour six tranches d'âge, la population de 2008, notée x , et celle de 2019, notée y , en milliers d'habitants.

Tranche d'âge	0-14	15-29	30-44	45-59	60-74	75 et plus
x en 2008	180	200	220	210	150	80
y en 2019	170	210	200	220	180	120

La méthode suivante permet de construire la **droite de Mayer**, qui peut approximer un nuage de points.

1. Ranger les points par abscisses croissantes.
2. Partager la série en deux groupes de même taille, ou de tailles aussi proches que possible.
3. Calculer les points moyens G_1 et G_2 des deux groupes.

4. Tracer la droite ($G_1 G_2$) et déterminer son équation.

Après avoir rangé les points par abscisses croissantes, on obtient :

$$G_1 \left(\frac{80 + 150 + 180}{3} ; \frac{120 + 180 + 170}{3} \right) = \left(\frac{410}{3} ; \frac{470}{3} \right),$$

$$G_2 \left(\frac{200 + 210 + 220}{3} ; \frac{210 + 220 + 200}{3} \right) = (210 ; 210).$$

Le coefficient directeur de la droite ($G_1 G_2$) vaut :

$$a = \frac{210 - \frac{470}{3}}{210 - \frac{410}{3}} = \frac{8}{11}.$$

Puisque G_2 est sur la droite, on a :

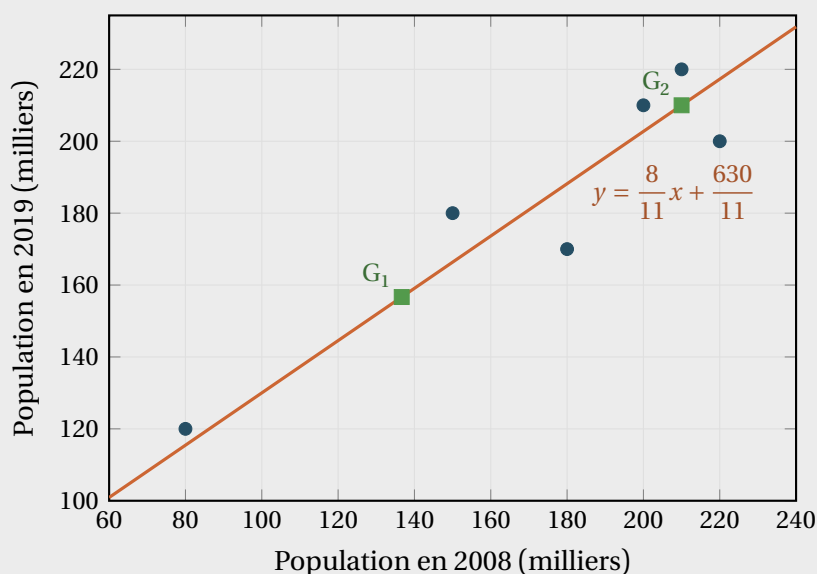
$$210 = \frac{8}{11} \times 210 + b,$$

donc

$$b = 210 - \frac{8}{11} \times 210 = \frac{630}{11}.$$

On obtient alors l'équation :

$$y = \frac{8}{11}x + \frac{630}{11}.$$

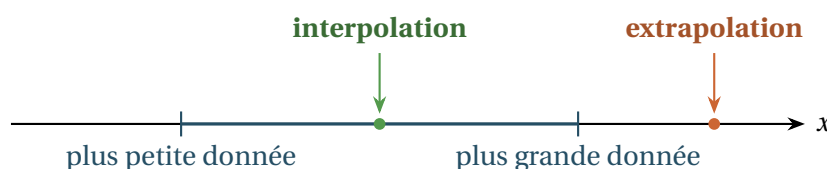


Remarque IV.1. On peut aussi utiliser un **tableur** pour tracer le nuage de points et obtenir directement l'équation d'une droite d'ajustement affine. Cette commande est souvent appelée « régression linéaire » dans les logiciels. La droite obtenue peut être légèrement différente de la droite de Mayer.

IV.4 Interpolation et extrapolation

Définition 22

- Une **interpolation** estime y pour une valeur de x située à l'intérieur de l'intervalle des abscisses observées.
- Une **extrapolation** estime y pour une valeur de x située à l'extérieur de cet intervalle.



Exercice 23 : Utiliser le modèle précédent

On reprend la droite d'ajustement $y = \frac{8}{11}x + \frac{630}{11}$.

1. Estimer y lorsque $x = 190$. S'agit-il d'une interpolation ou d'une extrapolation ?

$$y = \frac{8}{11} \times 190 + \frac{630}{11} = \frac{2150}{11} \approx 195,5.$$

La valeur 190 appartient à l'intervalle $[80 ; 220]$ couvert par les données : il s'agit d'une interpolation.

2. Même question lorsque $x = 250$.

$$y = \frac{8}{11} \times 250 + \frac{630}{11} = \frac{2630}{11} \approx 239,1.$$

La valeur 250 est en dehors de l'intervalle observé : il s'agit d'une extrapolation.

Cette estimation est plus fragile.

À retenir

Une droite d'ajustement est un **modèle**. Elle simplifie les données observées et ne garantit pas que la même tendance se poursuivra en dehors de la période ou du domaine étudié.