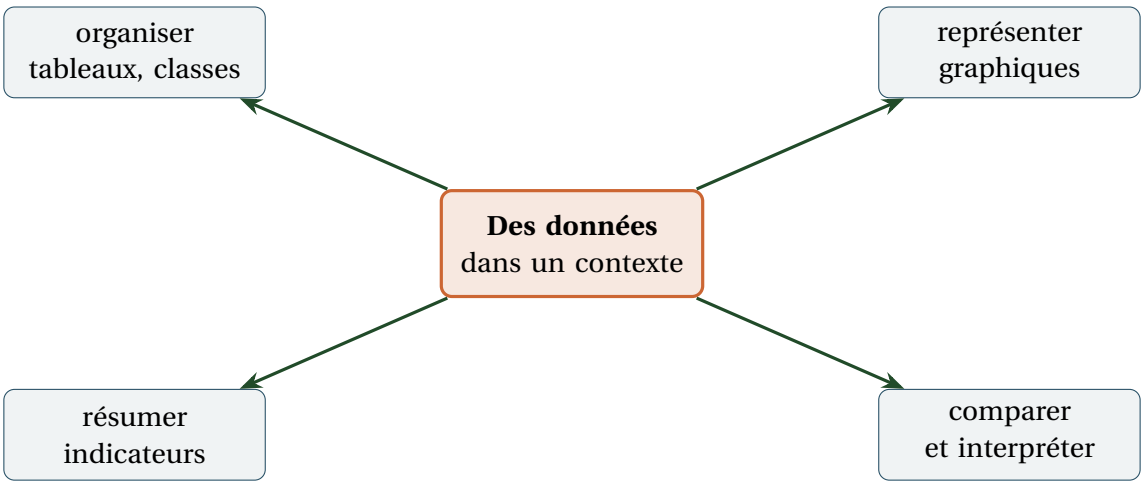
STATISTIQUES



Méthode 1 : Un indicateur n'est jamais une conclusion

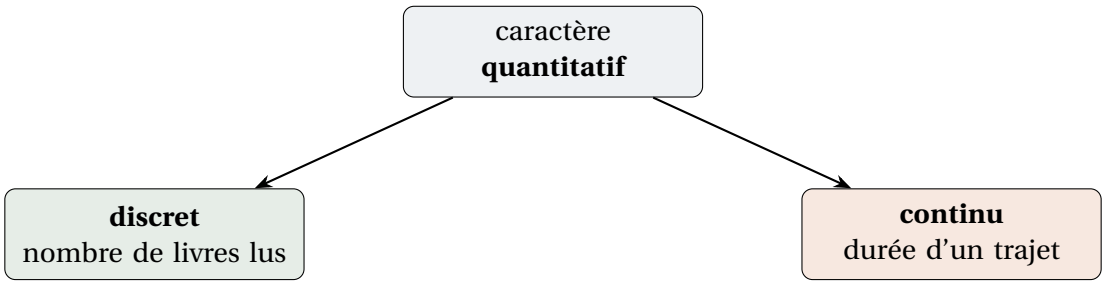
Une étude statistique doit toujours préciser la population, le caractère, l'unité et la source des données. Un nombre comme une moyenne ou un écart type ne prend sens qu'avec le graphique et le contexte. Deux séries peuvent partager un même centre tout en ayant des dispersions très différentes.

I Décrire une série quantitative

I.1 Vocabulaire et fréquences

Définition 2

- La **population** est l'ensemble des individus étudiés.
- Un **individu** est un élément de cette population.
- Le **caractère** est la propriété observée sur chaque individu.
- Un caractère est **quantitatif** si ses valeurs sont numériques. Il est **discret** si les valeurs sont isolées, et **continu** si elles peuvent parcourir un intervalle.
- L'**effectif** n_i d'une valeur x_i est le nombre d'individus qui la possèdent. Sa fréquence est $f_i = \dfrac{n_i}{\mathrm{N}}$, où N est l'effectif total.



Remarque I.1. Les effectifs totalisent N et les fréquences totalisent 1, soit 100 %.

I.2 Moyenne

Définition 3

Pour les valeurs $x_1, \dots, x_p$ d'effectifs $n_1, \dots, n_p$, la **moyenne pondérée** est

$$\boxed{\overline{x} = \frac{n_1 x_1 + \cdots + n_p x_p}{\mathrm{N}}}, \qquad \mathrm{N} = n_1 + \cdots + n_p.$$

Exemple 4 : Masse de dix paquets

Une ligne de production fournit les masses suivantes, en grammes :

$$496,\ 498,\ 499,\ 500,\ 500,\ 500,\ 501,\ 501,\ 502,\ 503.$$

Leur somme vaut 5 000, donc

$$\overline{x} = \frac{5\,000}{10} = 500 \text{ g}.$$

Propriété 5 : Linéarité de la moyenne

Si chaque valeur x_i est transformée en $y_i = ax_i + b$, alors

$$\boxed{\overline{y} = a\overline{x} + b.}$$

Exemple 6 : Changer d'unité sans tout recalculer

Une série de températures a une moyenne de 18 °C. La conversion en degrés Fahrenheit est $F = 1{,}8C + 32$. La moyenne convertie vaut donc

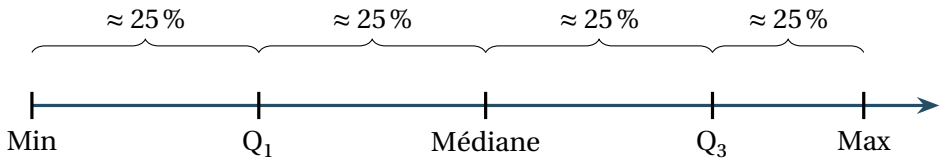
$$1{,}8 \times 18 + 32 = 64{,}4\ ^\circ\text{F}.$$

I.3 Médiane et quartiles : rappels

Définition 7

Les N valeurs sont rangées dans l'ordre croissant.

- Si N est impair, une **médiane** est la valeur centrale.
- Si N est pair, on choisit comme médiane la moyenne des deux valeurs centrales.
- Q_1 est la valeur de rang $\left\lceil \frac{N}{4} \right\rceil$.
- Q_3 est la valeur de rang $\left\lceil \frac{3N}{4} \right\rceil$.



Exemple 8 : Reprendre les dix paquets

Les valeurs sont déjà rangées. La médiane est la moyenne des valeurs de rangs 5 et 6, donc 500 g.

$$Q_1 = x_3 = 499 \text{ g}, \qquad Q_3 = x_8 = 501 \text{ g}.$$

II Mesurer la dispersion

II.1 Étendue et écart interquartile

Définition 9

- L'**étendue** est maximum − minimum.
- L'**écart interquartile** est $Q_3 - Q_1$. Il mesure l'étalement de la moitié centrale des données et dépend peu des valeurs extrêmes.

Exemple 10

Pour les masses précédentes,

$$\text{étendue} = 503 - 496 = 7 \text{ g}, \qquad Q_3 - Q_1 = 501 - 499 = 2 \text{ g}.$$

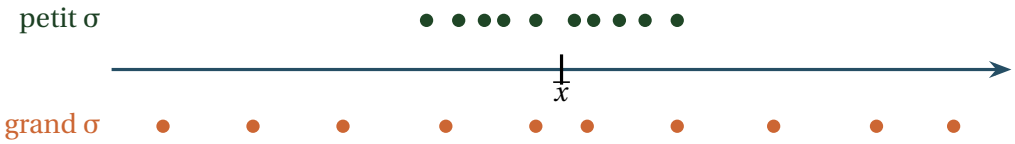
II.2 Écart type

Définition 11

L'**écart type** d'une série de moyenne $\overline{x}$ est

$$\sigma = \sqrt{\frac{n_1(x_1-\overline{x})^2 + \cdots + n_p(x_p-\overline{x})^2}{N}}.$$

Il s'exprime dans la même unité que le caractère et mesure la dispersion des données autour de la moyenne.



Remarque II.1. L'écart type n'est pas un intervalle qui contiendrait automatiquement toutes les valeurs. C'est une échelle de distance à la moyenne : plus il est grand, plus la série est dispersée.

Exemple 12 : Calcul de l'écart type

Pour les dix paquets, $\overline{x} = 500$ g et la somme des carrés des écarts vaut

$$16+4+1+0+0+0+1+1+4+9 = 36.$$

Ainsi

$$\sigma = \sqrt{\frac{36}{10}} \approx 1{,}90 \text{ g}.$$

Méthode 13 : Choisir un couple d'indicateurs

- Le couple $(\overline{x}\,;\sigma)$ décrit la position autour de la moyenne et la dispersion autour de celle-ci ; il est sensible aux valeurs extrêmes.
- Le couple $(\text{médiane}\,;Q_3-Q_1)$ décrit le centre et la moitié centrale ; il est moins sensible aux valeurs extrêmes.

Exercice 14 : Calculer et interpréter

Pour la série

$$8, 9, 9, 10, 10, 10, 10, 11, 11, 13,$$

calculer la moyenne, la médiane, Q_1, Q_3 et l'écart interquartile. La somme vaut 101, donc $\overline{x} = 10{,}1$. La médiane est 10. Comme

$$\left\lceil \frac{10}{4} \right\rceil = 3, \qquad \left\lceil \frac{30}{4} \right\rceil = 8,$$

on a $Q_1 = 9$, $Q_3 = 11$, puis $Q_3 - Q_1 = 2$.

III Ajouter ou supprimer une valeur

Propriété 15

Une série de N valeurs a pour moyenne $\overline{x}$, donc sa somme vaut $N\overline{x}$.

- Après l'ajout d'une valeur y,

$$\overline{x}_{\text{nouvelle}} = \frac{N\overline{x} + y}{N+1}.$$

- Après la suppression d'une valeur y, si $N > 1$,

$$\overline{x}_{\text{nouvelle}} = \frac{N\overline{x} - y}{N-1}.$$

Exemple 16 : Une valeur extrême

Les dix paquets ont une moyenne de 500 g et une médiane de 500 g.

- Si l'on ajoute la valeur 520, la moyenne devient

$$\frac{10 \times 500 + 520}{11} \approx 501{,}82 \text{ g},$$

  tandis que la médiane reste 500 g.
- Si l'on retire 496, la moyenne devient

$$\frac{10 \times 500 - 496}{9} \approx 500{,}44 \text{ g},$$

  et la médiane reste 500 g.

Remarque III.1. Pour la médiane, il n'existe pas de formule analogue : il faut réordonner la série et déterminer les nouveaux rangs centraux. Une valeur ajoutée ou retirée peut donc modifier ou non la médiane selon sa position.

Exercice 17 : Anticiper avant de calculer

Une série de 20 valeurs a pour moyenne 12. Dans les deux situations suivantes, on repart de cette série initiale : l'ajout et le retrait sont indépendants.

1. On ajoute 33. La moyenne augmente-t-elle ou diminue-t-elle ? La calculer.
2. On retire une valeur égale à 7. Calculer la nouvelle moyenne.
3. Peut-on prévoir la nouvelle médiane avec les seules informations données ?

La somme initiale vaut 240. Ainsi : 1. $\frac{240+33}{21} = 13$, donc la moyenne augmente ; 2. $\frac{240-7}{19} = \frac{233}{19} \approx 12{,}26$; 3. non, car il faut connaître l'ordre et la position des valeurs.

IV Série continue regroupée en classes

IV.1 Tableau et histogramme

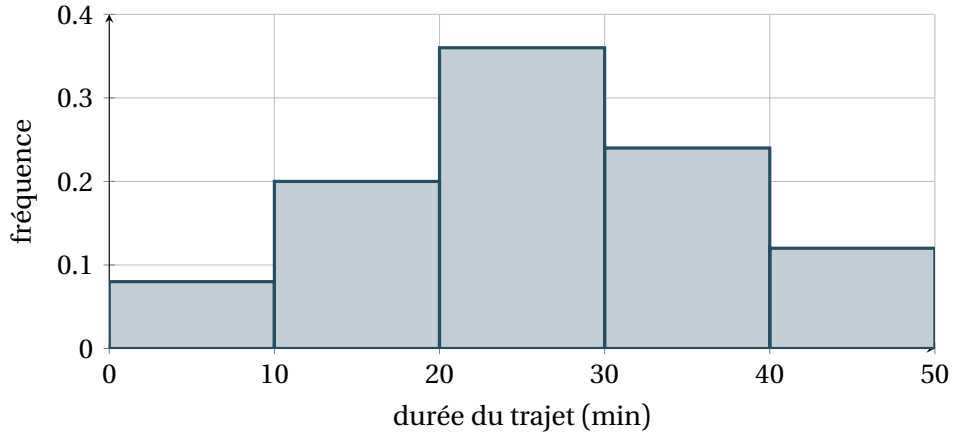
Exemple 18 : Durées de trajet anonymisées

On regroupe les durées de trajet, en minutes, de 50 personnes dans des classes de même amplitude 10.

Durée (min)	[0 ; 10[	[10 ; 20[	[20 ; 30[	[30 ; 40[	[40 ; 50[	**Total**
Effectif	4	10	18	12	6	50
Fréquence	0,08	0,20	0,36	0,24	0,12	1
Fréquence cumulée	0,08	0,28	0,64	0,88	1	

Définition 19

Pour un caractère continu, une **classe** est un intervalle de valeurs. Dans un histogramme, l'**aire** de chaque rectangle est proportionnelle à l'effectif ou à la fréquence de la classe.



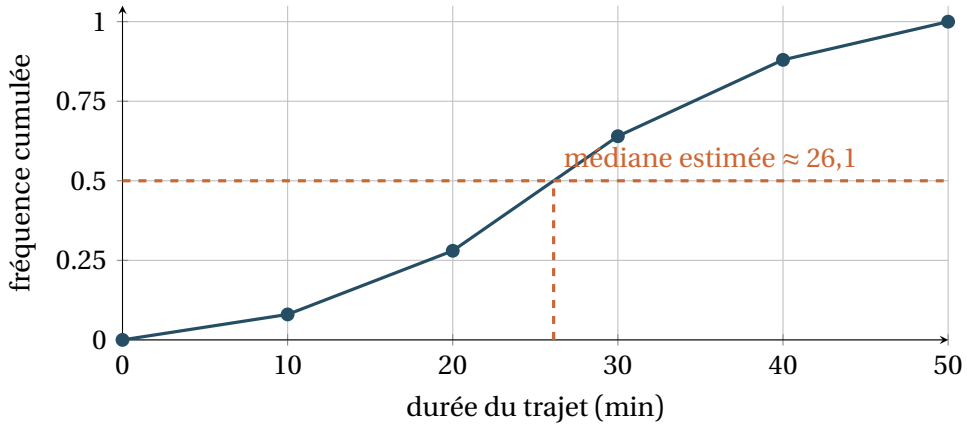
Remarque IV.1. Ici les classes ont la même amplitude : comparer les aires revient à comparer les hauteurs. Cette simplification ne serait plus valable avec des amplitudes différentes.

IV.2 Polygone des fréquences cumulées

Définition 20

La **fréquence cumulée croissante** à une borne b est la fréquence des valeurs inférieures à b. Le polygone relie les points dont :

- l'abscisse est une borne de classe ;
- l'ordonnée est la fréquence cumulée correspondante.



Méthode 21 : Construire le polygone

1. Calculer les fréquences cumulées de gauche à droite.
2. Placer le point de départ à la borne inférieure avec la fréquence 0.
3. À chaque borne supérieure, placer la fréquence cumulée de la classe terminée.
4. Relier les points et contrôler que le dernier a pour ordonnée 1.

V Estimer des indicateurs à partir des classes

V.1 Moyenne à partir de groupes

Propriété 22 : Moyenne de plusieurs groupes

Des groupes d'effectifs $N_1, \dots, N_p$ ont pour moyennes $\overline{x}_1, \dots, \overline{x}_p$. La moyenne de l'ensemble est

$$\boxed{\overline{x} = \frac{N_1\overline{x}_1 + \cdots + N_p\overline{x}_p}{N_1 + \cdots + N_p}.}$$

Exemple 23 : Réunir deux groupes

Un groupe de 18 personnes a une durée moyenne de trajet de 24 min ; un autre groupe de 12 personnes a une moyenne de 31 min. La moyenne des 30 personnes est

$$\frac{18 \times 24 + 12 \times 31}{30} = 26{,}8 \text{ min.}$$

On ne fait pas $\frac{24+31}{2}$, car les groupes n'ont pas le même effectif.

V.2 Estimation sous l'hypothèse d'uniformité

Définition 24 : Hypothèse d'uniformité

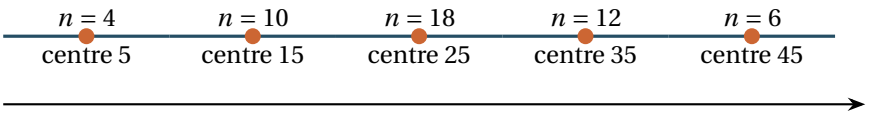
Lorsque seules les classes sont connues, on peut **supposer** que les valeurs sont réparties uniformément dans chaque classe. On remplace alors chaque classe par son centre pour estimer la moyenne.

Exemple 25 : Estimer la durée moyenne

Les centres des classes sont 5, 15, 25, 35 et 45. Sous l'hypothèse d'uniformité :

$$\overline{x}_{\text{estimée}} = \frac{4 \times 5 + 10 \times 15 + 18 \times 25 + 12 \times 35 + 6 \times 45}{50} = \frac{1\,310}{50} = 26{,}2 \text{ min.}$$

Ce résultat est une **estimation**, car les valeurs exactes ont été perdues lors du regroupement.



V.3 Classe médiane et médiane estimée

Définition 26

La **classe médiane** contient les valeurs centrales lorsque la moitié de l'effectif est strictement comprise entre deux cumuls successifs. Si un cumul vaut exactement la moitié du total, les deux valeurs centrales encadrent une borne de classe : leur moyenne ne peut pas être déterminée par le seul tableau.

Exemple 27 : Reprendre les trajets

La moitié de 50 vaut 25. Les effectifs cumulés sont

$$4,\ 14,\ \boxed{32},\ 44,\ 50.$$

La classe médiane est donc $[20\,;30[$.
Sous l'hypothèse d'une répartition uniforme dans cette classe, il faut parcourir $25 - 14 = 11$ des 18 valeurs

de la classe, soit la fraction $\frac{11}{18}$ de son amplitude 10 :

$$\text{médiane estimée} = 20 + \frac{11}{18} \times 10 \approx 26{,}1 \text{ min.}$$

Remarque V.1. Sans hypothèse sur la répartition à l'intérieur de [20 ; 30[, on peut seulement affirmer que la médiane appartient à cette classe. L'estimation 26,1 min n'est pas une valeur exacte.

Exercice 28 : Données regroupées

Les durées d'une tâche, en minutes, sont regroupées ainsi :

classe	[0 ; 5[	[5 ; 10[	[10 ; 15[	[15 ; 20[
effectif	6	14	16	4

1. Calculer les fréquences cumulées.
2. Localiser les deux valeurs centrales. La médiane exacte est-elle déterminée ?
3. Estimer la moyenne avec les centres de classes.
4. Estimer la médiane sous l'hypothèse d'uniformité.

L'effectif total est 40. Les fréquences cumulées sont 0,15, 0,50, 0,90, 1. Le cumul 50 % est atteint à la borne 10. La 20^{e} valeur appartient à [5 ; 10[et la 21^{e} à [10 ; 15[: leur moyenne n'est pas déterminée par le tableau. Les centres sont 2,5, 7,5, 12,5, 17,5, donc

$$\overline{x}_{\text{estimée}} = \frac{6 \times 2{,}5 + 14 \times 7{,}5 + 16 \times 12{,}5 + 4 \times 17{,}5}{40} = 9{,}75 \text{ min.}$$

Sous l'hypothèse d'uniformité, l'interpolation au niveau 0,5 donne la borne commune 10 : la médiane **estimée** vaut 10 min. Cette estimation n'est pas nécessairement la médiane exacte des données brutes.

VI Comparer deux distributions

Exemple 29 : Deux lignes de conditionnement

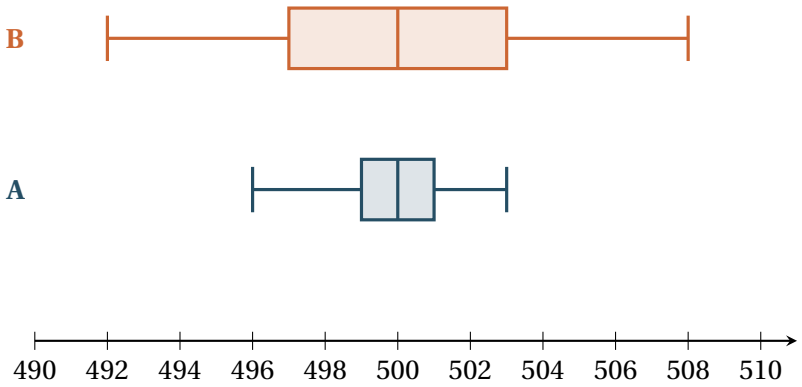
Les lignes A et B produisent chacune dix paquets. Leurs masses sont :

A : 496, 498, 499, 500, 500, 500, 501, 501, 502, 503,

B : 492, 494, 497, 499, 500, 500, 501, 503, 506, 508.

Les deux séries ont la même moyenne 500 g et la même médiane 500 g.

	A	B
σ	$\sqrt{3{,}6} \approx 1{,}90$	$\sqrt{22} \approx 4{,}69$
Q_1	499	497
Q_3	501	503
$Q_3 - Q_1$	2	6



Méthode 30 : Rédiger une comparaison

1. Comparer d'abord des indicateurs de même nature et munis de leur unité.
2. Décrire ensuite la dispersion à l'aide de σ, de $Q_3 - Q_1$ et du graphique.
3. Conclure dans le contexte, sans employer « meilleur » sans critère explicite.

Ici les centres sont identiques, mais la ligne A est plus régulière : son écart type et son écart interquartile sont plus petits, et sa boîte est plus resserrée.

Exercice 31 : Comparer avec des indicateurs

Deux temps de réaction, en millisecondes, sont résumés ainsi :

	$\overline{x}$	σ	médiane	$Q_3 - Q_1$
C	248	18	246	22
D	241	31	239	45

Rédiger une comparaison argumentée. Le groupe D est plus rapide en moyenne (241 ms contre 248 ms) et en médiane (239 ms contre 246 ms). En revanche, ses résultats sont moins réguliers : $\sigma_D = 31$ ms et son écart interquartile vaut 45 ms, contre 18 ms et 22 ms pour C. Le choix dépend donc du critère retenu : rapidité centrale ou régularité.

VII Croiser deux variables qualitatives

VII.1 Tableau croisé

Définition 32

Une variable **qualitative** décrit une catégorie. Un tableau croisé répartit les mêmes individus selon deux variables qualitatives :

- les cases intérieures sont les effectifs conjoints ;
- les totaux de lignes et de colonnes sont les effectifs marginaux.

Exemple 33 : Mobilité et distance

Une enquête anonyme porte sur 200 déplacements domicile–études.

Distance \ mode	**Vélo**	**Transports publics**	**Voiture**	**Total**
Moins de 5 km	42	18	20	80
Au moins 5 km	18	54	48	120
Total	60	72	68	200

VII.2 Fréquences marginales et conditionnelles

Définition 34

- Une **fréquence marginale** rapporte un total de ligne ou de colonne à l'effectif total.
- Une **fréquence conditionnelle** rapporte un effectif conjoint au total de la ligne ou de la colonne imposée par « parmi » ou « sachant ».

marginale
sur les 200 déplacements

conditionnelle
dans une ligne ou une colonne

$\frac{60}{200} = 30\,\%$ à vélo

$\frac{54}{120} = 45\,\%$ parmi les longs trajets

Exemple 35 : Deux conditionnements différents

- Parmi les déplacements d'au moins 5 km, la fréquence des transports publics est

$$\frac{54}{120} = 45\,\%.$$

- Parmi les déplacements en transports publics, la fréquence de ceux d'au moins 5 km est

$$\frac{54}{72} = 75\,\%.$$

Les numérateurs sont identiques, mais les populations de référence diffèrent.

Méthode 36 : Choisir le bon dénominateur

1. Traduire la question sous la forme « parmi …, quelle part … ? ».
2. Entourer la ligne ou la colonne correspondant à « parmi ».
3. Son total est le dénominateur ; l'intersection recherchée est le numérateur.

Exercice 37 : Compléter puis calculer

Une enquête croise le type de repas et le jour de présence.

	sur place	à emporter	total
lundi	36	14	50
mardi	24	…	40
total	60	…	90

1. Compléter le tableau.
2. Calculer la fréquence marginale des repas à emporter.
3. Parmi les repas du mardi, calculer la fréquence des repas à emporter.
4. Parmi les repas à emporter, calculer la fréquence de ceux du mardi.

Il y a $40 - 24 = 16$ repas à emporter le mardi et $14 + 16 = 30$ au total.

$$f(\text{à emporter}) = \frac{30}{90} = \frac{1}{3}, \qquad f_{\text{mardi}}(\text{à emporter}) = \frac{16}{40} = 40\,\%,$$

tandis que

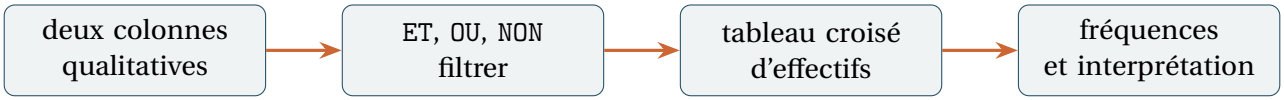
$$f_{\text{à emporter}}(\text{mardi}) = \frac{16}{30} \approx 53{,}3\,\%.$$

VIII Filtrer et croiser dans un tableur

Méthode 38 : De deux colonnes au tableau croisé

Si chaque ligne du tableur représente un individu :

1. utiliser `ET` pour imposer simultanément deux critères ;
2. utiliser `OU` lorsqu'au moins un critère suffit ;
3. utiliser `NON` pour exclure une modalité ;
4. compter les lignes filtrées, compléter les marges, puis calculer les fréquences.



Exercice 39 : Interpréter un filtre

Dans une base, la colonne B contient le mode de transport et la colonne C la catégorie de distance. Expliquer ce que sélectionnent :

1. `=ET(B2=« Vélo »;C2=« Moins de 5 km »)`
2. `=OU(B2=« Vélo »;B2=« Transports publics »)`
3. `=NON(B2=« Voiture »)`

1. Les déplacements simultanément effectués à vélo et inférieurs à 5 km.
2. Les déplacements à vélo ou en transports publics.
3. Tous les déplacements qui ne sont pas effectués en voiture.

À retenir

- Toujours annoncer population, caractère et unité avant de calculer.
- La linéarité donne $\overline{ax+b} = a\overline{x} + b$.
- σ mesure la dispersion autour de la moyenne ; $Q_3 - Q_1$ mesure celle de la moitié centrale.
- Après ajout ou retrait d'une valeur, la somme se met à jour ; pour la médiane, il faut recalculer les rangs.
- Pour des classes, centres et interpolation ne donnent que des **estimations sous une hypothèse d'uniformité**.
- Une comparaison associe indicateurs, graphiques, unités et conclusion contextualisée.
- Dans un tableau croisé, une fréquence marginale utilise le total général ; une fréquence conditionnelle utilise le total imposé par « parmi ».

Activité numérique AN3 du carnet partagé. Construire une feuille de calcul avec critères logiques, tableau croisé et fréquences conditionnelles.