

Statistique et probabilités

Première Gestion-Administration

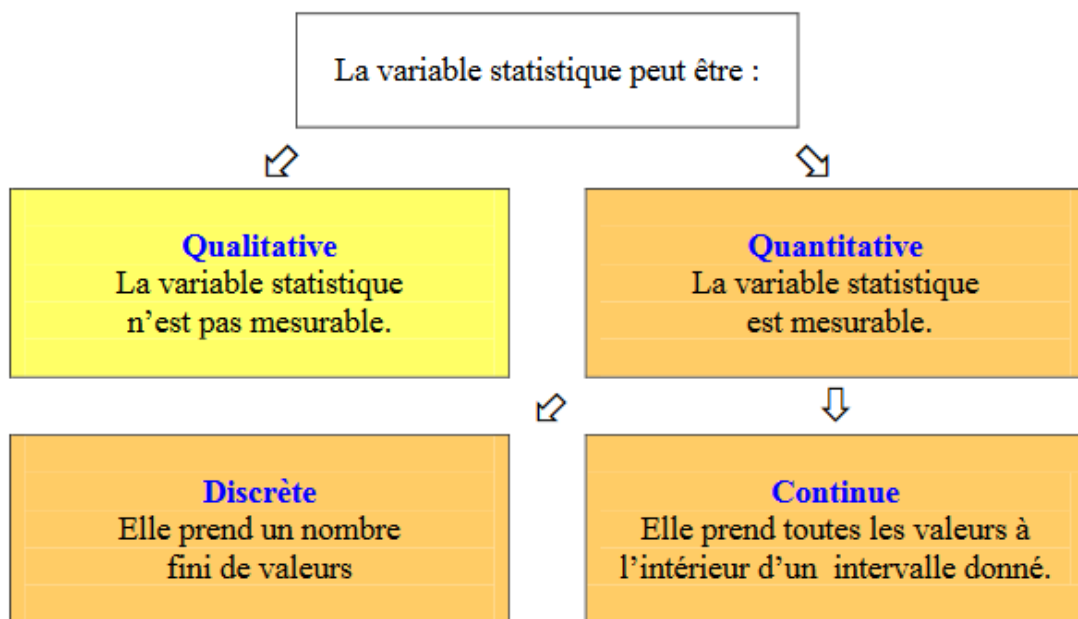
Pr. Marchetti

1. Rappel : Vocabulaire

L'ensemble sur lequel porte l'étude statistique est nommé **population** : comme l'ensemble des élèves de la classe qui ont répondu au questionnaire.

Chaque élément de la population étudiée est : une **unité statistique** ou un **individu** (élève, trajet journalier ou autres).

Le caractère ou variable statistique d'une population est la propriété sur laquelle porte l'étude statistique.



L'étude statistique d'une population par rapport à une variable continue impose de regrouper le grand nombre de valeurs en tranches ou **classes**. (classes d'âge pour une population de personnes comme dans la classe d'école).

Une classe, c'est la portion de l'intervalle auquel appartiennent les valeurs de caractères.

Une série statistique associe à chaque valeur x_i du caractère le nombre d'individus correspondant, appelé **effectif partiel** et noté n_i . L'**effectif total** de la population est noté N .

La **fréquence** d'une valeur x_i du caractère est le quotient de l'effectif n_i de ce caractère par l'effectif total N : $f_i = \frac{n_i}{N}$.

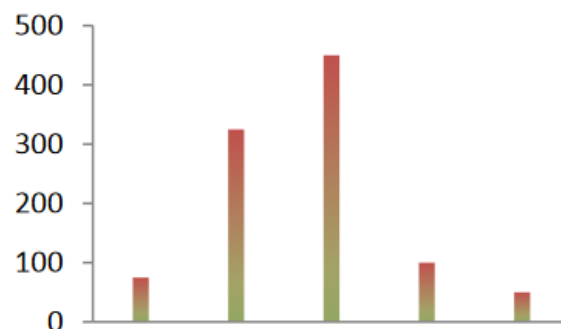
La somme des fréquences est égale à 1.

Les fréquences sont souvent exprimées en pourcentage après multiplication par 100 du rapport n_i/N .

2. Les types de représentation graphiques

Il existe différents types de représentations graphiques en statistiques :

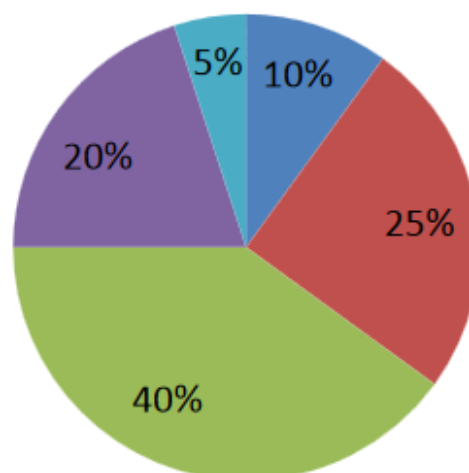
- **Diagramme en bâtons** : On l'utilise pour les séries à caractère discret. Pour celles qui utilisent un repère cartésien :
 - sur l'axe des abscisses : valeur du caractère ;
 - sur l'axe des ordonnées : valeurs des effectifs ou des fréquences.Les hauteurs des différents bâtons sont proportionnelles aux effectifs correspondants.



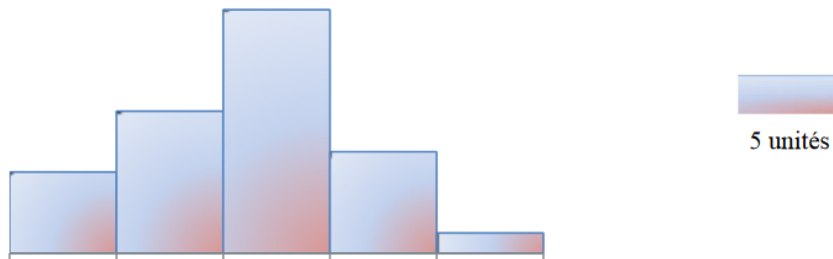
- **Diagramme à secteurs circulaires** : On l'utilise souvent dans le cas d'une variable qualitative ou disposant de peu de valeurs. Chaque secteur a un angle au centre de mesure proportionnelle à la fréquence de la classe correspondante exprimée en pourcentage. Pour ce type de diagramme on obtient la valeur de l'angle au centre α_i en effectuant le calcul :

$$\alpha_i = 360^\circ \times f_i \quad (1)$$

avec α_i = angle au centre de la i^e valeur. f_i = fréquence de la i^e valeur.



- **Histogramme** : On l'utilise pour les séries à caractère quantitatif continu, lorsque les valeurs de la variable sont réparties en classes. Les aires des différents rectangles sont proportionnelles aux effectifs (aux fréquences) correspondantes.



3. Indicateur à tendance centrale

3.1 Le mode

Le **mode** est la valeur de la variable dont l'effectif est le plus grand.

Exemple : Lors d'une course cycliste, on a relevé la vitesse moyenne, en km/h, des 100 participants.

Vitesse (en km/h) x_i	36	37	38	39	40
Nombre de participants n_i	16	28	23	28	5

Dans ce cas, l'effectif le plus important est 28, donc les modes sont 37 et 39. Cela signifie que parmi ces cyclistes, les vitesses 37 km/h et 39 km/h sont les plus fréquentes.

3.2 La moyenne

La **moyenne** \bar{x} est le quotient de la somme des valeurs de la variable par l'effectif total.

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{n_1 + n_2 + \dots + n_p} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N} \quad (2)$$

où x_i désigne le centre de la classe et N l'effectif total.

3.3 La médiane

La **médiane** M_e est la valeur de la variable qui partage la série statistique en deux parties de même effectif.

Exemple : Le relevé du nombre de pièce d'un immeuble de 21 appartements a donné les résultats suivants.

Nombre de pièces x_i	1	2	3	4	5
Nombre d'appartements n_i	2	6	5	7	1

En utilisant la calculatrice on trouve : $\bar{x} \approx 2,95$ et la médiane $Me = 3$. Cela signifie que les appartements de cet immeuble ont en moyenne à peu près trois pièces et que 50% des appartements de cet immeuble ont trois pièces ou moins (ou que 50% des appartements de cet immeuble ont trois pièces ou plus).

4. Indicateurs de dispersion

4.1 Étendue

L'**étendue** est la différence entre la plus grande valeur et la plus petite valeur de la série.

4.2 Quartiles

Les trois **quartiles** sont les trois valeurs du caractère qui partagent la population totale en quatre parties d'effectifs égaux.

Le premier quartile Q_1 correspond à 25% de l'effectif totale.

Le deuxième quartile Q_2 correspond à la médiane (50% de l'effectif totale).

Le troisième quartile Q_3 correspond à 75% de l'effectif totale.

L'**écart** ou **intervalle interquartile** est la différence entre les quartiles extrêmes et a pour valeur : $Q_3 - Q_1$.

4.3 L'écart type

L'**écart type** est un indicateur de la dispersion de la série statistique : c'est la moyenne des écarts autour de la valeur moyenne \bar{x} . On la note sigma : σ .

Pour certaines distributions statistiques, on constate qu'environ 95% des valeurs de la variable appartiennent à l'intervalle $[\bar{x} - 2\sigma; \bar{x} + 2\sigma]$.

Exemple : Dans une entreprise, le contrôle de la quantité de peinture sur un échantillon de 200 pots de contenance 3L (300cL) a donné les résultats suivants.

Classe des quantités de peinture (en cL)	[295;297[[297;299[[299;301[[301;303[[303;305[
Nombre de participants n_i	18	34	64	50	34

Calculons le centre de chaque classe :

Classe des quantités de peinture (en cL)	[295;297[[297;299[[299;301[[301;303[[303;305[
Centre des classes x_i	296	298	300	302	304
Nombre de participants n_i	18	34	64	50	34

En utilisant la calculatrice on trouve : $\bar{x} \approx 300,48$; $\sigma \approx 2,37$; $Q_1 = 298$; $Me = 300$ et $Q_3 = 302$.

Cela signifie que :

- La quantité moyenne de peinture d'un pot est 300,48 cL avec un écart type d'environ 2,37 cL.
- Environ 25% des pots contiennent une quantité de peinture inférieure ou égale à 298 cL.

- Environ 50% des pots contiennent une quantité de peinture inférieure ou égale à 300 cL.
- Environ 75% des pots contiennent une quantité de peinture inférieure ou égale à 302 cL.

5. Diagrammes en boîtes à moustaches

Le **diagramme en boîte en moustaches** permet de visualiser 5 données : le **minimum** et le **maximum**, le **premier** et le **troisième quartile**, la **médiane** d'une série statistiques.

