CHAPITRE III: INDICATEURS STATISTIQUES

I- Introduction

Méthodologie d'étude d'une série statistique:

- En pratique on est gêné en présence d'un grand nombre de données.
- Si l'intégralité de ces valeurs forme l'information complète, il n'est pas aisé de les manipuler ensemble.
- Il faut donc caractériser une variable statistique par un ensemble de paramètres.
- Les plus utilisées sont les caractéristiques : de position, de dispersion, de forme.

II- Caractéristiques de position

Introduction:

- La meilleure façon de résumer un ensemble de données par une seule valeur est de trouver la valeur la plus représentative, celle qui indique où se situe le **centre de la distribution.**

- C'est ce que l'on appelle **la tendance centrale**. Les trois mesures de tendance centrale les plus courantes sont :
  - Le mode
  - La moyenne arithmétique
  - La médiane.

II- Caractéristiques de position

II.1 Mode:

- Soit (x_j, n_j) une série statistique de caractère qualitatif ou quantitatif discret, associée à une population E d'effectif total n.

- **Le mode** est la valeur ou la modalité x_j du caractère qui a **le plus grand effectif**.

Exemple 1:

Les effectifs cumulés et les fréquences cumulées de la série des '*personnes à charge*' sont :

Valeurs x_i	Effectifs n_i	Fréquences f_i	Effectifs cumulés N_i	Fréquences cumulées F_i
0	5	0.10	5	0.10
1	8	0.16	13	0.26
2	15	0.30	28	0.56
3	10	0.20	38	0.76
4	7	0.14	45	0.90
5	5	0.10	50	1
Total	50	1	–	–

Le mode de la série des 'Personnes à charge' est $x_3 = 2$ car max $n_j = 15$.

Graphiquement, le mode 2 est la valeur de la variable associé au plus grand bâton dans le Diagramme en bâtons

II- Caractéristiques de position

II.1 Mode: (suite)

Exemple 2:

Reprenons l'exemple de la série statistique des états civils pour des valeurs prises sur 20 personnes.

x_j	n_j	f_j
C	9	0.45
M	7	0.35
V	2	0.10
D	2	0.10
	$n = 20$	1

Le mode de cette série est x_1= C car max n_j = 9.

II- Caractéristiques de position

II.2 Classe Modale:

- Soit (I_j, n_j) une série statistique de caractère quantitatif continu.
- **La classe modale** I_j est la classe qui a **le plus grand effectif corrigé**.

Exemple 3: La série des “Salaires horaires”

Classes I_i	Effectifs n_i	Amplitudes a_i	Effectifs corrigés n_i^*	Effectifs cumulés N_i
$[47, 52[$	10	5	2	10
$[52, 57[$	30	5	6	40
$[57, 60[$	60	3	20	100
$[60, 63[$	72	3	24	172
$[63, 67[$	40	4	10	212
$[67, 77[$	38	10	3.8	250
Total	250	30	–	–

effectif corrigé: $h_j = n'_j = \dfrac{n_j}{a_j}$

La classe modale de la série des ‘Salaires Horaires est la classe $I_4 = [60, 63[$ car max $n'_j = 24$

Graphiquement, la classe modale est la modalité associée au plus grand rectangle dans l’Histogramme des effectifs corrigés

II- Caractéristiques de position

II.2 Classe Modale: (suite)

Remarque:

1- Lorsque les classes sont de même amplitudes alors la classe modale est la classe qui a le plus grand effectif.

2- Le mode (ou classe modale) n'est pas nécessairement unique : 'Défaut' de cette caractéristique

II- Caractéristiques de position

II.3 Moyenne arithmétique

- C’est une valeur autour de la quelle se repartissent les observations.
- La moyenne peut être calculée pour les séries statistiques à caractère quantitatif
- Une série à caractère qualitatif ne possède pas de moyenne arithmétique.

a) *Pour une série discrète (x_j, n_j):*

La moyenne arithmétique est: $\bar{X} = \dfrac{x_1n_1 + x_2n_2 + \cdots + x_pn_p}{n}$ $\quad avec\ n: effectif\ total$

La formule ci-dessus peut être écrite :

$$\boxed{\bar{X} = \frac{1}{n}\sum_{i=1}^{p} x_in_i} \qquad \boxed{ou\ \bar{X} = \sum_{i=1}^{p} f_ix_i}$$

II- Caractéristiques de position

II.3 Moyenne arithmétique (suite)

b) Pour une série continue (I_j, n_j):

La moyenne arithmétique est définie de la même façon en retenant pour xj les centres des classes Ij:

Exemple 3: La série des “Salaires horaires”

Classes I_i	Effectifs n_i	Amplitudes a_i	Effectifs corrigés n_i^*	Effectifs cumulés N_i
$[47, 52[$	10	5	2	10
$[52, 57[$	30	5	6	40
$[57, 60[$	60	3	20	100
$[60, 63[$	72	3	24	172
$[63, 67[$	40	4	10	212
$[67, 77[$	38	10	3.8	250
Total	250	30	–	–

Les centres x_j:

$$x_1 = \frac{52 + 47}{2} = 49{,}5$$

$$x_2 = 54{,}5$$

$$x_3 = 58{,}5$$

$$x_4 = 61{,}5$$

$$x_5 = 65$$

$$x_6 = 73$$

$$\bar{X} = \frac{49{,}5 * 10 + 54{,}5 * 30 + 58{,}5 * 60 + 61{,}5 * 72 + 65 * 40 + 73 * 38}{250}$$

$$\bar{X} = 61{,}768$$

II- Caractéristiques de position

II.3 Moyenne arithmétique (suite)

Exemple 4 : *calculer la moyenne arithmétique de la série des personnes à charge*

Les effectifs cumulés et les fréquences cumulées de la série des '*personnes à charge*' sont :

Valeurs x_i	Effectifs n_i	Fréquences f_i	Effectifs cumulés N_i	Fréquences cumulées F_i
0	5	0.10	5	0.10
1	8	0.16	13	0.26
2	15	0.30	28	0.56
3	10	0.20	38	0.76
4	7	0.14	45	0.90
5	5	0.10	50	1
Total	50	1	–	–

$$\bar{X} = \frac{0 * 5 + 1 * 8 + 2 * 15 + 3 * 10 + 4 * 7 + 5 * 5}{50}$$

$\bar{X} = 2{,}42$ Un salarié a en moyen environ 2.42 personnes à sa charge

II.3 Moyenne arithmétique (suite)

c) Propriétés de la moyenne arithmétique

- *La moyenne arithmétique est associative*

  - La moyenne globale d'une variable statistique sur l'agrégation de plusieurs populations est la moyenne pondérée des moyennes partielles.
  - Soit Pa et Pb deux populations d'effectifs na et nb, x une v. s. de moyennes arithmétiques $\bar{X}_a$et $\bar{X}_b$sur Pa et Pb respect.
  - La moyenne arithmétique de x sur l'agrégation des deux populations P= Pa U Pb est:

$$\bar{X} = \frac{n_a\bar{X}_a + n_b\bar{X}_B}{n_a + n_B}$$

L'intérêt pratique: La mise à jour facile de la moyenne dans le cas d'ajout d'une (des) observation(s)

II- Caractéristiques de position

II.3 Moyenne arithmétique (suite)

c) Propriétés de la moyenne arithmétique

- *La moyenne arithmétique est associative*

Exemple 5
Le salaire moyen des 8 employés d'une entreprise est $\bar{X} = 26585$
Si l'entreprise recrute deux nouveaux employés qualifiés dont le revenu moyen est 100 000 Dh

Le nouveau revenu moyen des employés $\overline{X'}$ est:

$$\bar{x}' = \frac{8 \times 26585 + 2 \times 100000}{8+2} = \boxed{41\ 268\ \text{Dh}}$$

II- Caractéristiques de position

II.3 Moyenne arithmétique (suite)

c) Propriétés de la moyenne arithmétique

- *La moyenne arithmétique est sensible à la présence des valeurs aberrantes*

  - Une valeur **aberrante** est une valeur qui n'est pas du même ordre de grandeur que la plus part des autres observations

Exemple:

Échantillon $\{1, 1, 2, 2, 2, 2, 3, 3\} \longrightarrow \bar{x} = 2$

Échantillon $\{1, 1, 2, 2, 2, 2, 3, 300\} \longrightarrow \bar{x} = 39.125.$

Cette propriété est en faite un 'défaut' de la moyenne arithmétique

II- Caractéristiques de position

II.4 Médiane:

a) Définition:

- La **médiane** est une valeur du caractère, notée $x_{1/2}$ (ou Me), partageant une série ordonnée en deux sous-ensembles à tailles égales (d'effectifs égaux.).
- La médiane peut être calculée pour les séries statistiques à **caractère quantitatif ou qualitatif ordinal**
- Une série à caractère qualitatif nominal ne possède pas de médiane.

II.4 Médiane:

b) Médiane d'une série discrète (x_j, n_j) et d'effectif total n:

- S'il existe une valeur x_j telle que son effectif cumulé $N_j = n/2$ alors: $x_{1/2} = \dfrac{x_{j-1} + x_j}{2}$

- Sinon, la valeur médiane est la plus petite valeur x_j dont l'effectif cumulé dépasse la moitié de l'effectif total:

$$x_{1/2} = x_j \Leftrightarrow N_{j-1} < \frac{n}{2} < N_j$$

II- Caractéristiques de position

On reprend l'exemple 4 : calculer la médiane de la série des personnes à charge

Les effectifs cumulés et les fréquences cumulées de la série des '*personnes à charge*' sont :

Valeurs x_i	Effectifs n_i	Fréquences f_i	Effectifs cumulés N_i	Fréquences cumulées F_i
0	5	0.10	5	0.10
1	8	0.16	13	0.26
2	15	0.30	28	0.56
3	10	0.20	38	0.76
4	7	0.14	45	0.90
5	5	0.10	50	1
Total	50	1	–	–

$n = 50$

Le premier effectif cumulé qui dépasse n/2 =25 est $N_2 = 28$

$\Rightarrow x_{1/2} = x_3 = 2$

Dans cet exemple, la médiane (2) est un peu plus petite que la moyenne (2,42).

II- Caractéristiques de position

II.4 Médiane:

Avantage:

- L’avantage d’utiliser la médiane plutôt que la moyenne est qu’elle est plus robuste aux valeurs extrêmes qui pourraient surgir à l’une des extrémités de la distribution.
- Il est donc important de vérifier si les données comptent des valeurs extrêmes avant de choisir quelle mesure de tendance centrale doit être utilisée.

II- Caractéristiques de position

Exercice 1:

Nous avons demandez aux 30 élèves d'une classe combien de personnes vivent dans leur ménage. Nous avons obtenu le tableau des effectifs suivant:

Valeurs x_j	Effectifs n_j
2	3
3	4
4	10
5	4
6	2
7	3
8	1
9	2
10	1

a) Compléter le tableau des effectifs (fréq, eff. Cum., fréq, cum.)
b) Dessiner sa *Courbe cumulative*.
c) Calculer la moyenne arithmétique
d) Calculer la médiane
e) Un nouvel élève est récemment inscrit en classe. Son ménage compte 18 résidents. Calculer la nouvelle moyenne et médiane.
f) Laquelle des ces 2 caractéristiques est la plus robuste aux valeurs extrêmes?

II- Caractéristiques de position

II.4 Médiane:

c) Médiane d'une série continue:

- La détermination de la médiane $x_{1/2}$ se fait en 2 étapes:

1) Détermination de la classe médiane $I_m = [x_m^-,\ x_m^+[$: c'est la 1ère classe dont l'éffectif cumulé dépasse n/2

2) Détermination de la valeur médiane par une **interpolation linéaire** suivant la formule suivante:

$$\boxed{x_{1/2} = x_m^- + (x_m^+ - x_m^-) \times \frac{n/2 - N_{m-1}}{N_m - N_{m-1}}}$$

- **Exemple 3:** Considérons par exemple la distribution des salaires observés en continu:

Salaires	Effectifs n_i	Effectifs cumulés N_i
$[1000, 2000[$	20	20
$[2000, 4000[$	50	70
$[4000, 6000[$	30	100
Total	100	–

la classe médiane est $I_2 = [2000, 4000[$ ayant N2 =70 > 50

$$\Rightarrow x_{1/2} \in I_m = I_2$$

la médiane de la série des salaires est: $x_{1/2} = 2000 + (4000 - 2000) \times \frac{50 - 20}{70 - 20} = \boxed{3\ 200 \text{ Dh}}$

II- Caractéristiques de position

II.4 Médiane:

c) Médiane d'une série continue: (suite)

- La détermination de la médiane $x_{1/2}$ peut se faire graphiquement:

1) On trace la courbe cumulative N(x) (une fonction linéaire par morceaux)
2) On localise sur l'axe des abscisses la valeur correspondante à la valeur n/2 sur l'axe des ordonnées ; cette valeur est la médiane

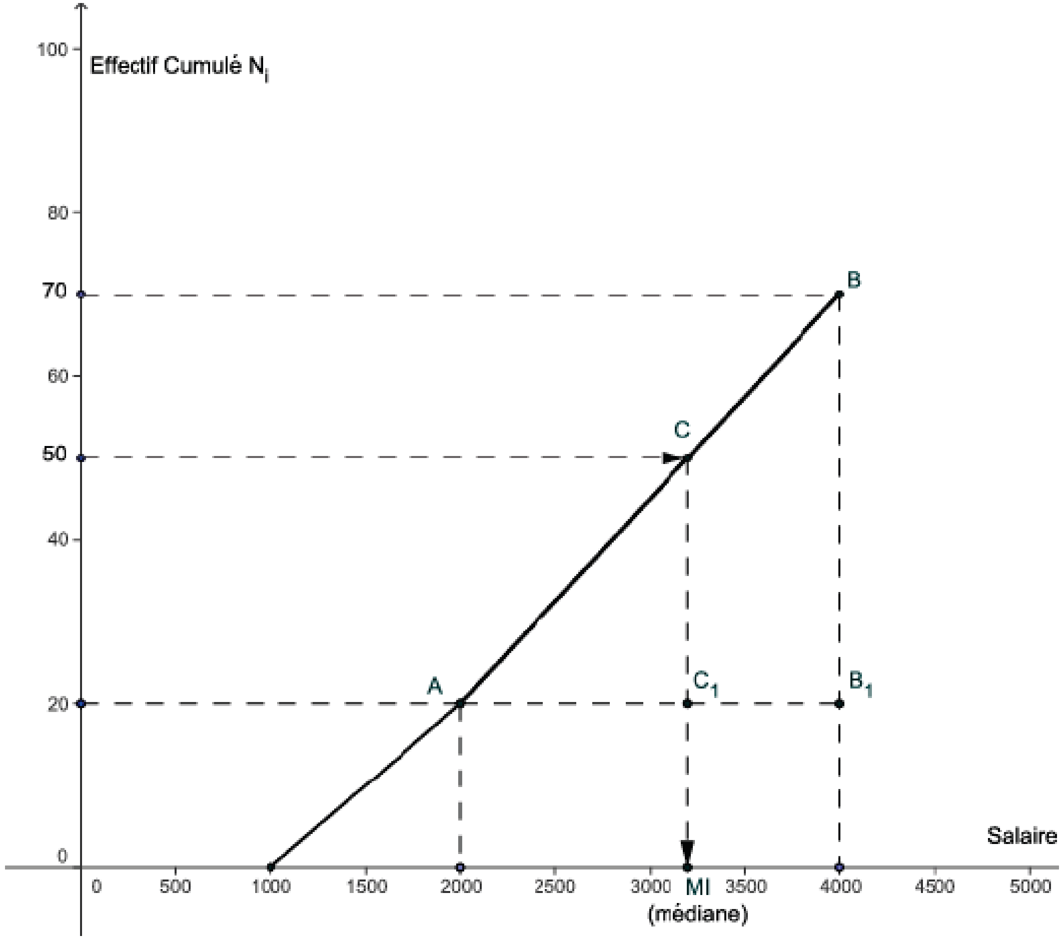
Courbe cumulative:

On interpole les effectifs cumulés par une ligne brisée joignant les points d'abscisses (x_j^+, Nj)

II- Caractéristiques de position

- **Exemple 3:** Considérons par exemple la distribution des salaires observés en continu:

Salaires	Effectifs n_i	Effectifs cumulés N_i
$[1000, 2000[$	20	20
$[2000, 4000[$	50	70
$[4000, 6000[$	30	100
Total	100	–



II- Caractéristiques de position

II.5 Quantiles:

- On généralise l'idée de médiane, en utilisant la notion de quantile où l'on va partager la distribution en plusieurs sous distributions de même taille.

- On appelle **quantile d'ordre α** (où $0 < \alpha < 1$) la valeur x_α de la variable telle que au moins (α *100) % des observations sont inférieures ou égales à x_α.

- La médiane est un quantile particulier d'ordre $\alpha = 0.5$

a) Quartiles:

- Pour une série statistique dont les valeurs sont classées par ordre croissant, la médiane partage la série en deux parties de même effectif.

- On peut aussi partager en **4 parties de même effectif**. On obtient ainsi trois valeurs **Q_1, Q_2, Q_3 appelés quartiles.**

b) Déciles:

- Lorsque l'étude porte sur une population très importante, on utilise souvent **les déciles** qui partagent la population en **10 parties de même effectif.**

II.5 Quantiles: (suite)

- Résumons les quantiles usuels :

Quantiles	Ordres α	Notations	Sous-groupes
Médiane	0.5	Me	$2 \times 50\%$
Quartiles	$(0.25, 0.50, 0.75)$	(Q_1, Q_2, Q_3)	$4 \times 25\%$
Deciles	$(0.10, 0.20, \ldots, 0.90)$	$(D_1, \ldots, D_9)$	$10 \times 10\%$
Centiles	$(0.01, 0.02, \ldots, 0.99)$	$(C_1, \ldots, C_{99})$	$100 \times 1\%$

Remarques:

- *Le premier quartile Q_1 est la plus petite valeur de la série telle que 25% des valeurs de la série lui soient inférieures ou égales.*

- *Le deuxième quartile Q_2 est la médiane*

- *Le troisième quartile Q_3 est la plus petite valeur de la série telle que 75% des valeurs de la série lui soient inférieures ou égales.*

II- Caractéristiques de position

II.5 Quantiles: (suite)

Exemple 6: Observons la série statistique des lancers de javelot suivante :

Longueur (en m)	37	39	40	41	42	43	44	48
Effectif	4	3	4	2	2	4	5	2
Effectif cumulés	4	7	11	13	15	19	24	26

-Le premier quartile Q1: Le premier effectif cumulé qui dépasse n/4 =26/4=6,5 est $N_2 = 7$

Q1=39m

- Le 2ème quartile Q1 ou La médiane: Me = $40, 5m$.

- Le 3ème quartile Q3 Le premier effectif cumulé qui dépasse 3*n/4 =3*6,5=19,5 est $N_7 = 24$

Q3=44m

Enfin, l'intervalle $[Q_1;Q_3]$ = [39;44] contient 50% de la population.

II- Caractéristiques de position

II.5 Quantiles: (suite)

c) Cas d'une série continue

La détermination du quantile x_α est identique à celle de la médiane :

1. Détermination de la classe $I_m := [x_m^-, x_m^+[$ qui inclue le quantile: c'est la première classe dont l'effectif cumulé dépasse $\alpha * n$
2. Détermination du quantile d'ordre α par l'interpolation linéaire :

$$x_\alpha = x_m^- + (x_m^+ - x_m^-) \times \frac{\alpha\, n - N_{m-1}}{N_m - N_{m-1}}.$$

II- Caractéristiques de position

II.5 Quantiles: (suite)

$$x_\alpha = x_m^- + (x_m^+ - x_m^-) \times \frac{\alpha n - N_{m-1}}{N_m - N_{m-1}}.$$

c) Cas d'une série continue

- **Exemple 3:** Considérons par exemple la distribution des salaires observés en continu:

Salaires	Effectifs n_i	Effectifs cumulés N_i
$[1000, 2000[$	20	20
$[2000, 4000[$	50	70
$[4000, 6000[$	30	100
Total	100	–

Déterminons les **quartiles** Q1,Q2,Q3

- Premier quartile $Q_1 = x_{1/4}$: Le premier effectif cumulé qui dépasse n/4 =100/4=25 est $N_2 = 70$

Alors $Q_1 \in I_2 = [2000, 4000[$ $$Q_1 = 2000 + (4000 - 2000) \times \frac{25 - 20}{70 - 20} = \boxed{2\,200}.$$

- troisième quartile $Q_3 = x_{3/4}$: Le premier effectif cumulé qui dépasse 3*n/4 =75 est $N_3 = 100$

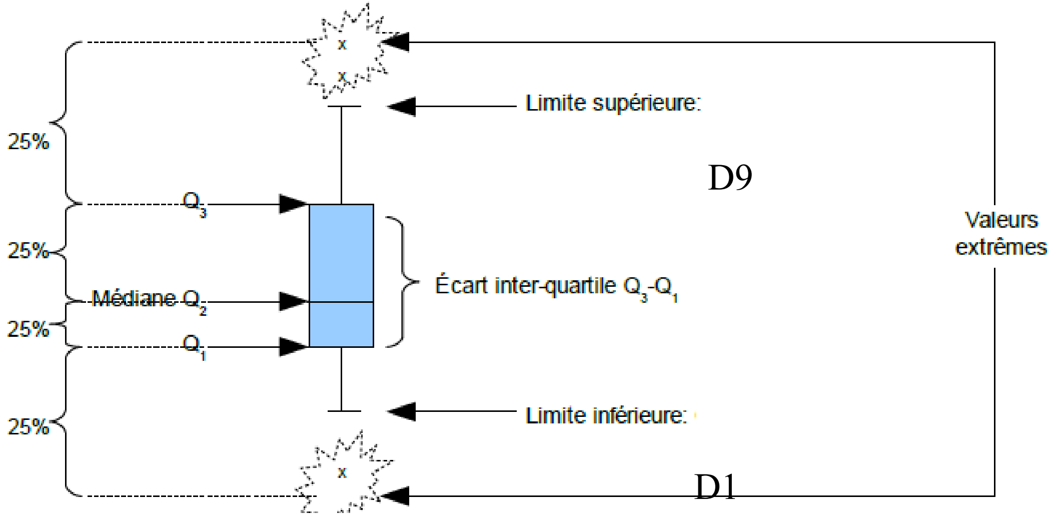
Alors $Q_3 \in I_3 = [4000, 6000[$ $$Q_3 = 4000 + (6000 - 4000) \times \frac{75 - 70}{100 - 70} \approx \boxed{4\,333.33}.$$

II- Caractéristiques de position

II.6 Diagramme de Tukey

- Un **diagramme de Tukey** (aussi appelé « **boîte à moustache** ») est un résumé, sur un axe gradué, des quantiles.

- Ce diagramme est constitué:
  1- d'une boîte (dont la hauteur est prise de manière arbitraire) délimitée par le 1er et 3ème quartiles (Q1 et Q3).
  2- Cette même boite est ensuite partagée par la médiane.
  3- Deux segments « moustaches » sortent des deux côtés du rectangle et sont délimitées par les valeurs D1 et D9



II.7 Conclusion

Le tableau ci-dessous résume les paramètres de position envisageable en fonction du type de la variable étudiée :

	Quantitative	Ordinale	Nominale
Moyenne	OUI	NON	NON
Médiane	OUI	OUI	NON
Quantiles	OUI	OUI	NON
Modes	OUI	OUI	OUI