Séance 4

Les paramètres statistiques de centralité

Objectifs de la séance

Une série numérique peut être résumée par deux paramètres statistiques :

- le **centre** d'une distribution des valeurs, représentant leur tendance d'ensemble;

- La **dispersion** des valeurs, représentant leur variabilité.

Exemple introductif

2 élèves obtiennent les 5 notes suivantes aux examens :

Pierre : 10, 11, 11, 10, 12

Paul : 4.5, 20, 4.5, 20, 5

Moyenne de Pierre = (10+11+11+10+12)/5= 10,8

Moyenne de Paul = (4.5+20+4.5+20+5)/5= 10,8

La moyenne est une mesure de la tendance centrale d'une distribution

La moyenne apporte une information incomplète sur une distribution car les deux élèves, qui ont la même moyenne, ont deux profils très différents : le premier est régulier, le second très irrégulier. C'est pourquoi l**es mesures de tendance centrale doivent toujours être accompagnées d'une mesure de dispersion.**

Amplitude des notes du 1^{er} élève = 12 - 10 = 2

Amplitude des notes du $2^{ème}$ élève = 20 - 4.5 = 15,5

L'étendue est une mesure de la dispersion (de la variabilité) d'une distribution

➢ MOYENNE ARITHMÉTIQUE (MEAN)

C'est le paramètre de tendance centrale le plus utilisé.
Il peut être un résumé de la distribution d'un caractère quantitatif

Notation:

$$\overline{X}$$

Calcul sur un tableau complet :

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \Rightarrow \boxed{\overline{X} = \frac{1}{n}\sum_{i=1}^{i=n} X_i}$$

Calcul sur un tableau condensé

$$\overline{X} = \frac{n_1 c_1 + n_2 c_2 + \ldots + n_k c_k}{n} \Rightarrow \boxed{\overline{X} = \frac{1}{n}\sum_{j=1}^{j=k} n_j c_j}$$

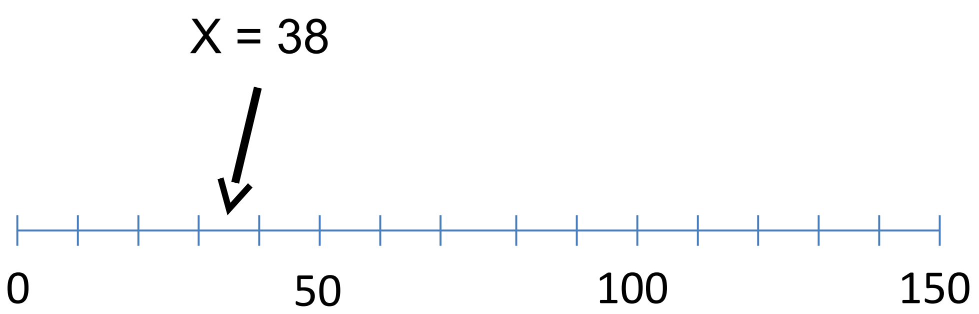
Ou nj = l'effectif de la classe
cj = le centre de la classe

MOYENNE ARITHMÉTIQUE (MEAN)

Propriétés :

- \- La moyenne est le **centre de gravité** d'une distribution;
  \- La moyenne arithmétique est très sensible aux valeurs extrêmes.

Considérons la série statistique suivante : 10;10;10;10;150



Alors que l'essentiel des valeurs est 10, la moyenne est de 38. Forte sensibilité à la valeur extrême 150.

MOYENNE ARITHMÉTIQUE (MEAN)

Calcul sur un tableau complet

Salariés	Salaires mensuels nets (€)
Dupond	2400
Claude	1350
Garisson	1800
Toto	4500
Martin	4900
Steen	1350
Jefferson	1600
Douglas	1500
Bryan	2400
Marteau	1500
Pertus	2000
Carrière	1300
Bistouri	1700
Birhut	1900
Vasquez	1500
Urena	5000
Ndione	1820
Pauli	1350
Sanchez	5000
Muller	2000
Norma	4900

$$\overline{X} = \frac{2400 + 1350 + 1800 + \ldots + 4900}{21} = 2465$$

Calcul sur un tableau condensé

Salaire	Effectifs	Centre de la classe
[1300 ; 1600[	7	(1300+1600)/2=1450
[1600 ; 1900[	4	(1600+1900)/2=1750
[1900; 2400[	5	(1900+2400)/2=2150
[2400; 5000]	5	(2400+5000)/2=3700
Total	21	

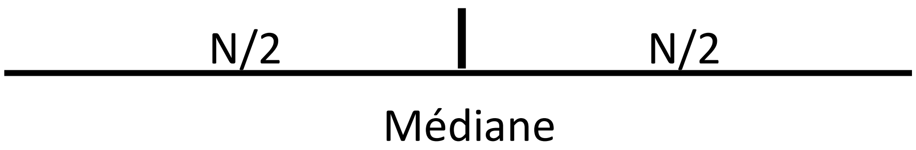
$$\overline{X} = \frac{7*1450 + 4*1750 + 5*2150 + 5*3700}{21}$$

$$\overline{X} = 2210$$

LA MÉDIANE (MEDIAN)

Elle peut être calculée sur des caractères qualitatifs ordinaux ou quantitatifs

Définition : La médiane est la valeur telle que la moitié des valeurs lui est inférieure et l'autre moitié supérieure.



Notation: Q_2

Propriétés : La médiane est déterminée par le classement des valeurs. Elle est donc peu sensible aux valeurs extrêmes et résume bien les distributions fortement dissymétriques.

Considérons la série statistique suivante : 10;10;10;10;150

Aucune sensibilité à la valeur 150

➢ LA MÉDIANE (MEDIAN)

Calcul sur un tableau statistique complet

Salariés	Salaires mensuels nets (€)
Carrière	1300
Claude	1350
Steen	1350
Pauli	1350
Douglas	1500
Marteau	1500
Vasquez	1500
Jefferson	1600
Bistouri	1700
Garisson	1800
Ndione	1820
Birhut	1900
Pertus	2000
Muller	2000
Dupond	2400
Bryan	2400
Toto	4500
Martin	4900
Norma	4900
Urena	5000
Sanchez	5000

1er cas : n est impair

Q_2 est une valeur de la variable de rang : (n+1)/2

10 valeurs inférieures

(21+1)/2 = 11ème valeur de la série

Q2=1820

10 valeurs supérieures

LA MÉDIANE (MEDIAN)

Calcul sur un tableau statistique complet

Salariés	Salaires mensuels nets (€)
Carrière	1300
Claude	1350
Steen	1350
Pauli	1350
Douglas	1500
Marteau	1500
Vasquez	1500
Jefferson	1600
Bistouri	1700
Garisson	1800
Ndione	1820
Birhut	1900
Pertus	2000
Muller	2000
Dupond	2400
Bryan	2400
Toto	4500
Martin	4900
Norma	4900
Urena	5000

10 valeurs inférieures

Q2=(1800+1820)/2=1810

10 valeurs supérieures

2ème cas : n est pair

Q_2 Correspond au milieu de l'intervalle entre les valeurs de rangs n/2 et (n+1)/2

20/2 = 10ème valeur de la série
(20+1)/2 = 10,5 on arrondit au dessus : 11ème valeur

➢ LA MÉDIANE (MEDIAN)

Calcul sur un tableau statistique condensé

Salaire	Effectifs	Fréquences simples	Fréquences cumulées
[1300 ; 1600[	7	0,33	0,33
[1600 ; 1900[	4	0,19	0,52
[1900; 2400[	5	0,24	0,76
[2400; 5000]	5	0,24	1,00
Total	21	1,00	

On commence par chercher la classe comprenant la fréquence cumulée 50 % = **[1600;1900[**

X_j = **1600**

F_{j-1} = **0,33**

f_j = **0,19**

a_j = **1900-1600 = 300**

$$Q_2 = X_{j.} + \frac{0,5 - F_{j-1}}{f_j} \text{ x } a_j$$

$X_{j.}$ Valeur de la borne inférieure de la classe

F_{j-1} Fréquence cumulée de la classe j-1

f_j Fréquence simple de la classe j (contenant la fréquence cumulée 0,5)

a_j Amplitude de la classe j ((contenant la fréquence cumulée 0,5)

$$Q_2 = 1600 + \frac{(0,5-0,33)}{0,19} \text{x } 300 \quad = 1868$$

LE MODE OU LA CLASSE MODALE

Il peut être déterminé pour des caractères de toute nature (quantitatif, qualitatif nominal ou ordinal).

Définition :
Le mode d'un caractère quantitatif discret est la valeur la plus fréquente.
Dans une distribution connue par classes ou catégories:
- Si le caractère est quantitatif continu, la classe modale est celle de plus grande densité de fréquence;
- Si le caractère est qualitatif, c'est la modalité la plus fréquente

Notation: Mo

Propriétés : C'est un résumé assez pauvre d'une distribution mais le seul disponible pour les variables qualitatives nominales

LE MODE OU LA CLASSE MODALE

Variable quantitative discrète (valeur la plus fréquente)

Âge (xj)	Effectifs (nj)
20	5
24	2
25	3
28	2
36	2
40	4
51	1
60	2
Total	21

Mode (Mo) = 20

Variable qualitative (modalité la plus fréquente)

Locomotion	Effectifs (nj)
Voiture	12
Transport en commun	4
2 roues	2
A pied	3
Total	21

Classe modale = Voiture

Variable quantitative continue connue par des **classes d'égale amplitude** (classe possédant l'effectif le plus important)

Âge (xi)	Effectifs (ni)
[20; 30[	12
[30; 40[	2
[40; 50[	4
[50;60]	3
Total	21

Classe modale (Mo) = [20;30[

Variable quantitative continue connue par des **classes d'inégale amplitude** (classe possédant la densité de fréquences la plus importante)

Âge	Effectifs (nj)	Fréquence simples (fj)	Amplitude de la classe (aj)	Densité de fréquences (dfj)
[10; 20[	10	0,019	10	0,0019
[20; 30[	40	0,077	10	0,0077
[30; 50[	220	0,423	20	0,0212
[50; 90[	240	0,462	40	0,0115
[90; 100[	10	0,019	10	0,0019
Total	520	1,000		

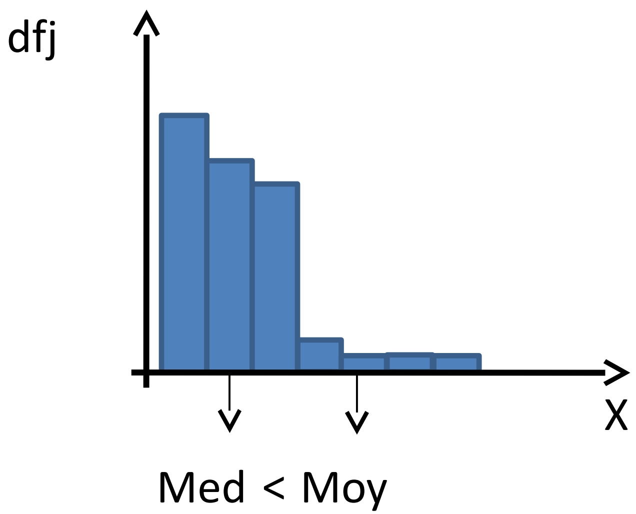
Classe modale (Mo) = [30;50[

1) Si médiane < moyenne : la moyenne est influencée par les fortes valeurs de X. La distribution est **dissymétrique à gauche**.

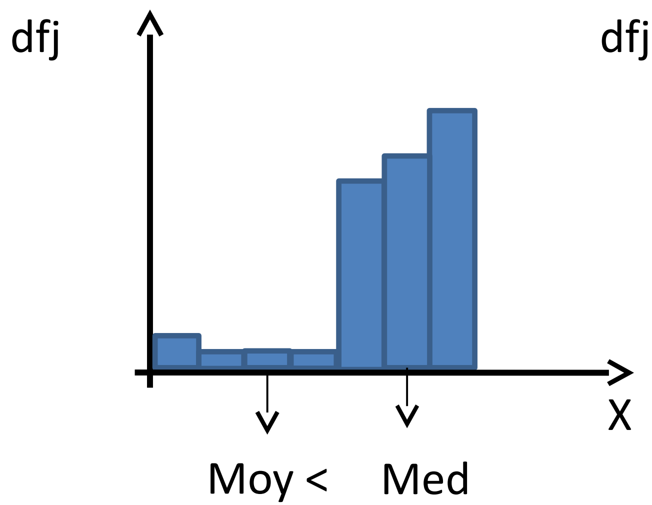
2) Si médiane > moyenne : la moyenne est influencée par les faibles valeurs de X. **La distribution est dissymétrique à droite.**

3) Si médiane = moyenne, le mode l'est aussi (sauf pour des distributions à plusieurs modes) et **la distribution est symétrique**.

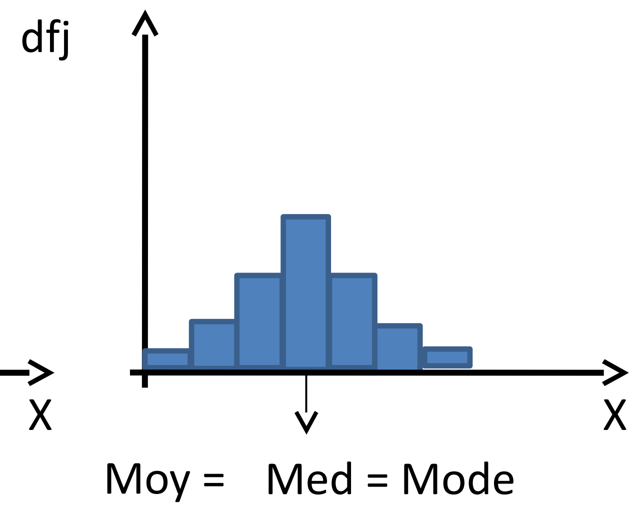
1) Dissymétrie à gauche



2) Dissymétrie à droite



3) Distribution symétrique



	Nom	PIB/Hab	Population ou Effectifs (nj)
1	Moldavie	3698	3583288
2	Ukraine	7618	44291413
3	Bosnie-Herzégovine	8307	3871643
4	Albanie	8850	3020209
5	Macédoine	10790	2091719
6	Serbie	11161	7209764
7	Monténégro	11429	650036
8	Roumanie	12918	21729871
9	Bulgarie	15105	6924716
10	Biélorussie	15654	9608058
11	Croatie	17649	4470534
12	Russie	17920	142470272
13	Lettonie	17952	2165165
14	Chypre	18440	1172458
15	Lituanie	19234	3505738
16	Hongrie	19820	9919128
17	Pologne	21228	38346279
18	Portugal	22499	10813834
19	Estonie	23801	1257921
20	Slovaquie	24506	5443583
21	Grèce	24788	10775557
22	République tchèque	26874	10627448
23	Malte	27771	412655
24	Slovénie	28416	1988292
25	Espagne	29096	47737941
26	Italie	29264	61680122
27	France	34305	66259012
28	Andorre	37012	85458
29	Finlande	37105	5268799
30	Royaume-Uni	37306	63742977
31	Danemark	37942	5569077
32	Irlande	39398	4832765
33	Allemagne	39841	80996685
34	Saint-Marin	39888	32742
35	Belgique	40357	10449361
36	Suède	40499	9723809
37	Pays-Bas	41256	16877351
38	Islande	41311	317351
39	Autriche	43901	8223062
40	Suisse	45934	8061516
41	Norvège	54820	5147792
42	Luxembourg	81952	520672
43	Liechtenstein	85761	37313
44	Monaco	188410	30508

Sur l'exemple « PIB/Hab des pays d'Europe »

1/ Calculer graphiquement l'indice de GINI

2/ Sans prendre en compte la population :
calculer moyenne (ou PIB moyen des pays d'Europe), médiane et mode.

3/ En prenant en compte la population :
calculer la moyenne (ou PIB moyen d'un européen).

4/ Comparer les deux moyennes