

Séance 5

Les paramètres statistiques
de dispersion

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.

➤ ÉTENDUE OU AMPLITUDE (RANGE)

Définition: Différence entre la plus grande et la plus petite des observations d'une série statistique *

* : suppose que la variable soit quantitative ou pseudo-quantitative

Notation: ω

Calcul:

$$\omega = X_{\max} - X_{\min}$$

Limite : la dispersion est mesurée en prenant en compte seulement les deux valeurs extrêmes et non pas toutes les valeurs de la distribution

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

$$\omega = 60 - 20 = 40 \text{ ans}$$

➤ VARIANCE (VARIANCE)

Définition: Moyenne des carrés des écarts à la moyenne. Plus cette valeur est importante et plus la dispersion est importante.

Notation: Var ou σ^2

Calcul:

Sur la population entière

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

Sur un échantillon

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Sur un tableau statistique complet

Sur un tableau statistique condensé

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^{j=k} (c_j - \bar{x})^2 n_j$$

$$\sigma^2 = \frac{1}{n-1} \sum_{j=1}^{j=k} (c_j - \bar{x})^2 n_j$$

➤ VARIANCE (VARIANCE)

Calcul sur un tableau statistique complet

Salariés	Salaires mensuels nets (X_i)	$(x_i - \bar{x})^2$
Carrière	1300	$(1300-2465)^2= 1357780$
Claude	1350	$(1350-2465)^2=1243756$
Steen	1350	1243756
Pauli	1350	1243756
Douglas	1500	931685
Marteau	1500	931685
Vasquez	1500	931685
Jefferson	1600	748637
Bistouri	1700	585589
Garrison	1800	442542
Ndione	1820	416332
Birhut	1900	319494
Pertus	2000	216446
Muller	2000	216446
Dupond	2400	4256
Bryan	2400	4256
Toto	4500	4140256
Martin	4900	5928066
Norma	4900	5928066
Urena	5000	6425018
Sanchez	5000	6425018
		39684524

Sur la population entière
 $\sigma^2 = 39684524/21 = 1889739$

Sur un échantillon
 $\sigma^2 = 39684524/(21-1) = 1984226$

$$\bar{X} = 2465$$

➤ ÉCART-TYPE (STANDARD DEVIATION)

Définition : Racine carrée (positive) de la variance

Notation : σ

σ = moyenne quadratique des écarts à la moyenne.

σ , comme la moyenne, est exprimé dans la même unité que la variable d'origine.

Calcul :
$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

➤ **COEFFICIENT DE VARIATION (COEFFICIENT OF VARIATION)**

Définition : écart-type divisé par la moyenne

Notation : c_v

c_v est sans dimension (unité).

Ainsi est-il possible de comparer les C.V. de variables exprimées dans des unités différentes

Calcul :
$$c_v = \frac{\sigma}{\mu}$$

➤ COEFFICIENT DE VARIATION (COEFFICIENT OF VARIATION)

Définition : écart-type divisé par la moyenne

Notation : c_v

c_v est sans dimension (unit c_v)
Ainsi est-il possible de comparer les C.V. de variables exprimées dans des unités différentes

Calcul :
$$c_v = \frac{\sigma}{\mu}$$

!!! : Quand la moyenne tend vers zéro, le coefficient de variation tend vers l'infini

➤ **QUARTILES** (Q1; Q2; Q3)

Définition: Les 3 valeurs (Q1; Q2; Q3) divisant la série triée en 4 parties d'effectifs égaux tel que chaque partie représente $\frac{1}{4}$ de l'échantillon de la population;

25 % des valeurs \leq Q1

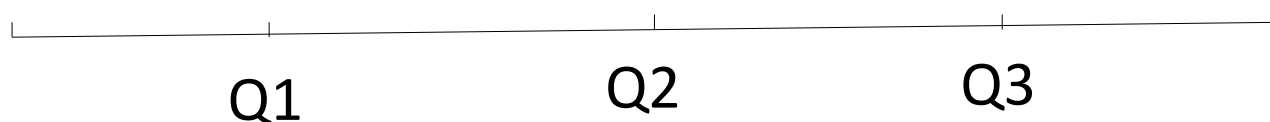
25 % comprises entre Q1 et Q2

25 % entre Q2 et Q3

et 25 % supérieures à Q3 .

L'intervalle interquartile est égal à (Q3-Q1)

=> 50 % des individus sont compris dans cet intervalle



➤ QUARTILES

Calcul sur un tableau statistique complet

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

- 1ere étape, on classe les individus par ordre croissant

- 2ème étape, on cherche le rang en calculant : $n \times F$

n désigne le nombre d'individus et F la fréquence cumulée recherchée (0,25 pour Q1 et 0,75 pour Q3)

Exemple pour Q1

$$21 \times 0,25 = 5,25$$

Le premier quartile est le terme de rang 6

$$Q1 = 1500$$

Exemple pour Q3

$$21 \times 0,75 = 15,75$$

Le troisième quartile est le terme de rang 16

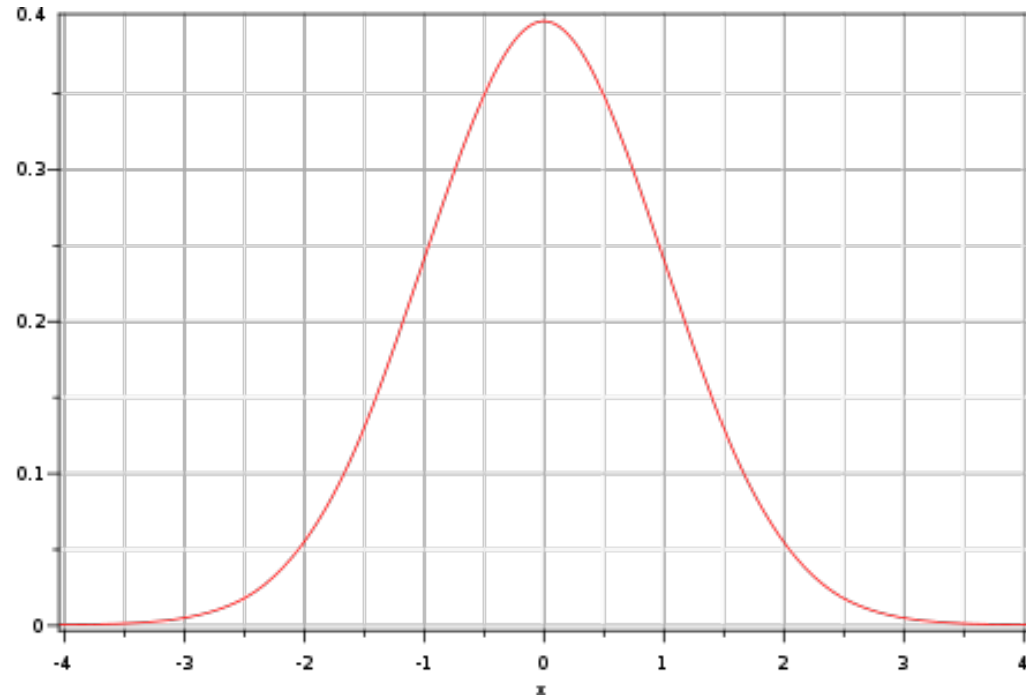
$$Q3 = 2400$$

➤ DISTRIBUTION GAUSSIENNE et LOI NORMALE

En théorie des probabilités et en statistique, la loi normale est l'une des lois de probabilité les plus adaptées pour modéliser des phénomènes naturels issus de plusieurs événements aléatoires.

Une variable répondant à la loi normale présente une distribution caractéristique. Sur l'histogramme on note :

- la forme en cloche ;
- la proximité mode/moyenne/médiane ;
- des conditions relative à dispersion qui s'expriment en fonction de la Moyenne et de la médiane.



$$\mathbb{P}(\mu - \sigma \leq x \leq \mu + \sigma) \approx 0.6827$$

$$\mathbb{P}(\mu - 2\sigma \leq x \leq \mu + 2\sigma) \approx 0.9545$$

$$\mathbb{P}(\mu - 3\sigma \leq x \leq \mu + 3\sigma) \approx 0.9973$$

	Nom	PIB/Hab	Population ou Effectifs (n _j)
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 », sans prendre en compte la population :

1/ Calculer l'étendue, la variance, l'écart-type et le coefficient de variation

2/ Calculer les quartiles et l'intervalle interquartile

3/ Quel est le % d'individus compris dans les intervalles suivants :

$Q3 - Q1$

$$\mathbb{P}(\mu - \sigma \leq x \leq \mu + \sigma) \approx 0.6827$$

$$\mathbb{P}(\mu - 2\sigma \leq x \leq \mu + 2\sigma) \approx 0.9545$$

$$\mathbb{P}(\mu - 3\sigma \leq x \leq \mu + 3\sigma) \approx 0.9973$$

4/ La distribution est-elle gaussienne ?