

Objectif On dispose d'une grande série de données (10 ; 100 ; 1 000 000 ; ...), ou série statistique, d'un caractère dans une population.

Cet ensemble de données n'est pas humainement compréhensible ou interprétable.

L'objectif des statistiques descriptives est de réduire ce "grand ensemble" de données à seulement quelques valeurs pertinentes et qui permettent à elles seules de décrire, ou caractériser, le grand ensemble de départ.

I Description par la moyenne et l'écart type

Exercice Six élèves ont obtenu les notes suivantes. Ce sont six séries statistiques. Calculer la moyenne de chacune de ces séries.

Elève A : 8 - 7 - 10 - 14 - 12 - 9

Elève B : 9 - 9,5 - 10 - 10,5 - 10,5 - 10,5

Elève C : 10 - 10 - 10 - 10 - 10 - 10

Elève D : 0 - 0 - 0 - 20 - 20 - 20

Elève E : 5 - 9 - 15 - 4 - 11 - 16

Elève F : 10

Tous ces élèves ont la même moyenne $\bar{x} = 10$, ce qui montre bien que si la moyenne est une caractéristique statistique descriptive et pertinente, elle est insuffisante pour caractériser la série complète. On a donc besoin de compléter cette caractéristique par une mesure de la dispersion des valeurs autour de cette moyenne.

Cadre général : Série statistique

Valeurs (x_i)	x_1	x_2	x_3	...	x_p
Effectifs (n_i)	n_1	n_2	n_3	...	n_p

L'effectif total est : $N = n_1 + n_2 + \dots + n_p$.

Exemple : Elève G

Notes (x_i)	7	10	4	2	14
coefficients (n_i)	2	2	1	1	5

L'effectif total est : $N = 2 + 2 + 1 + 1 + 5 = 11$.

Définition: La moyenne de la série, notée en général \bar{x} est : $\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{N}$.

La variance de la série est la moyenne des carrés des écarts à la moyenne :

$$V = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{N}$$

L'écart type, noté σ , est alors : $\sigma = \sqrt{V}$.

Exercice Calculer la moyenne, la variance et l'écart type des notes de l'élève G.

Algorithme de calcul de la moyenne d'une série

Moyenne simple (sans coefficient)

```

Algorithme
Entrer N
0 → S
Pour I de 1 à N
    Entrer X
    S+X → S
Fin Pour
S/N → M
Afficher M
    
```

Moyenne pondérée (coefficientée)

```

Algorithme
Entrer N
0 → S
0 → T
Pour I de 1 à N
    Entrer X
    Entrer C
    S+C*X → S
    T+C → T
Fin Pour
S/T → M
Afficher M
    
```

Propriété: La variance est la moyenne des carrés moins le caré de la moyenne :

$$V = \overline{x^2} - \bar{x}^2$$

Démonstration :

$$\begin{aligned}
 V &= \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{N} \\
 &= \frac{n_1(x_1^2 - 2x_1\bar{x} + x_1^2) + n_2(x_2^2 - 2x_2\bar{x} + x_2^2) + \dots + n_p(x_p^2 - 2x_p\bar{x} + x_p^2)}{N} \\
 &= \underbrace{\frac{n_1x_1^2 + n_2x_2^2 + \dots + n_px_p^2}{N}}_{\overline{x^2}} - 2\bar{x} \underbrace{\frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N}}_{\bar{x}} + \bar{x}^2 \underbrace{\frac{n_1 + n_2 + \dots + n_p}{N}}_{\frac{N}{N}=1} \\
 &= \overline{x^2} - 2\bar{x} \times \bar{x} + \bar{x}^2 \\
 &= \overline{x^2} - \bar{x}^2
 \end{aligned}$$

Exercice Modifier l'algorithme du calcul de la moyenne afin qu'il calcule et affiche la moyenne, la variance et l'écart type d'une série de valeur.

Exercice 1 Le tableau suivant donne les tailles de 34 élèves d'une classe.

taille(cm)	151	152	155	160	165	170	172	176	180	186	188
effectif	1	2	3	4	5	4	6	3	3	2	1

Calculer la moyenne et l'écart type de cette série.

Propriété: Propriétés de la moyenne.

— Si \bar{x}_1 et \bar{x}_2 sont les moyennes des deux sous groupes d'effectifs respectifs n_1 et n_2 , alors la moyenne de l'ensemble est

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

— La moyenne de la série des valeurs $ax_i + b$ est $a\bar{x} + b$.

Exercice 2 Dans une classe de 35 élèves, la moyenne des 20 filles de la classe est de 13. La moyenne des garçons est de 11. Quelle est la moyenne de la classe ?

Exercice 3 Un relevé de température (très précis) a donné les valeurs, en degré Celsius :

$$37,2408 - 37,2407 - 37,2410 - 37,2414 - 37,2412 - 37,2409.$$

1. Calculer (sans calculatrice!) la moyenne de ces valeurs.
2. On convertit les degrés Celsius en degrés Kelvin en ajoutant 273,15 à la température en degrés Celsius. Calculer la moyenne, en degré Kelvin, des températures précédentes.

II Description par la médiane et les quantiles

Définition: Caractéristiques de position, ou de tendance centrale

On considère une série statistique d'effectif total N .

- La **médiane** M_e est une valeur qui partage la série ordonnée en deux séries de même effectif.
Si N est impair, $N = 2p + 1$, alors la médiane est la $(p + 1)^{\text{ème}}$ valeur de la série ordonnée.
Si N est pair, $N = 2p$, alors la médiane est la moyenne de la $p^{\text{ème}}$ et de la $(p + 1)^{\text{ème}}$ valeur.
- Le **premier quartile** Q_1 est la plus petite valeur de la série telle que 25 % des valeurs (un quart) de la série lui soient inférieures.
Le **troisième quartile** Q_3 est la plus petite valeur de la série telle que 75 % des valeurs (les trois quarts) de la série lui soient inférieures.

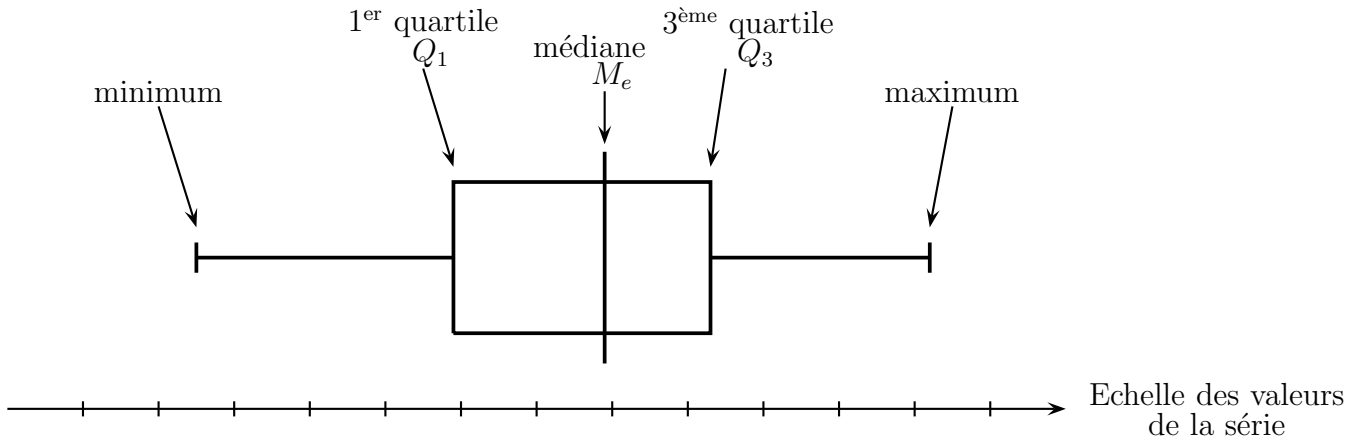
- De même avec les déciles (D_1 avec 10 %, D_9 avec 90 %) et les centiles (C_1 avec 1 % ; ...).

Définition: Caractéristiques de dispersion

- **L'étendue** d'une série est la différence entre sa plus grande et sa plus petite valeur.
- **L'écart inter-quartile** est la différence entre le troisième et le premier quartile : $Q_3 - Q_1$.

Exercice Donner la médiane, l'étendue, les quartiles et l'écart inter-quartile des notes de l'élève G.

Diagrammes en boîte (boîtes à moustaches) On peut alors représenter les données de la série statistique par un diagramme en boîte, aussi connu sous le nom de "boîte à moustaches" :



Exercice 4 Le tableau suivant donne les notes des élèves d'une classe.

Elèves	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Notes	15	10	12	8	10	18	12	8	8	15	10	8	6	18	12	8	12

On ordonne la série :

Notes x_i																	
Effectifs n_i																	
Effectifs cumulés croissants																	

L'effectif total de la série : $N = \dots$

La médiane de la série : $M_e = \dots$

Les 1^{er} et 3^{ème} quartiles sont : $Q_1 = \dots$, $Q_3 = \dots$

L'étendue de la série est : \dots

L'écart inter-quartile est : \dots

Tracer le diagramme en boîte de cette série.

Exercice 5 On compare les températures moyennes (en ° C) de chaque mois de l'année pour deux communes de Haute-Savoie situées à 1000 m d'altitude : Chamonix et La Clusaz.

Mois	1	2	3	4	5	6	7	8	9	10	11	12
Chamonix	1,5	4	7,5	12	15,5	20	23	22	19	14	6,5	2
La Clusaz	2,5	3,5	6	9,5	14	17	20,5	20,0	17	13	7	3,5

Déterminer pour ces deux communes la médiane et les quartiles des températures.

Tracer ensuite les diagrammes en boîte de ces deux séries en utilisant la même échelle, de manière à pouvoir les comparer.

III Fréquence des valeurs d'une série statistique

Définition: La fréquence, ou pourcentage, f_i de la valeur x_i du caractère est égale au quotient de l'effectif n_i par l'effectif total : $f_i = \frac{n_i}{N}$

Exemple :

Etude du nombre de garçons et de filles dans une classe de 34 élèves :

	Garçons	Filles
Effectifs (n_i)	14	20
Fréquences (f_i)	$\frac{14}{34} = \frac{7}{17} \sim 0,41$	$\frac{20}{34} = \frac{10}{17} \sim 0,59$
Pourcentages	$\sim 41\%$	$\sim 59\%$

Propriété: La somme des fréquences d'une série statistique est égale à 1.

Démonstration : $f_1 + f_2 + \dots + f_p = \frac{n_1}{N} + \frac{n_2}{N} + \dots + \frac{n_p}{N} = \frac{n_1 + n_2 + \dots + n_p}{N} = \frac{N}{N} = 1$

Propriété: La moyenne d'une série statistique dont le caractère prend les valeurs x_1, x_2, \dots, x_p avec les fréquences f_1, f_2, \dots, f_p est $\bar{x} = f_1x_1 + f_2x_2 + \dots + f_px_p$

Démonstration : $\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N} = \frac{n_1}{N}x_1 + \frac{n_2}{N}x_2 + \dots + \frac{n_p}{N}x_p = f_1x_1 + f_2x_2 + \dots + f_px_p$

Exercice 6 Dans un établissement, 350 élèves ont participé à une épreuve. Cette épreuve était notée de telle façon que seulement 4 notes étaient possibles : 0 ; 8 ; 14 ou 20. Les résultats ont été les suivants : 8 % des élèves ont eu 0 ; 28 % ont eu 8 ; 48 % ont eu 14 ; et 16 % ont eu 20

Quelle est la moyenne à cette épreuve dans l'établissement ?

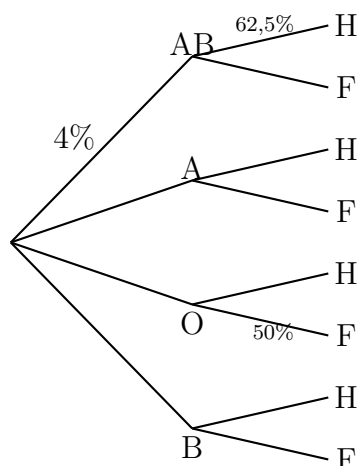
Exercice 7 Calculer la moyenne, la variance et l'écart type de la série :

Valeur du caractère	-5	3	8
Fréquence	0,2	0,3	0,5

Exercice 8 On donne la répartition des individus constituant un échantillon d'une population suivant deux critères qualitatifs : le sexe et le groupe sanguin.

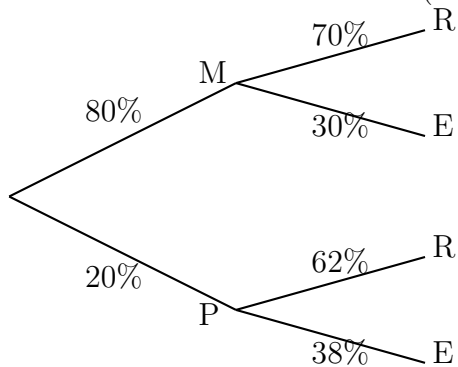
groupe \ sexe	sexe		total
	masculin	féminin	
AB	25	15	
A	250	200	
O	200	200	
B	60	50	
total			

1. Quel est le pourcentage d'hommes du groupe O dans l'échantillon ?
2. Quel est le pourcentage de femmes du groupe AB dans l'échantillon ?



3. Compléter l'arbre ci-contre.
4. Quel est le pourcentage d'hommes du groupe AB ?
Le pourcentage de femmes du groupe A ?
5. Parmi les personnes du groupe B, quel est le pourcentage d'hommes ?
de femmes ?

Exercice 9 L'arbre ci-dessous donne la répartition des réussites (R) et des échecs (E) au permis de conduire des moins de 25 ans (M) et plus de 25 ans (P) dans une auto-école.



Compléter le tableau à double entrée suivant :

	M	P	total
R			
E			
total			500

Pourcentages d'évolutions :

Exercice 10

- Un article coûte 24€. On lui applique une augmentation de 5%. Calculer son nouveau prix.
- Un article coûte initialement 112 €. Calculer son nouveau prix, après une réduction de 20%.

Propriété: Augmenter une quantité de $t\%$ revient à multiplier cette quantité par $(1 + t\%)$.
 Diminuer une quantité de $t\%$ revient à multiplier cette quantité par $(1 - t\%)$.

Exercice 11

- Un article coûte 25 €. Le vendeur décide de l'augmenter successivement de 10%, puis de 15%, et finalement de le diminuer de 25%. Quel est le prix final de cet article ?
- Après une réduction de 15%, un article coûte 32€. Quel était son prix initial ?

Exercice 12

- J'ai acheté une voiture 12 000 euros. La première année, le prix d'une voiture perd 30% de sa valeur.
Combien pourrais-je espérer revendre mon véhicule au bout d'un an ?
- La deuxième année, ainsi que les suivantes, le prix d'une voiture perd 20% par rapport à l'année précédente.
Que vaudra alors ma voiture dans 4 ans ?
A partir de combien d'années, ma voiture vaudra moins de 2 000 euros ?

Exercice 13 Dans une entreprise, le salaire moyen est de 1300 euros.

La direction décide d'accorder à tous les employés une augmentation de 2% et une prime de 10 euros.

Tous les salaires sont donc multipliés par 1,02 et augmentés de 10 euros.
Calculer le nouveau salaire moyen.

Exercice 14 Commenter cette annonce d'un journaliste : "Une nouvelle hausse de 15% du prix du tabac sera appliquée dès le 1er janvier qui, ajoutée à la hausse de 10% survenue 1er Septembre précédent, aura augmenté d'un quart le prix du tabac en quatre mois".

Exercice 15 On dispose de 5 000 euros d'économies. Quel est le placement le plus avantageux :

- 5000 euros à 10%
- 2000 euros à 9% et 3000 euros à 11%
- 2000 euros à 5% et 3000 euros à 15%

Exercice 16 Un article à 825 euros augmente de 25% puis baisse de 20%.

- Quel est son prix final ?

2. Quel est le pourcentage d'évolution global, c'est-à-dire entre le prix initial et le prix final ?

Exercice 17 Le bénéfice d'une entreprise était, il y deux ans de 125 000 euros. Ce bénéfice a baissé l'année dernière de 10 % avant de réaugmenter de 10 % cette année.

1. Calculer le pourcentage global d'évolution du bénéfice sur ces deux dernières années.
2. On appelle pourcentage moyen d'évolution, le pourcentage d'évolution qui serait constant, identique à chaque évolution successive, et qui permettrait d'aboutir à la même valeur finale. Calculer le pourcentage moyen annuel d'évolution du bénéfice.

Exercice 18 On augmente la longueur d'un rectangle de 20 % et on diminue sa largeur de 20 %. Son aire a-t-elle varié ? Si oui, préciser cette variation en pourcentage.

Exercice 19 Effet de structure

Lors d'un discours au cours duquel il a donné les résultats des examens de fin d'études des deux universités du pays, le ministre a déclaré :

“Dans l'université du Nord, 82 % des garçons et 80 % des filles ont réussi.

Dans l'université du Sud, 56 % des garçons et 52 % des filles ont réussi.

Je ne suis pas sexiste, mais il faut bien reconnaître que dans notre pays, les garçons réussissent mieux que les filles.”

Compléter le tableau :

	Garçons		Filles	
	Total	Admis	Total	Admises
Université du Nord	500	410	500	400
Université du Sud	800	448	200	104
Total	1300	858, soit 66 % des garçons	700	504, soit 72 % des filles

Calculer les proportions de filles et de garçons qui ont réussi dans le pays. La conclusion du ministre est-elle exacte ?

Exercice 20 Répartition des salaires dans une entreprise

Le tableau ci-dessous donne la répartition en 2005 et 2010 des ouvriers et cadres dans une entreprise, ainsi que le salaire de chacun.

	2005		2010	
	Effectifs	Salaires	Effectifs	Salaires
Cadres	30	2500	20	2600
Ouvriers	40	1200	50	1300
Total	70	123 000	70	117 000

Le directeur affirme : “dans mon entreprise, en 5 ans, tous les salaires ont augmenté, ceux des cadres et ceux des ouvriers”.

Le responsable syndical affirme : “dans l'entreprise, en 5 ans, le salaire moyen a diminué.”

Qui a raison ?

Calculer le pourcentage d'évolution (augmentation ou diminution) du salaire des cadres, du salaire des ouvriers, et du salaire moyen entre 2005 et 2010.

Définition: Effet de structure (D'après l'INSEE, Institut National de la Statistique et des Etudes Economiques)

Lorsqu'une population est répartie en sous-populations, il peut arriver qu'une grandeur évolue dans un sens pour chaque sous-population et dans le sens contraire sur l'ensemble de la population.

Ce paradoxe s'explique parce que les effectifs de certaines sous-populations augmentent alors que d'autres régressent : c'est l'effet de structure.