



Chapitre VI : statistique et échantillonnage

I Moyenne et écart-type

Définition I.1

On considère une série statistique de p valeurs différentes : x_1, x_2, \dots, x_p . On note n_1, n_2, \dots, n_p les effectifs associés :

Valeur	x_1	x_2	\dots	x_p
Effectif	n_1	n_2	\dots	n_p

La **moyenne** (pondérée) de cette série, notée \bar{x} est alors donnée par

Remarques :

- Chaque n_i est le « poids » de la valeur x_i .
- $n_1 + n_2 + \dots + n_p$ représente

Exemple 1. On considère les résultats de deux équipes de basket durant une saison. On a regroupé leurs matchs dans des plages/des classes de points.

Nombre de points	[40; 50[[50; 60[[60; 70[[70; 80[[80; 90[[90; 100[
Nombre de matchs de l'équipe A	2	5	15	9	7	2
Nombre de matchs de l'équipe B	3	8	10	8	7	4

1. Combien de matchs l'équipe A a-t-elle joué ?

.....

2. Même question pour l'équipe B.

.....



3. Combien de fois l'équipe *A* a-t-elle gagné entre 50 et 60 points ?

.....

4. Pour l'équipe *B*, quelle classe de points a l'effectif le plus grand ?

.....

5. Donner pour chaque classe un représentant :

Nombre de points	[40; 50[[50; 60[[60; 70[[70; 80[[80; 90[[90; 100[
Représentant						

6. Calculer le score moyen de chaque équipe.

.....

.....

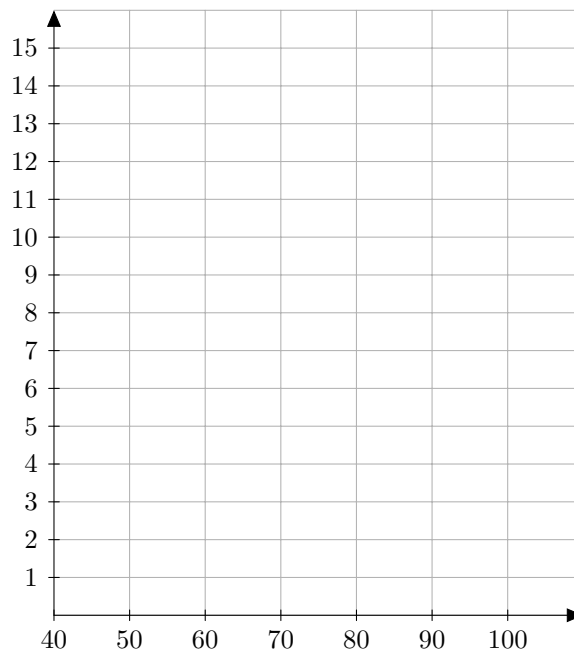
.....

.....

.....

.....

7. Représenter ci-dessous, en vert pour l'équipe *A* et en rouge pour l'équipe *B*, l'histogramme associé à ces résultats.



8. Comparer l'allure de ces deux séries statistiques.

.....

.....


.....



Définition I.2

On note σ l'**écart-type** d'une série statistique. Il se calcule à l'aide de la calculatrice et est une caractéristique de de la série.

Remarque : la moyenne est une caractéristique de position.

Exemple 2.  On reprend l'exemple 1.

- Sur votre calculatrice, appuyer sur [stats] puis dans le menu *éditer* choisissez 1 :*Modifier*.
- Dans la liste *L1* mettre les représentants des classes.
- Dans la liste *L2* mettre les effectifs de l'équipe *A*.
- Dans la liste *L3* mettre les effectifs de l'équipe *B*.
- Appuyer à nouveau sur [stats] puis choisir cette fois-ci dans *CALC* la première option : 1 :*Stats 1 Var*.
- Dans *XListe* : mettre *L1* ([2nde] [stats] 1 :*L1*).
- Dans *ListeFréq* : mettre *L2*.
- Valider.

1. Que retourne la calculatrice sur la première ligne ?

.....

2. L'écart-type, noté σ_x se trouve sur cinquième ligne. Combien vaut-il pour l'équipe *A* ?

.....

3. Dans cet affichage de la calculatrice, où se trouve l'effectif total ?

.....

4. A l'aide de la calculatrice toujours, déterminer l'écart-type de l'équipe *B*.

.....

5. Comparer ce résultat à celui de l'équipe *A*. Pouvait-on s'y attendre ? Pourquoi ?

.....



II Le diagramme en boîte

Définition II.1

La **médiane**, notée Me est un nombre qui découpe une série statistique en deux groupes de même

- Si l'effectif total N est pair, on prend la moyenne entre la valeur $\frac{N}{2}$ et la valeur $\frac{N}{2} + 1$.
- Si l'effectif total N est impair, on prend la valeur $\frac{N+1}{2}$.

Exemple 3. On regarde le prix d'une baguette en centimes d'euros dans différentes boulangeries.

prix	70	75	80	85	90	95	100	105	110	115	120
Effectifs	1	1	2	4	7	3	3	4	2	2	1
E.C.											

1. Remplir la dernière ligne du tableau par les effectifs cumulés de la séries.
2. Quel est le nombre total de boulangeries interrogées ?

.....

3. Calculer la médiane de cette série.

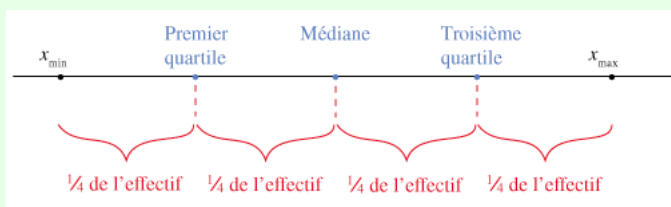
.....

.....

.....

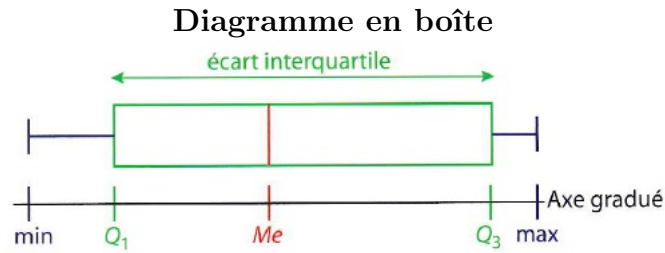
Définition II.2

Les **quartiles** Q_1 , $Q_2 = Me$ et Q_3 sont les trois nombres qui séparent la série statistique en quatre groupes de même effectif.



Définition II.3

L'**écart interquartile** est la différence et est une caractéristique de dispersion.



Exemple 4. On reprend l'exemple 3.

1. Calculer le premier et le troisième quartile de cette série.

.....

.....

.....

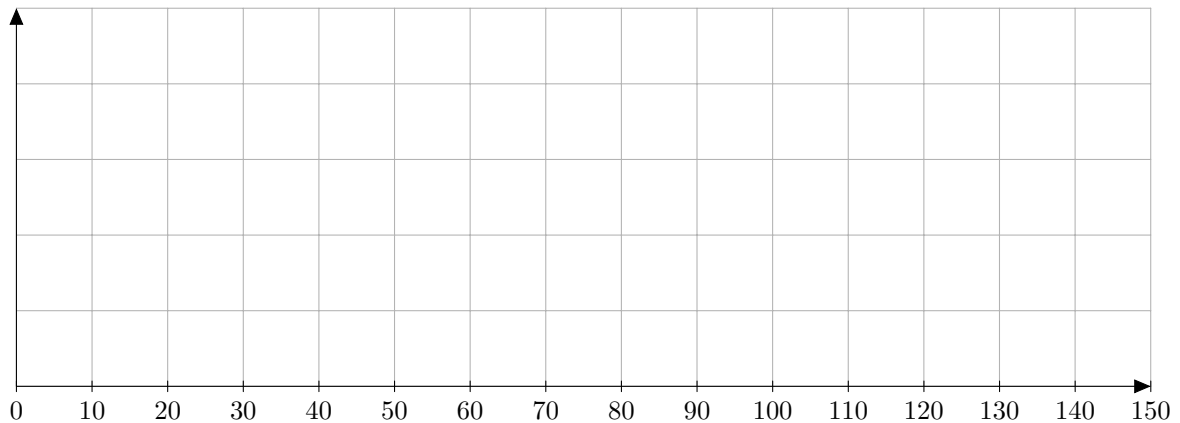
.....

2. En déduire l'écart interquartile.

.....

3. A l'aide de la calculatrice, contrôler vos résultats.

4. Tracer le diagramme en boîte associé.



III Echantillonnage

On considère une population d'individus présentant ou non un caractère particulier. On note p la proportion théorique des personnes qui présentent cette particularité.

Définition III.1

Un **échantillon** est un tirage au hasard, avec remise et de façon de n individus de cette population.

Proposition III.2

La loi X du nombre d'individus dans cet échantillon présentant le caractère est de paramètre n et p :

$$X \sim \dots\dots\dots$$

**Définition III.3**

Un intervalle de fluctuation à 95% est un intervalle auquel appartient X dans 95 réalisations sur 100.

Proposition III.4

Soit a le plus petit entier tel que $\mathbb{P}(X \leq a) \geq 0,025$ et b le plus grand entier tel que $\mathbb{P}(X \leq b) \geq 0,975$. Alors l'intervalle

est un intervalle de fluctuation à 95%.

Exemple 5. On publicité affirme que 80% des personnes utilisant leur produit ont leurs cheveux qui repoussent au cours du premier mois d'utilisation. On donne ci-dessous un extrait de la table de la fonction de répartition de la loi binomiale, calculée à l'aide de la commande LOI.BINOMIALE.N('valeur de a';50;0,8;VRAI) dans un tableur excel :

a	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
P(X < a)	0,0009	0,0025	0,0063	0,0144	0,0308	0,0607	0,1106	0,1861	0,2893	0,4164	0,5563	0,6927	0,8096	0,8966	0,952	0,9815	0,9943	0,9987	0,9998	1	1

1. A l'aide de ce tableau, déterminer un intervalle de fluctuation à 95%.

.....

.....

.....

.....

.....

Proposition III.5

On peut vérifier si la proportion p théorique est en accord avec une fréquence mesurée f en regardant si $f \in \left[\frac{a}{n}; \frac{b}{n} \right]$ ou non. Si c'est le cas, on l'hypothèse de la proportion p , sinon

Exemple 6. On reprend l'exemple 5. Sur les 50 personnes interrogées, 35 affirment avoir eu les cheveux qui ont repoussé. Accepte-t-on à 95% l'affirmation de la publicité ?

.....

.....

.....