I Définitions et vocabulaire des statistiques

La **population** est l'ensemble des individus sur lesquels portent l'étude statistique. Le **caractère** (ou **variable statistique**) d'une série statistique est une propriété étudiée sur chaque individu.

Lorsque le caractère ne prend que des **valeurs** (ou **modalités**) numériques, on dit qu'il est **quantitatif**. Sinon, on dit qu'il est **qualitatif**. Quand il est quantitatif, il est dit **discret** s'il ne peut prendre que des valeurs isolées (par exemple : série de notes lors d'un contrôle) ; il est dit **continu** dans le cas contraire, et dans ce cas on effectue souvent un **regroupement** des valeurs **par classes** (voir plus loin). Si le caractère n'est pas quantitatif, il est dit **qualitatif** : les modalités ne sont pas des nombres.

✥ A chaque valeur (ou classe) est associée un **effectif** n : c'est le nombre d'individus associés à cette valeur.
✥ De même à chaque valeur (ou classe) est associée une **fréquence** f : c'est la proportion d'individus associés à cette valeur. f est un nombre compris entre 0 et 1, que l'on peut écrire sous forme de pourcentage.

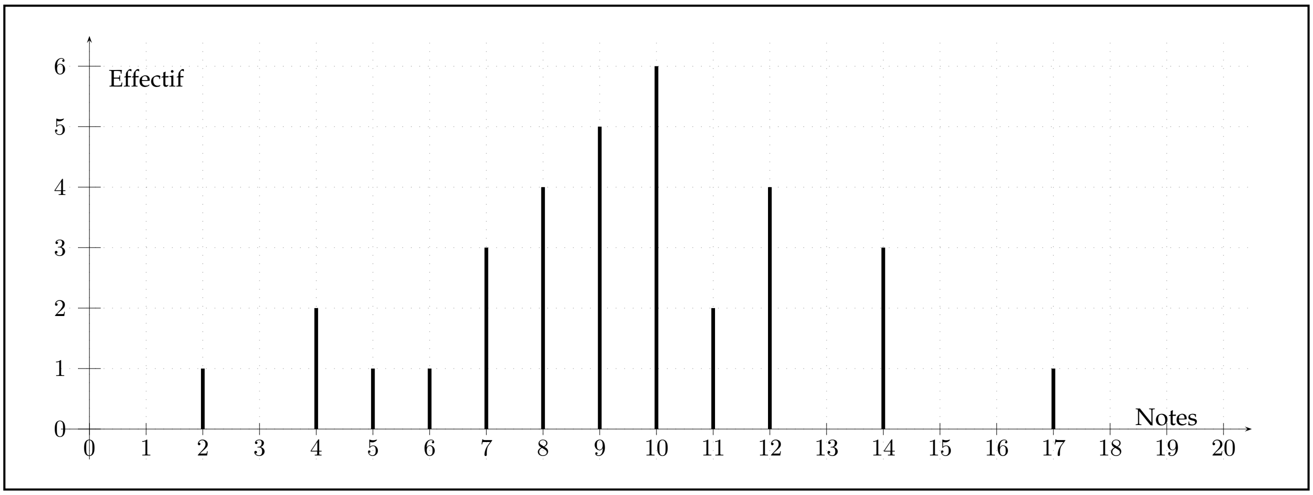
✥ Si N est l'effectif total (l'effectif de la population entière) alors on a $f = \dfrac{n}{N}$ (ou $f = \dfrac{n}{N} \times 100$ si on l'exprime sous forme de pourcentage).

Faire des **statistiques**, c'est recueillir, organiser, synthétiser, représenter et exploiter des données, numériques ou non, dans un but de comparaison, de prévision, de constat... Les plus gros "consommateurs" de statistiques sont les **assureurs** (risques d'accidents, de maladie des assurés), les **médecins** (épidémiologie), les **démographes** (qui étudient les populations et leur dynamique) et **sociologues** (qui étudient les phénomènes sociaux humains), les **économistes** (emploi, conjoncture économique), les **météorologues** . . .

II Représentation graphique d'une série statistique

Lorsque le caractère étudié est **quantitatif et discret**, on peut représenter la série statistique étudiée par un **diagramme en bâtons** : la hauteur de chaque bâton est alors proportionnelle à l'effectif (ou à la fréquence) associé à chaque valeur. Par exemple, voici le diagramme en bâtons représentant la série des notes obtenues par une classe à un contrôle :

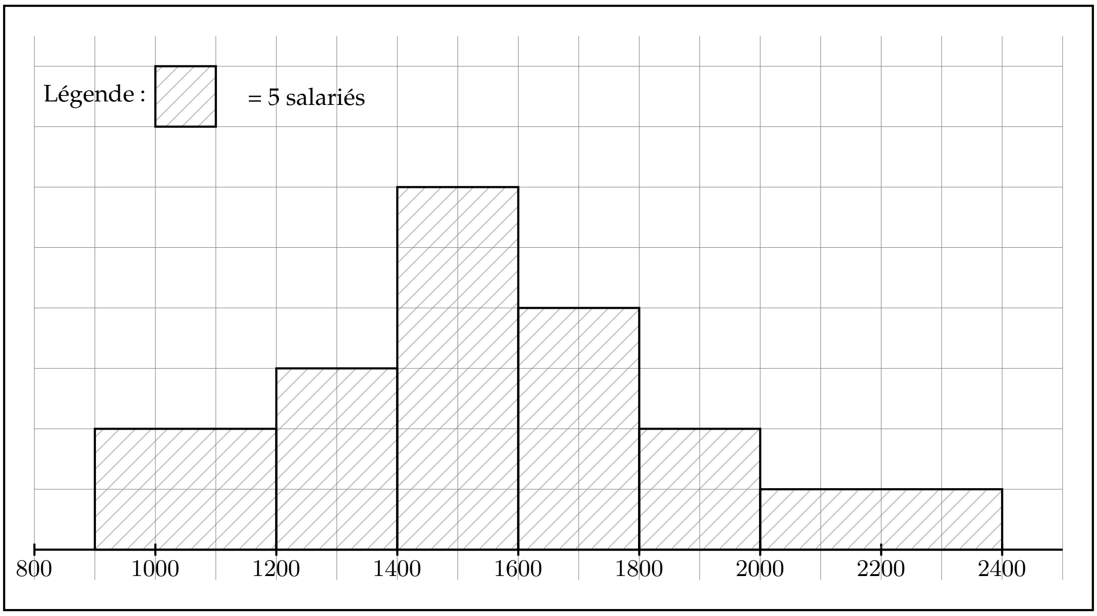
Série n°1 :



Notes	**2**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**14**	**17**	**Total**
Effectif													
Fréquence													

Lorsque le caractère étudié est **quantitatif et continu**, et lorsque les modalités sont regroupées en classes, on peut représenter la série par un **histogramme** : l'aire de chaque rectangle est alors proportionnelle à l'effectif (ou à la fréquence) associée à chaque classe. Lorsque les classes ont la même **amplitude**, c'est la hauteur de chaque rectangle qui est proportionnelle à l'effectif. Par exemple, voici un histogramme représentant la répartition des salaires dans une entreprise :

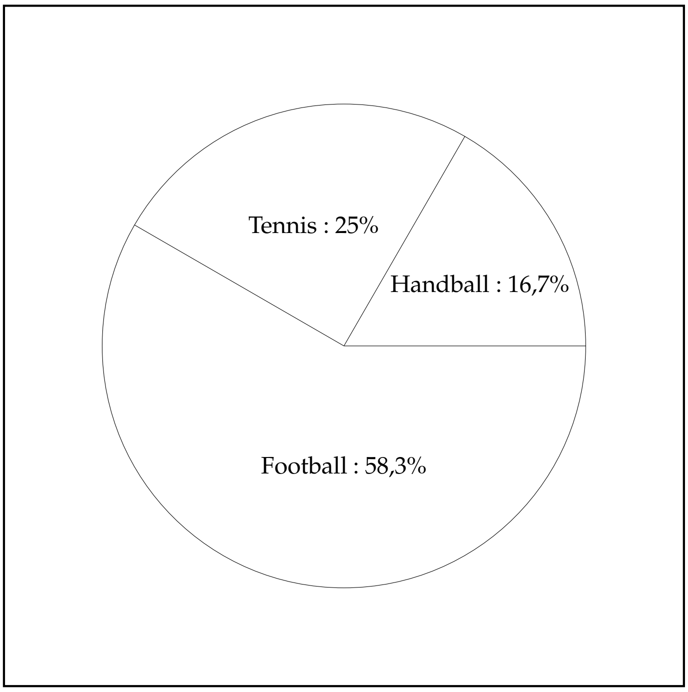
Série n°2 :



Salaires	[900 ; 1200]	[1200 ; 1400]	[1400 ; 1600]	[1600 ; 1800]	[1800 ; 2000]	[2000 ; 2400]	**Total**
Effectif							
Fréquence							

Enfin, lorsque le caractère est **qualitatif**, on représente la série par un **diagramme circulaire ou semi-circulaire** ("camemberts") : la mesure de chaque secteur angulaire est proportionnel à l'effectif (ou à la fréquence) associé. Par exemple, voici un diagramme circulaire représentant la répartition des adhérents à un club sportif :

Série n°3 :



Question : Sachant que ce club compte 240 adhérents, combien d'adhérents jouent :
- au tennis ? . . . .
- au football ? . . . .
- au handball ? . . . .

III Mode, classe modale

Définition :
❖ Un **mode** d'une série statistique est une valeur de la série pour laquelle l'effectif associé est le plus grand.
❖ Dans le cas d'un regroupement en classes, une **classe modale** est une classe pour laquelle l'effectif associé est le plus grand.

Par exemple :
- Dans la série n°1 le mode est la valeur . . . .
- Dans la série n°2 la classe modale est . . . .
- Dans la série n°3 le mode est la modalité . . . .

IV Moyenne d'une série statistique

Définition :
On considère une série statistique à caractère quantitatif, dont les N valeurs sont données par x_1, x_2, ..., x_p (d'effectifs associés n_1, n_2, ..., n_p avec $n_1 + n_2 + ... + n_p = N$).
La **moyenne pondérée** de cette série est le nombre noté $\overline{x}$ et qui vaut

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + ... + n_px_p}{n_1 + n_2 + ... + n_p} = \frac{1}{N}\sum n_ix_i$$

Remarque : le symbole $\sum$ signifie "somme sur tous les...",

Par exemple :
- Dans la série n°1 la moyenne du contrôle est égale à

  . . . .

  Important :
  Lorsque la série est regroupée en classes, on calcule la moyenne en prenant pour valeurs x_i le **centre de chaque classe** ; ce centre est obtenu en faisant la moyenne des deux extrémités de la classe.

– Dans la série n°2 une estimation du salaire moyen est donné par

. . . .

Remarque :
On peut également calculer une moyenne à partir de la distribution de fréquences :

$$\overline{x} = f_1x_1 + f_2x_2 + \cdots + f_px_p = \sum f_ix_i$$

Vous pouvez vérifier avec la série n°1, et faire le calcul de moyenne à partir des fréquences :

. . . .

Linéarité de la moyenne :
✥Si on ajoute (ou soustrait) un même nombre k à toutes les valeurs d'une série, alors la moyenne de cette série se trouve augmentée (resp. diminuée) de k.
✥ Si on multiplie (ou divise) par un même nombre non nul k toutes les valeurs d'une série, alors la moyenne de cette série se trouve multipliée (resp. divisée) par k.

Par exemple,
– Si on ajoute $1,5$ points à chaque note du contrôle de la série n°1, alors la moyenne de classe devient

. . . .

– Si on augmente chaque salarié de 5% dans la série n°2, cela revient à multiplier chaque salaire par $1,05$ (voir par ailleurs) ; le salaire moyen passe alors à

. . . .

Moyenne par sous-groupes
Soit une série statistique, d'effectif total N, de moyenne $\overline{x}$.
On divise cette série en deux sous-groupes **disjoints** (sans individus communs) d'effectifs respectifs p et q (avec $p + q = N$).
Si $\overline{x_1}$ est la moyenne du premier sous-groupe, et $\overline{x_2}$ est celle du second sous-groupe, alors on a

$$\overline{x} = \frac{p}{N}\overline{x_1} + \frac{q}{N}\overline{x_2}$$

Par exemple, supposons que les 15 garçons de la classe (dans la série n°1) aient obtenu une moyenne globale de $9,40/20$, quelle est la moyenne du groupe formé par les filles de la classe ?

. . . .

. . . .

V Médiane et quartiles d'une série statistique

V.1 médiane

Définition :
Soit une série statistique **ordonnée** dont les N valeurs sont $x_1 \leqslant x_2 \leqslant x_3 \leqslant \cdots \leqslant x_N$ (où les x_i ne sont pas forcément distincts les uns des autres. . .).
Grossièrement, la médiane est un nombre M qui permet de diviser cette série en deux sous-groupes de même effectif ; plus précisément, cela dépend de la parité de N :

✥ **Si N est impair**
alors M est la valeur de cette série qui est située "au milieu", à savoir la valeur dont le rang est $\dfrac{N+1}{2}$, notée $x_{\frac{N+1}{2}}$.

✥ **Si N est pair**
alors M est le centre de ce que l'on appelle l'**intervalle médian**, qui est l'intervalle formé par les deux nombres situés "au milieu" de la série, à savoir $\left[x_{\frac{N}{2}}\ ;\ x_{\frac{N}{2}+1}\right]$.

Clarifions cela grâce à des exemples simples :

➥ Si la série ordonnée est 2 – 5 – 6 – 8 – 9 – 9 – 10
alors $N = 7$ est **impair**, et donc la médiane M est la "valeur du milieu", à savoir $M = 8$.

➥ Si la série ordonnée est 2 – 5 – 6 – 8 – 9 – 9
alors $N = 6$ est **pair**, l'intervalle médian est $[6\ ;\ 8]$, et donc $M = \dfrac{6+8}{2} = 7$.

➥ Si la série ordonnée est 2 – 5 – 6 – 6 – 9 – 10
alors $N = 6$ est **pair**, l'"intervalle médian" (ce n'est pas vraiment un intervalle) est $[6\ ;\ 6]$, et donc $M = 6$.

Questions :

– Quelle est la médiane de la série n°1 ? ....

– On ne peut pas calculer précisément la médiane de la série n°2, mais dans quelle classe se situe-t-elle ? .

V.2 Quartiles

Le **premier quartile** Q_1 est le plus petit nombre de la série tel qu'au moins 25% des données sont inférieures ou égales à ce nombre.
Le **troisième quartile** Q_3 est le plus petit nombre de la série tel qu'au moins 75% des données sont inférieures ou égales à ce nombre.

Méthode pour la détermination des quartiles

Dans le cas d'une série statistique discrète $(x_1, x_2, \cdots, x_n)$ rangée dans l'ordre croissant.

1. Le premier Quartile Q_1 est la valeur x_i dont l'indice i est le plus petit entier supérieur ou égal à $\dfrac{n}{4}$
2. Le troisième Quartile Q_3 est la valeur x_j dont l'indice j est le plus petit entier supérieur ou égal à $\dfrac{3n}{4}$

Exemples

Les notes a un devoir ont été :
7, 8, 9, 10, 10, 11, 12, 13, 14, 15, 16 et 18.
Déterminer Q_1 et Q_3

Les notes a un devoir ont été :
5, 7, 8, 9, 10, 10, 11, 12, 13, 14, 15, 16 et 18.
Déterminer Q_1 et Q_3

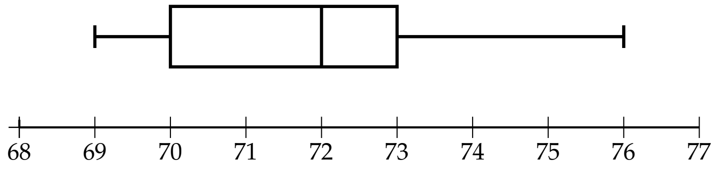
✂ ....

V.3 Diagramme en boite

Un diagramme en boîte est un diagramme qui fait apparaître la médiane, les quartiles et les valeurs extrêmes du caractère.
La **boîte** est délimitée par les premier et troisième quartiles et partagée par la médiane, sa "longueur" est donc égale à l'intervalle interquartile. Les moustaches sont les segments reliant les quartiles aux valeurs extrêmes de lma série de la série. Souvent lorsqu'elles sont connues, on indique les valeurs extrêmes de la série.
Les boîtes à moustache ont été crées en 1977 par J.W. Turkey, elles constituent un résumé à la fois visuel et numérique d'une série statistique.

Exemple 1

On a mesuré l'envergure de neuf papillons d'une espèce européenne rare et on a obtenu les données suivantes exprimées en millimètres : 69, 70, 70, 72, 72, 72, 73, 75 et 76.
En déduire Q_1, M_e et Q_3, les retrouver sur le diagramme en boite donnée ci-dessous.



✂ ....

Exemple 2
Les répartitions suivantes donnent les précipitations moyennes mensuelles en millimètres à Nice et à Paris.

	J	F	M	A	M	J	J	A	S	O	N	D
Nice	67	83	71	70	39	37	21	38	83	109	158	92
Paris	53	48	40	45	53	57	54	61	54	50	58	51

1. Construire le diagramme en boîte de chaque série avec une même graduation
2. Comparer les deux séries à l'aide de leurs diagrammes en boîte.

✂ ....

VI Etendue d'une série statistique

Définition :
L'**étendue** d'une série statistique (à caractère quantitatif) est la différence entre la plus grande valeur de la série et la plus petite.

C'est un indicateur - sommaire - de la **dispersion** des valeurs de la série.
Par exemple, l'étendue de la série n°1 est ....