Chapitre 8 : Statistiques

I Indicateurs d'une série statistique

I.1 Moyenne

Définition (moyenne)
Soit une série quantitative discrète.
Notons x_1, x_2, …, x_p les valeurs et n_1, n_2, …, n_p les effectifs respectifs.
L'effectif total est alors $N = n_1 + n_2 + \dots n_p$.
La moyenne (pondérée) est le nombre noté $\overline{x}$ (prononcer "x barre") défini par :

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + \cdots + n_px_p}{N}$$

Remarque (moyenne à partir des fréquences)
Si l'on note f_1, f_2, …, f_p les fréquences associées respectivement aux valeurs x_1, x_2, …, x_p, alors

$$\overline{x} = f_1x_1 + f_2x_2 + \dots f_px_p$$

Exercice 1
Déterminer la moyenne de la série dans chacun des cas suivants.

1. Les valeurs sont données par le tableau suivant :

Valeurs x_i	7	11	12	17
Effectifs n_i	1	4	3	2

. . . .
. . . .

2. Les valeurs de la série sont : 3 ;1 ;-4 ; 0 ;1 ;2 ;1 ;0.

. . . .
. . . .

Remarque
Dans le cas des séries à caractère continu (regroupée en classes), on prend les centres des classes pour jouer le rôle des valeurs. On obtient alors une estimation de la moyenne.

Propriété (linéarité de la moyenne)
Si l'on ajoute un même nombre a à chacune des valeurs de la série, alors la moyenne augmente de a.
Si l'on multiplie chaque valeur de la série par un même nombre b, alors la moyenne est multipliée par b.

Exercice 2
En comparant cette série à celle de l'exercice 1 question 1., déterminer directement la moyenne $\bar{y}$.

Valeurs y_i	9	13	14	19
Effectifs n_i	1	4	3	2

. . . .
. . . .

I.2 Médiane

Définition (Médiane d'une série discrète)
On considère une série statistique discrète x_1, x_2, ..., x_N de N valeurs .
On appelle médiane de la série et on note Me tout nombre tel que :
au moins 50% des valeurs de la série sont inférieures ou égales à Me,
et au moins 50% des valeurs de la série sont supérieures ou égales à Me.

Méthode de détermination
On range les N valeurs x_1, x_2, ..., x_N dans l'ordre croissant.
Si l'effectif total N est pair, on choisit pour médiane la demi-somme des deux valeurs centrales.
Si l'effectif total N est impair, la médiane est la valeur centrale de la série.

Remarque
Si N est pair et que les deux valeurs centrales sont distinctes, alors tout nombre compris strictement entre ces deux valeurs centrales peut être appelé médiane (il n'y a pas unicité). Dans les autres cas, il n'y a qu'une médiane.

Exercice 3
Dans chaque cas, compléter les effectifs cumulés croissants puis déterminer la médiane.

1.

Valeurs x_i	35	36	37	38	39	40
Effectifs n_i	3	4	2	1	1	2
ECC						

....
....
....
....

2.

Valeurs x_i	5	11	27	33	42
Effectifs n_i	4	3	1	5	3
ECC					

....
....
....
....

I.3 Quartiles

Définition
On considère une série statistique à caractère quantitatif discret.
Le premier quartile Q_1 est la plus petite valeur de la série telle qu'au moins 25 % des valeurs soient inférieures ou égales à Q_1.
Le troisième quartile Q_3 est la plus petite valeur de la série telle qu'au moins 75 % des valeurs soient inférieures ou égales à Q_3.
L'intervalle interquartile est $[Q_1; Q_3]$. Il contient environ 50 % des valeurs.
L'écart interquartile est le nombre $Q_3 - Q_1$.

Méthode pour déterminer les quartiles :
Soit N l'effectif total de la série. On range les N valeurs x_1, x_2, ..., x_N dans l'ordre croissant.
Pour Q_1, on calcule $\dfrac{N}{4}$, ce qui donne, en arrondissant si besoin à l'entier supérieur, le rang de Q_1.
Pour Q_3 on procède de même en remplaçant $\dfrac{N}{4}$ par $\dfrac{3N}{4}$.

Remarque
Les quartiles sont toujours des valeurs de la série.
L'écart interquartile est un indicateur de dispersion. Plus l'écart interquartile est grand, plus les données sont dispersées.

Exercice 4
Compléter

Valeurs x_i	35	41	46	65	81
Effectifs n_i	5	3	2	3	2
Effectifs cumulés croissants					

L'effectif total est $N = 5 + 3 + 2 + 3 + 2 = 15$.
Déterminons le 1er quartile Q_1.
$\dfrac{N}{4} = \dfrac{15}{4} = 3.75$. Donc Q_1 est la ... valeur. $Q_1 = \ldots$.
Déterminons le 3e quartile Q_3.
$\dfrac{3N}{4} = \dfrac{3 \times 15}{4} = 11.25$. Donc Q_3 est la ... valeur. $Q_3 = \ldots$.
L'intervalle interquartile est donc ...
L'écart interquartile est ...

I.4 Écart-type

Définition
La variance de la série statistique est le nombre positif V défini par :

$$V = \frac{n_1(\bar{x} - x_1)^2 + n_2(\bar{x} - x_2)^2 + \cdots + n_p(\bar{x} - x_p)^2}{N} = \frac{1}{N}\sum_{i=1}^{p} n_i(\bar{x} - x_i)^2$$

Autrement dit, la variance est la moyenne des carrés des écarts la moyenne.

Définition
L'écart-type de la série est le nombre positif noté σ ("sigma") défini par $\sigma = \sqrt{V}$.

Exercice 5
On considère la série statistique suivante.

Valeurs y_i	7	11	13	15
Effectifs n_i	1	4	3	2

1. Calculer la moyenne $\bar{x}$ de cette série.
2. Déterminer la variance V de la série.
3. En déduire l'écart-type σ de la série.

Remarque
L'écart-type est un indicateur sur la dispersion des valeurs par rapport à la moyenne.
Plus précisément, plus l'écart-type est grand, plus les valeurs ont tendance à être éloignées de la moyenne.

Remarque
On résume souvent une série par un couple associant un indicateur de tendance centrale avec un indicateur de dispersion. On choisit en général :
— (moyenne $\bar{x}$; écart-type s) ou (moyenne ; étendue)
— (médiane Me ; écart interquartile $Q_3 - Q_1$).

I.5 Utilisation de la calcultatrice pour les statistiques

Texas

1. Entrer les données de la série :
   $\boxed{\text{Stats}}$, puis EDIT, (edite...).
   Entrer les valeurs dans la liste L_1, et les effectifs dans la liste L_2.
2. Obtenir les indicateurs de la série :
   $\boxed{\text{Stats}}$ CALC Stats 1-Var
   Xliste : L_1
   ListeFreq : L_2

Casio

1. Entrer les valeurs :
   Aller dans le menu $\boxed{\text{Statistiques}}$.
   Entrer les valeurs dans `List 1`, et les effectifs dans `List 2`.
2. Obtenir les indicateurs de la série :
   Aller dans le menu $\boxed{\text{CALC}}$.
   suivre $\boxed{\text{SET}}$ pour paramétrer les calculs de la façon suivante :
   — 1Var X list : `List 1`
   — 1Var Freq :`List 2`
   Alors on obtient les indicateurs de la série par : $\boxed{\text{CALC}}$ $\boxed{\text{1 VAR}}$

Remarque
L'étendue d'une série est la différence entre la plus grande valeur et la plus petite valeur.

Exercice 6
Voici un tableau présentant les salaires dans une entreprise.

Salaires mensuels (euros)	1200	1650	2100	2400	6500
Effectifs	5	8	4	3	1
Effectifs cumulés croissants					

1. Compléter les effectifs cumulés. Donner l'effectif total.
2. Déterminer, à l'aide de la calculatrice, la moyenne et l'écart-type, la médiane et les quartiles, l'étendue.
3. Que deviennent tous ces indicateurs si l'on remplace la plus grande valeur par 500 000 ? Que peut-on dire de la moyenne dans ce dernier cas ?

Remarque
La médiane et les quartiles ne sont pas influencés par les valeurs extrêmes de la série.
La moyenne, l'écart-type, et l'étendue sont sensibles aux valeurs extrêmes.

II Complément : diagramme en boîte

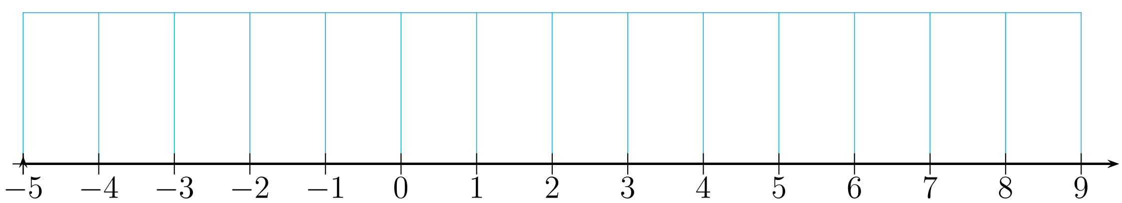
Ce diagramme permet de résumer une série en plaçant certains indicateurs sur un même axe. C'est notamment utile pour comparer deux séries statistiques.

Exercice 7
Voici un relevé des températures durant le mois de février :

Température observée	-4	-3	-2	0	2	3	4	5	6	7	8	9
Nombre de jours	1	2	1	1	1	1	2	3	4	5	4	3
Effectifs cumulés croissants												

1. Compléter les effectifs cumulés croissants.
2. L'effectif total est $N = \dots$
3. $\dfrac{N}{4} =\dots$, donc Q_1 est la …$^{\text{e}}$ valeur : $Q_1 =\dots$.
4. $\dfrac{3N}{4} =\dots$, donc Q_3 est la …$^{\text{e}}$ valeur : $Q_3 =\dots$.
5. Détermination de la médiane Me
   ……………………………………………………………………………
   ……………………………………………………………………………
   ……………………………………………………………………………

On résume la série par le diagramme en boîte (ou boîte à moustache) suivant :



Cette série a pour valeur minimale -4, pour maximum 9.
Le premier quartile est 3, la médiane est 6, et le troisième quartile est 7.