I MÉDIANE ET QUANTILES

1 LA MÉDIANE

La médiane d'une série statistique est une valeur telle qu'il y ait autant d'observations ayant une valeur supérieure à la médiane que d'observations ayant une valeur inférieure à la médiane.

La médiane d'une série statistique de N valeurs rangées par ordre croissant est le nombre M_e défini par :

— si l'effectif N est impair, la médiane M_e est la valeur centrale du caractère c'est à dire la valeur de rang $\frac{N+1}{2}$ de la série ordonnée.

— si l'effectif N est pair, la médiane M_e est la demi-somme des deux valeurs centrales du caractère c'est à dire la moyenne des valeurs de rangs $\frac{N}{2}$ et $\frac{N}{2}+1$ de la série ordonnée.

EXEMPLE

Dans un service de maintenance, on a répertorié le nombre d'interventions par jour sur un mois. On a obtenu la distribution suivante :

Nombre d'interventions x_i	3	5	6	7	8	9
Nombre de jours n_i	2	5	8	6	3	1

L'effectif total $N = 25$ donc la médiane est la valeur du caractère de rang 13 soit $Me = 6$.

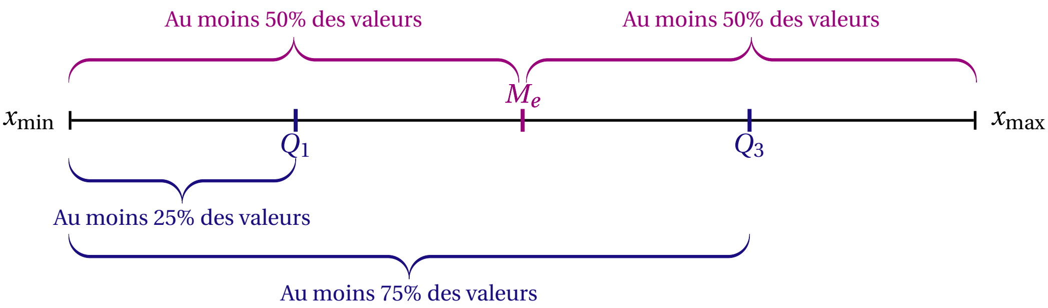
2 LES QUANTILES

LES QUARTILES

Les quartiles au nombre de trois Q_1, Q_2 et Q_3 partagent l'ensemble étudié de N éléments préalablement classés par valeurs croissantes, en quatre sous ensembles.

— Le premier quartile noté Q_1 est la plus petite valeur de la série statistique telle qu'au moins 25 % des valeurs de la série sont inférieures ou égales à Q_1.

— Le troisième quartile noté Q_3 est la plus petite valeur de la série statistique telle qu'au moins 75 % des valeurs de la série sont inférieures ou égales à Q_3.



REMARQUE

L'intervalle interquartile $[Q_1 ; Q_3]$ contient au moins 50% des valeurs de la série.

EXEMPLE

Dans la série précédente, l'effectif total $N = 25$.

— $25 \times \frac{1}{4} = 6{,}25$ donc le premier quartile est la valeur du caractère de rang 7 soit $Q_1 = 5$.

— $25 \times \frac{3}{4} = 18{,}75$ donc le troisième quartile est la valeur du caractère de rang 19 soit $Q_3 = 7$.

LES DÉCILES

Les déciles au nombre de neuf D_1, D_2, …, D_9 partagent l'ensemble étudié de N éléments préalablement classés par valeurs croissantes, en dix sous ensembles.

— Le premier décile noté D_1 est la plus petite valeur de la série statistique telle qu'au moins 10 % des valeurs de la série sont inférieures ou égales à D_1.

— Le neuvième décile noté D_9 est la plus petite valeur de la série statistique telle qu'au moins 90 % des valeurs de la série sont inférieures ou égales à D_9.

3 CARACTÉRISTIQUES DE DISPERSION

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

— L'écart interquartile est égal à la différence entre le troisième et le premier quartiles.

— L'écart interdécile est égal à la différence entre le neuvième et le premier déciles.

4 DIAGRAMME EN BOÎTE

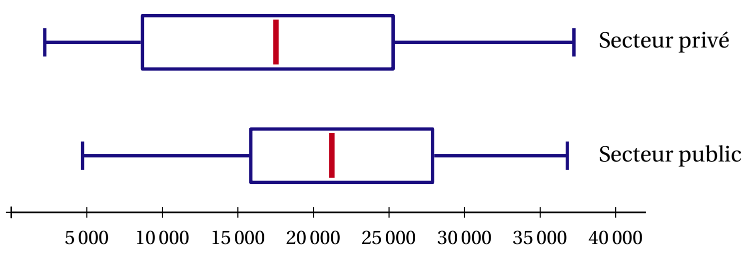
La représentation graphique de la dispersion d'une série statistique se fait à l'aide de diagramme en boîte appelés aussi « boîte à moustaches » ou « box-plot ».
Pour une catégorie donnée, on construit, en face d'un axe permettant de repérer les quantiles de la variable étudiée, un rectangle dont la longueur est égale à l'écart interquartile $Q_3 - Q_1$, la médiane est représentée par un trait. On ajoute alors des segments aux extrémités menant jusqu'aux valeurs extrêmes, ou jusqu'aux premier et neuvième déciles.

EXEMPLE

Le tableau suivant donne la distribution du revenu salarial par secteur d'activité en France en 2014.

	D1	Q1	Médiane	Q3	D9
Secteur privé	2 218	8 570	17 520	25 377	37 234
Secteur public	4 716	15 744	21 221	27 996	36 797

Source : INSEE



II MOYENNE, VARIANCE ET ÉCART-TYPE

1 LA MOYENNE

On considère la série statistique donnée par le tableau ci-contre.
On note $N = n_1 + n_2 + \cdots + n_p$ l'effectif total

Valeur x_i	x_1	x_2	$\cdots$	x_p
Effectif n_i	n_1	n_2	$\cdots$	n_p

La moyenne d'une série statistique est le quotient noté $\overline{x}$ de la somme de toutes les valeurs de cette série par l'effectif total.

$$\overline{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + \cdots + n_p \times x_p}{N} = \frac{1}{N}\sum_{i=1}^{p} n_i x_i$$

REMARQUE

Soit $f_i = \dfrac{n_i}{N}$ la fréquence de la valeur x_i alors, la moyenne $\overline{x} = f_1 \times x_1 + f_2 \times x_2 + \cdots + f_p \times x_p$.

EXEMPLE

Avec la série statistique précédente :

Nombre d'interventions x_i	3	5	6	7	8	9
Nombre de jours n_i	2	5	8	6	3	1
Fréquence f_i	0,08	0,2	0,32	0,24	0,12	0,04

Le nombre moyen d'interventions par jour est :

$$\overline{x} = \frac{2 \times 3 + 5 \times 5 + 8 \times 6 + 6 \times 7 + 3 \times 8 + 1 \times 9}{25} = 6{,}16$$

ou en utilisant les fréquences :

$$\overline{x} = 0{,}08 \times 3 + 0{,}2 \times 5 + 0{,}32 \times 6 + 0{,}24 \times 7 + 0{,}12 \times 8 + 0{,}04 \times 9 = 6{,}16$$

2 VARIANCE ET ÉCART-TYPE

Soit $(x_i ; n_i)$, $1 \leqslant i \leqslant p$, une série statistique de moyenne $\overline{x}$ et d'effectif total N.

— La variance de cette série est le nombre V défini par :

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

— L'écart-type, noté σ, de cette série est égal à la racine carrée de la variance :

$$\sigma = \sqrt{V}$$

REMARQUE

La variance est la moyenne des carrés des écarts à la moyenne $\overline{x}$. Les valeurs $\left(x_i - \overline{x}\right)$ sont les « écarts à la moyenne » ; les « carrés des écarts à la moyenne » sont donc $\left(x_i - \overline{x}\right)^2$.
En faisant la moyenne des carrés des écarts à la moyenne, on trouve la variance.

EXEMPLE

Dans la série précédente de moyenne $\overline{x} = 6{,}16$ la variance est :

$$V = \frac{2 \times (3-6{,}16)^2 + 5 \times (5-6{,}16)^2 + 8 \times (6-6{,}16)^2 + 6 \times (7-6{,}16)^2 + 3 \times (8-6{,}16)^2 + 1 \times (9-6{,}16)^2}{25} = 1{,}9744$$

L'écart-type de cette série est donc $\sigma = \sqrt{1{,}9744} \approx 1{,}4$

PROPRIÉTÉ

Soit $(x_i ; n_i)$, $1 \leqslant i \leqslant p$, une série statistique de moyenne $\overline{x}$ et d'effectif total N.
La variance de cette série est le nombre V défini par :

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

✻ DÉMONSTRATION

$$\begin{aligned}
V &= \frac{n_1 \times \left(x_1 - \overline{x}\right)^2 + n_2 \times \left(x_2 - \overline{x}\right)^2 + \cdots + n_p \times \left(x_p - \overline{x}\right)^2}{N} \\
&= \frac{n_1 \times \left({x_1}^2 - 2x_1\overline{x} + \overline{x}^2\right) + n_2 \times \left({x_2}^2 - 2x_2\overline{x} + \overline{x}^2\right) + \cdots + n_p \times \left({x_p}^2 - 2x_p\overline{x} + \overline{x}^2\right)}{N} \\
&= \frac{n_1 \times {x_1}^2 + n_2 \times {x_2}^2 + \cdots + n_p \times {x_p}^2}{N} - 2\frac{n_1 \times x_1 + n_2 \times x_2 + \cdots + n_p \times x_p}{N} \times \overline{x} + \frac{n_1 + n_2 + \cdots + n_p}{N} \times \overline{x}^2 \\
&= \frac{n_1 \times {x_1}^2 + n_2 \times {x_2}^2 + \cdots + n_p \times {x_p}^2}{N} - 2\overline{x} \times \overline{x} + \overline{x}^2 \\
&= \frac{n_1 \times {x_1}^2 + n_2 \times {x_2}^2 + \cdots + n_p \times {x_p}^2}{N} - \overline{x}^2
\end{aligned}$$

Ainsi, la variance d'une série statistique est égale à la différence entre « la moyenne des carrés » et « le carré de la moyenne ».

EXERCICE 1

Pour les trois séries statistiques ci dessous, la médiane est égale à 10. Compléter le tableau ci-dessous par des données pour chacune des séries sachant que :

— La moyenne de la série 1 est égale à 10.

— La moyenne de la série 2 est la plus petite possible.

— La moyenne de la série 3 est la plus grande possible.

indice i	1	2	3	4	5	6	7	8	9	10
Série 1	5									18
Série 2	5									18
Série 3	5									18

EXERCICE 2

Dans une entreprise, il y a 28 cadres et 92 ouvriers. Le salaire moyen des cadres est de 3 450 € et celui des ouvriers est de 1 320 €.

1. Calculer le salaire moyen de l'ensemble des salariés de cette entreprise.
2. a) Quel est le pourcentage d'augmentation du salaire moyen si on verse une prime de 35 € à chaque salarié ?

   b) On augmente le salaire de chaque cadre de 2 % et celui de chaque ouvrier de 4 %.
   Le salaire moyen dans l'entreprise a-t-il augmenté de 3% ?

EXERCICE 3

Un concours est organisé dans deux centres d'examens. Dans le premier centre, les garçons ont obtenu 13 de moyenne et les filles 12 de moyenne. Dans le second centre, les garçons ont obtenu 9 de moyenne et les filles 8 de moyenne.
Le président du jury en déduit que les garçons ont eu de meilleurs résultats que les filles Est-ce si sûr ?
Sachant qu'il y avait 58 garçons et 104 filles dans le premier centre, et 87 garçons et 32 filles dans le second centre, calculer la moyenne générale des garçons puis celle des filles. Conclure.

EXERCICE 4

PARTIE A

Le tableau suivant donne la distribution des salaires mensuels nets, en euros, en France en 2015.

Déciles	D1	D2	D3	D4	D5	D6	D7	D8	D9
Hommes	1 262	1 427	1 573	1 728	1 906	2 130	2 451	2 996	3 990
Femmes	1 171	1 288	1 396	1 512	1 650	1 830	2 073	2 432	3 149
Ensemble	1 213	1 357	1 490	1 630	1 797	2 004	2 286	2 752	3 646

Souce : Insee Première (*octobre 2017)*

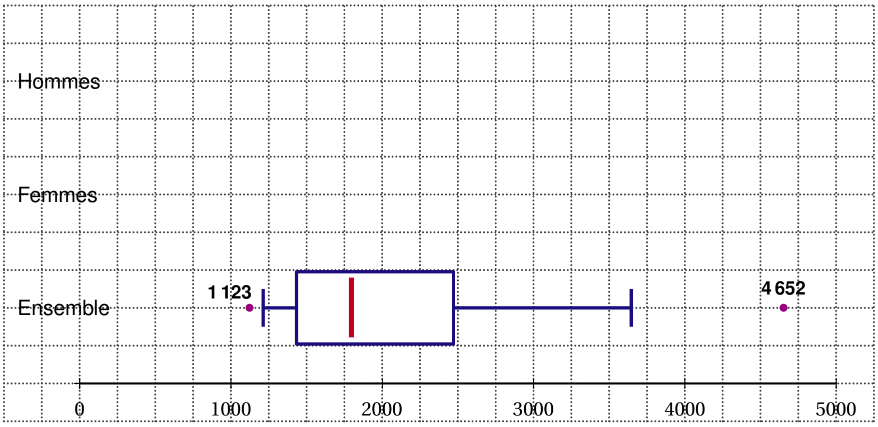
1. a) Donner le salaire net médian des salariés hommes et des salariées femmes.

   b) Calculer les variations en pourcentage des déciles des femmes par rapport à ceux des hommes

Déciles	D1	D2	D3	D4	D5	D6	D7	D8	D9
Hommes	1 262	1 427	1 573	1 728	1 906	2 130	2 451	2 996	3 990
Femmes	1 171	1 288	1 396	1 512	1 650	1 830	2 073	2 432	3 149
Écarts en %	−7,2								

   c) Recopier et compléter la phrase :
   « Les écarts de salaire entre femmes par rapport aux hommes ... le long de l'échelle salariale : de −7,2% pour le 1^{er} décile à ... pour le 9^{e} décile. »

2. Le montant en euros du premier quartile est de 1 488 euros pour les salariés hommes et de 1 332 euros pour les salariées femmes. Le troisième quartile est de 2 678 euros pour les hommes et de 2 208 euros pour les femmes.
   La distribution des salaires mensuels nets de l'ensemble des salariés est représentée ci-dessous (centiles C5 à C95).



   a) Donner une interprétation du nombre 4 652.
   b) Sur le même graphique, représenter la distribution des salaires nets des hommes et des femmes.

PARTIE B

Le tableau ci-dessous donne le montant en euros du salaire mensuel moyen net selon les catégories socioprofessionnelles en France en 2015.

	Hommes		**Femmes**	
	Salaires nets (en euros)	Effectifs (en %)	Salaires nets (en euros)	Effectifs (en %)
Cadres	4 451	20,6	3 561	15,6
Professions intermédiaires	2 420	18,9	2 081	20,8
Employés	1 739	16,2	1 591	50,8
Ouvriers	1 765	44,3	1 483	12,8

Souce : Insee Première (*octobre 2017)*

1. a) Calculer le salaire moyen des hommes et celui des femmes.
   b) Recopier et compléter la phrase :
   « En 2015, une salariée gagne, en moyenne en . . . % de moins qu'un salarié. »
2. 58,5% des salariés sont des hommes.
   De quel pourcentage, le salaire net médian de l'ensemble des salariés est-il inférieur au salaire net moyen ?

EXERCICE 5

1. Compléter le tableau ci-dessous qui donne la distribution des salaires mensuels bruts des 100 salariés d'une entreprise.

Salaires en euros	1500	1600	1900	2400	2700	3200	5000
Effectifs	30	25	15	12	8	6	4
Fréquences							

2. a) Donner le montant du salaire mensuel brut médian.
   b) Calculer le pourcentage de la masse salariale totale perçue par les 10% des salariés les mieux rémunérés.

3. a) Calculer le montant du salaire mensuel brut moyen.
   b) Calculer le pourcentage des salariés dont le salaire mensuel brut est compris dans l'intervalle $\left[\overline{x}-2\sigma;\overline{x}+2\sigma\right]$.

EXERCICE 6

Le tableau suivant donne le montant des salaires annuels exprimés en milliers d'euros d'une petite entreprise.

Salaires	16	18	20	25	30	40
Nombre de salariés	6	9	10	8	5	2

1. Déterminer la médiane, le premier et le troisième quartiles. Interpréter ces résultats et les traduire à l'aide d'un diagramme en boîte.
2. Calculer le montant en euros du salaire moyen annuel de cette entreprise.
3. A l'aide de la calculatrice, donner la valeur arrondie à l'euro près de l'écart-type s.
4. Soit S la fonction définie sur $\mathbb{R}$ par :

$$S(x)=6\times(16-x)^2+9\times(18-x)^2+10\times(20-x)^2+8\times(25-x)^2+5\times(30-x)^2+2\times(40-x)^2$$

   a) Vérifier que $S(x)=40x^2-1776x+21152$
   b) Déterminer le sens de variations de la fonction S et en déduire sa valeur minimale.
   c) Calculer la variance à partir de la somme S. En déduire la valeur exacte de l'écart-type σ.

EXERCICE 7

Une entreprise de produits alimentaires fabrique et distribue une marque de café dans des sachets de 250 grammes. On suppose que le poids du sachet vide est négligeable.
La machine utilisée pour remplir les sachets est contrôlée selon la procédure suivante.
À chaque heure, un échantillon aléatoire de 40 sachets est prélevé dans la production ; on mesure la masse de chaque sachet et on calcule la masse moyenne $\overline{x}$ de l'échantillon.
Un réglage de la machine est nécessaire si l'un des critères suivants n'est pas vérifié :

— les 40 sachets ont une masse supérieure ou égale à 245 grammes ;

— 50% au moins des sachets ont une masse en grammes comprise dans l'intervalle $[248;252]$;

— 95% au moins des sachets ont une masse en grammes comprise dans l'intervalle $\left[\overline{x}-2\sigma;\overline{x}+2\sigma\right]$.

Au cours de la production, l'échantillon suivant a été prélevé.

248	256	253	246	252	250	248	253	248	251
255	252	252	254	250	250	250	251	250	252
256	252	251	247	251	249	253	250	251	245
250	252	247	249	250	249	249	249	254	249

1. Recopier et compléter le tableau suivant :

Masse	245	246	247	248	249	250	251	252	253	254	255	256
Effectif												

2. Représenter la dispersion de cette série à l'aide d'un diagramme en boîte.
3. Faut-il effectuer un réglage de la machine ?