Module 2:
Mesures de tendance centrale et de dispersion

ÉLABORÉ PAR AFEF BEN ZINE EL ABIDINE

Principales mesures descriptives

Mesures de tendance centrale	Moyenne arithmétique Mode Médiane
Mesures de dispersion	Étendue Écart-type Coefficient de variation
Mesures de position (module 3)	Quantiles Cote z

Mesures de tendance centrale : moyenne, mode et médiane (manuel, p.39)

Deux cas peuvent se présenter:

- Données non groupées (série statistique)

Exemple : 25, 13, 15, 50, 23, ……………., 20.

- Données groupées (tableau de distribution)

Poids (kg)	Effectif
20≤X<25	12
25≤X<30	19
30≤X<35	27
35≤X<40	41
45≤X<50	11

Formule de la moyenne μ (données non groupées)

Écriture symbolique

Nous noterons la moyenne par le symbole μ, qui se lit «mu» (m dans l'alphabet grec), et le nombre de données par la lettre N.

En représentant chaque donnée de la série statistique par les symboles : x_1, x_2, x_3, x_4 et ainsi de suite, on obtient la formule suivante pour décrire le calcul d'une moyenne avec des données brutes :

$$\mu = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{N}$$

On peut simplifier l'écriture de cette formule à l'aide de la notation sigma, symbolisée par $\sum$ (S dans l'alphabet grec). Ce symbole indique que l'on doit faire la somme de tous les termes de forme x_i, l'indice i variant de 1 à N.

$$\mu = \frac{\sum x_i}{N}, \text{ pour } i \text{ variant de 1 à } N$$

Calcul de la moyenne lorsque les données sont non groupées (série statistique)

Exemple:
Soit la distribution statistique suivante concernant le nombre de programmes en techniques administratives offerts par 18 collèges :

Le nombre de programmes en techniques administratives offerts par 18 collèges

2	5	5	2	5	6
3	1	3	2	7	4
8	6	1	4	3	3

Moyenne 3,9

μ = 3,9 programmes.
Interprétation:
Les collèges offrent en moyenne 3,9 programmes en techniques administratives.

Moyenne d'un échantillon $\bar{x}$

NOTE

Dans le cadre d'un sondage, on emploie le symbole $\bar{x}$ (x barre) pour désigner la moyenne des données de l'échantillon et μ pour la moyenne des données de la population. De même, on utilise le symbole n pour désigner le nombre de données de l'échantillon et N pour le nombre de données de la population. La formule pour obtenir la moyenne des données d'un échantillon s'écrit donc :

$$\bar{x} = \frac{\sum x_i}{n}, \text{ pour } i \text{ variant de 1 à } n$$

Formule de la Moyenne μ (données groupées)

Moyenne avec les effectifs
$$\text{Moyenne} = \frac{\text{somme des produits de chaque valeur de la variable par son effectif}}{\text{nombre total de données}}$$

Écriture symbolique

En notant par la lettre n_i l'effectif de la valeur x_i, on obtient la formule suivante :

$$\mu = \frac{x_1 n_1 + x_2 n_2 + \cdots + x_k n_k}{N}, \text{ où } k \text{ représente le nombre de valeurs de la variable}$$

En utilisant la notation sigma, on obtient :

$$\mu = \frac{\sum x_i n_i}{N}, \text{ pour } i \text{ variant de 1 à } k$$

En notant f_i la fréquence relative de la valeur x_i, on obtient la formule suivante :

$$\boxed{\mu = x_1 f_1 + x_2 f_2 + \cdots + x_k f_k = \sum x_i f_i}$$

pour i variant de 1 à k, où k est le nombre de valeurs de la variable.

Calcul de la moyenne lorsque les données sont groupées (variable discrète)

Répartition de 40 collèges publics selon le nombre de programmes en techniques administrative, Québec, 2010

Nombre de programmes	Nombre de collèges
0	1
1	1
2	11
3	9
4	16
5	9
6	1

Démarche de calcul de la moyenne μ pour une variable discrète

Répartition de 48 collèges selon le nombre de programmes en techniques administratives, Québec, 2010

Nombre de programmes (xi)	Nombre de collèges (ni)	xini
0	1	0
1	1	1
2	11	22
3	9	27
4	16	64
5	9	45
6	1	6
Total	48	165

3,4

$$\mu = \frac{\Sigma xini}{N}$$

$\mu = 3{,}4$

<u>Interprétation:</u>
En 2010, les collèges publics québécois offrent en moyenne 3,4 programmes en techniques administratives.

Cette écriture n'est pas privilégiée.

$$\mu = \frac{0\times1+1\times1+2\times11+3\times9+4\times16+5\times9+6\times1}{48} = 3{,}4 \text{ programmes}$$

Exemple de calcul de la moyenne en utilisant le pourcentage (fréquence relative)

Répartition des ménages selon le nombre de vélos, Québec, 2010

Nombre de vélos	Pourcentage de ménages
0	21 %
1	20 %
2	27 %
3	13 %
4 et plus	19 %
Total	**100 %**

Source : Vélo Québec. *État de la pratique du vélo au Québec en 2010*, mai 2011.

Calculer la moyenne en remplaçant la catégorie « 4 et plus » par 4.

μ = 0x21% + 1x20% + 2x27% + 3x13% + 4x19% = 1,9 vélos

Dans les exercices et les évaluations, cette écriture n'est pas privilégiée, insérez une colonne *xifi comme indiqué dans ce tableau* !

Répartition des ménages selon le nombre de vélos, Québec, 2010

Nombre de vélos (xi)	0	1	2	3	4	Total
Pourcentage (fi)	21%	20%	27%	13%	19%	100%
xifi	0	0,2	0,54	0,39	0,76	1,9

$\mu = \sum x_i f_i$ = 1,9 vélos

Interprétation :
En 2010, les ménages au Québec ont en moyenne 1,9 vélo.

Calcul de la moyenne μ dans le cas d'une variable continue

Exemple (manuel, p. 43)

Pour calculer la moyenne, il faut calculer **le centre de chaque classe.**

Répartition des 40 clients de l'échantillon selon la durée du branchement au réseau Wi-Fi

Durée du branchement (en minutes)	Nombre de clients
$25 \leq X < 30$	2
$30 \leq X < 35$	4
$35 \leq X < 40$	9
$40 \leq X < 45$	12
$45 \leq X < 50$	8
$50 \leq X < 55$	3
$55 \leq X < 60$	2
Total	**40**

Démarche de calcul de la moyenne pour une variable continue

Répartition des clients selon la durée du branchement au réseau Wi-Fi

Durée du branchement (xi)	Nombre de clients (ni)	Centre de classe (ci)	cini	
25≤X<30	2	27,5	55	
30≤X<35	4	32,5	130	
35≤X<40	9	37,5	337,5	
40≤X<45	12	42,5	510	
45≤X<50	8	47,5	380	
50≤X<55	3	52,5	157,5	
55≤X<60	2	57,5	115	
Total	**40**		1685	42,1

$\mu = \frac{\sum c_i n_i}{N}$ = 1685/40

μ = 42,1 minutes

<u>Interprétation :</u>

La durée moyenne du branchement au réseau des 40 clients est de 42,1 minutes.

Le mode (variable discrète)

Définition, manuel p. 47.

Le **mode** est la valeur ou la catégorie qui revient le plus souvent dans une série statistique. La **classe modale** est la classe qui regroupe le plus de données. On considère le centre de cette classe comme une approximation du mode de la distribution.

Répartition des 48 collèges publics selon le nombre de programmes en techniques administratives offerts, Québec, 2010

Nombre de programmes en techniques administratives	0	1	2	3	4	5	6	Total
Nombre de collèges	1	1	11	9	16	9	1	**48**
Pourcentage de collèges	2,1 %	2,1 %	22,9 %	18,8 %	33,3 %	18,8 %	2,1 %	**100,0 %**

Solution

Le mode est 4 programmes.

Interprétation

En 2010, une pluralité de collèges publics québécois (33,3 %) offrent quatre programmes en techniques administratives.

La classe modale (variable continue)

Déterminer et interpréter la classe modale de la distribution suivante.

Répartition des 40 clients de l'échantillon selon la durée du branchement au réseau Wi-Fi

Durée du branchement (en minutes)	Nombre de clients	Pourcentage de clients
$25 \leq X < 30$	2	5,0 %
$30 \leq X < 35$	4	10,0 %
$35 \leq X < 40$	9	22,5 %
$40 \leq X < 45$	12	30,0 %
$45 \leq X < 50$	8	20,0 %
$50 \leq X < 55$	3	7,5 %
$55 \leq X < 60$	2	5,0 %
Total	**40**	**100,0 %**

La classe modale est $40 \leq X < 45$.

Interprétation

Une pluralité des clients de l'échantillon (30 %) ont été connectés au réseau pendant 40 à 45 minutes.

Médiane (Me): la valeur qui partage une série de données ordonnées en deux parties égales

Mise en situation:

Voici les salaires annuels de trois employés d'une entreprise: 130 000 \$ 40 000 \$ 37 000 \$

μ= (130 000 + 40 000 + 37 000)/3 = 70 000 \$.

Est-ce que Le salaire moyen est représentatif des salaires des employés de l'entreprise? (la médiane 40 000 \$ est plus représentative)

Exemple :

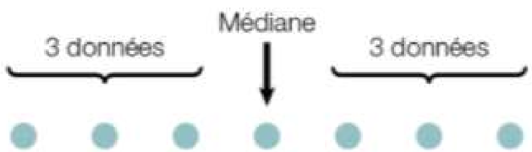
Déterminer la médiane (Me) de ces observation: 15   10   12   9   7

Si on ordonne les observations on aura: 7   9   10   12   15

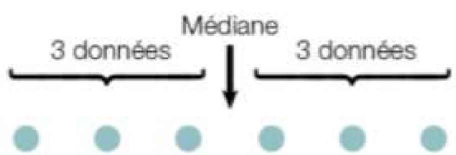
Me= 10, car il y a deux observations à gauche de 10 (inférieurs à 10) et deux observations à droite de 10 (supérieurs à 10)

Détermination de la médiane selon que le nombre de données est pair ou impair dans une série statistique

- Pour une série statistique comptant un nombre impair de données ordonnées, pour qu'il y ait autant de données à gauche qu'à droite de la médiane, cette dernière doit être égale à la donnée centrale de la série. À titre d'exemple, pour sept données ordonnées, la médiane est la 4[e] donnée.



- Pour une série statistique comptant un nombre pair de données ordonnées, pour qu'il y ait autant de données à gauche qu'à droite de la médiane, il faut choisir une valeur comprise entre les deux données centrales. Par convention, la médiane est égale à la valeur moyenne des deux données centrales de la série. À titre d'exemple, pour six données ordonnées, la médiane est la moyenne de la 3[e] et de la 4[e] donnée.



Exemple de détermination de la médiane pour des données non groupées (manuel, p. 50)

Trouver la médiane des séries statistiques suivantes.

a) Salaire de cinq ingénieurs travaillant pour une même entreprise :

41 500 $ 42 250 $ 58 550 $ 64 750 $ 120 800 $

La médiane est ________________. C'est la meilleure mesure de tendance centrale de la série, car elle n'est pas influencée par les valeurs extrêmes, contrairement à la moyenne.

Interprétation

Dans cette entreprise, au moins 50 % des ingénieurs gagnent 58 550 $ ou moins.

NOTE

On dit « au moins » 50 %, car trois ingénieurs sur cinq gagnent 58 550 $ ou moins.

Détermination de la médiane pour des données groupées (variable discrète) : Exemple (Manuel, p. 51)

4 4 4 5 5 5 5 $\boxed{5}$ 5 6 6 6 6 7 7

Caractère statistique étudié

Répartition des Centres de la petite enfance (CPE) d'une ville selon le nombre d'éducatrices

Nombre d'éducatrices	4	5	6	7	Total
Nombre de CPE	3	6	4	2	$\boxed{15}$
Pourcentage de CPE	20,0 %	40,0 %	26,7 %	13,3 %	100,0 %

15/2 = 7,5
La médiane correspond à la valeur de la 8ème donnée.

Effectif cumulé= 9

Solution

- Calcul de la médiane avec les effectifs :
  Avec un total de 15 données, un nombre impair, la médiane est la valeur de la 8e donnée.
  Médiane = 5 éducatrices
- Calcul de la médiane avec les pourcentages :
  Pour obtenir un pourcentage cumulé d'au moins 50 %, il faut additionner les pourcentages associés aux valeurs 4 et 5, soit 20 % + 40 %, ce qui donne 60 %, une valeur supérieure à 50 %.
  Médiane = 5 éducatrices

Détermination de la médiane dans le cas d'une variable continue, manuel p. 54 (détails de la méthode de détermination de la médiane)

Répartition des clients selon la durée du branchement au réseau Wi-Fi

Durée du branchement (xi)	Nombre de clients (ni)	Pourcentage	Pourcentage cumulé croissant
25≤X<30	2	5,0%	5,0%
30≤X<35	4	10,0%	15,0%
35≤X<40	9	22,5%	37,5%
40≤X<45	12	30,0%	67,5%
45≤X<50	8	20,0%	87,5%
50≤X<55	3	7,5%	95,0%
55≤X<60	2	5,0%	100,0%
Total	**40**		

Méthode de l'histogramme

Pour calculer la médiane:
- Calculer le pourcentage
- Calculer le pourcentage cumulé croissant

Classe médiane: [40;45[

En utilisant le pourcentage, on sait que :
50%= 5% + 10% + 22,5% + S
S = 12,5%
On doit déterminer le nombre de minutes correspondant à 12,5% selon la règle de 3.

30% ....5 min
12,5% ....X ?

X = (12,5% x 5)/30%
X = 2,1 minutes

Me= 40 + 2,1 = 42,1 minutes

Détermination de la médiane: méthode de l’interpolation linéaire

Méthode de l'interpolation linéaire: (on utilise le pourcentage cumulé)

La médiane (Me) appartient à la classe [40;45[, car c'est la première classe qui totalise un effectif cumulé de 50%

40................Me............45

37,5% ....50%..........67,5%

$$\frac{67{,}5\% - 37{,}5\%}{45-40} = \frac{67{,}5\% - 50\%}{45 - Me}$$

La résolution de cette équation :

0,3 (45-Me) = 0,875

45-Me = 2,91

Me = 42,1 minutes

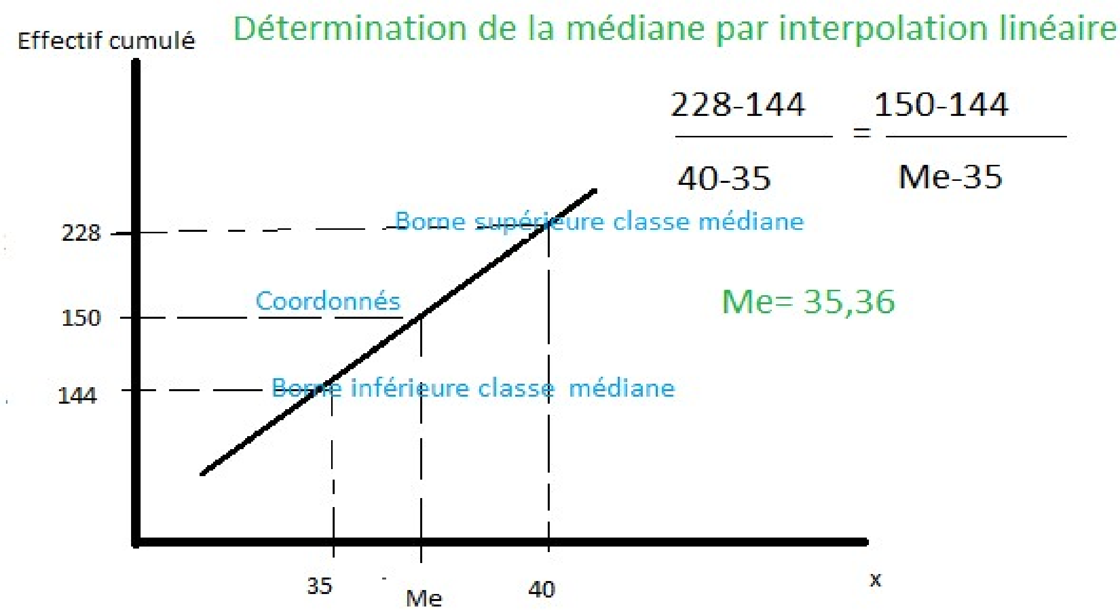
Exemple 2 pour déterminer la médiane

Répartition des employés d'une entreprise selon l'âge

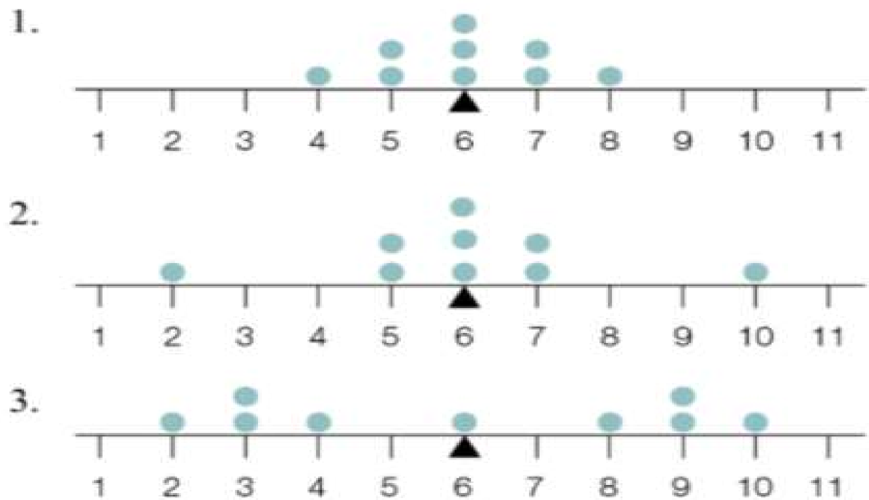
Âge	Effectif	Effectif cumulé
Moins de 25 ans	18	18
25≤ X <30	54	72
30≤ X <35	72	144
35≤ X <40	84	228
40≤ X <45	36	264
45≤ X <50	22	286
50 ans et plus	14	300
Total	300	

Classe médiane = classe où se situe la 150ème donnée (300/2).
C'est la classe [35; 40[

Détermination de la médiane par interpolation linéaire

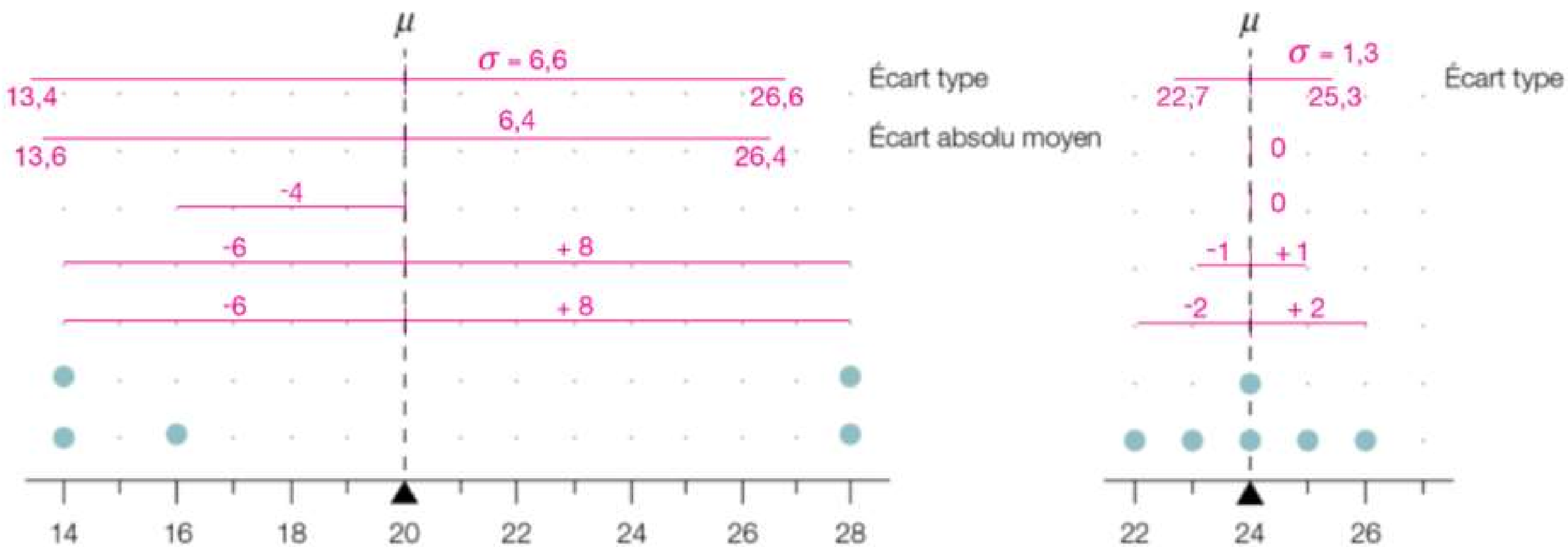


Mesures de dispersion: l’étendue et l’écart-type



Ces trois séries statistiques présentent des moyennes identiques, mais la dispersion des données autour de ces moyennes diffère d’une série à l’autre. Les données de la série 1 sont assez concentrées autour de la moyenne, alors que la série 2 comporte deux données qui sont passablement éloignées de la moyenne. Les données de la série 3 sont encore plus dispersées par rapport à la moyenne que celles de la série 2. Comment peut-on mesurer mathématiquement cette dispersion ?

L’écart type σ (sigma) dans le cas de données non groupées



Écart type = $\sqrt{[(-6)^2 + (-6)^2 + (-4)^2 + (8)^2 + (8)^2]/5}$=6,6 ans

La plupart des données (3 sur 5) sont situées à ± 6,6 ans de la moyenne, soit entre 13,4 ans et 26,6 ans.

Formule de calcul de l'écart type σ

Écart type et variance

$$\text{Écart type } (\sigma) = \sqrt{\text{variance } (\sigma^2)} = \sqrt{\frac{\text{somme des carrés des écarts à la moyenne}}{\text{nombre total de données}}}$$

$$\text{Écart type } (\sigma) = \sqrt{\frac{\sum (x_i - \mu)^2 n_i}{N}} \quad \left(\text{ou } \sigma = \sqrt{\sum (x_i - \mu)^2 f_i} \text{ avec les fréquences relatives}\right)$$

Dans le cas de données groupées, on multiplie par *ni*

Calcul de l'écart type pour une variable discrète

Valeur (xi)	6	8	14	20	**Total**	
Effectif (ni)	3	3	2	1	**9**	
xini	18	24	28	20	**90**	10
xi-μ	-4	-2	4	10		
(xi-μ)^2	16	4	16	100		
(xi-μ)^2 ni	48	12	32	100	192	21,3333333
						4,61880215

$\mu = \frac{\Sigma xini}{N}$

μ = 90/9 = 10

$\sigma^2 = \frac{\sum (x_i - \mu)^2 n_i}{N}$ = 192/9 = 21,3

Donc, $\boldsymbol{\sigma = \sqrt{21,3} = 4,6}$

<u>Interprétation:</u>

la plupart des données se situent entre ± 1σ de la moyenne μ. Dans ce cas, la plupart des données se situent entre 5,4 et 14,6.

Calcul de l'écart type pour une variable continue (exemple, p.65)

Exemple, p.65

Variable	Effectif	ci	cixi	ci-µ	$(ci-\mu)^2$	$(ci-\mu)^2 ni$
[3;5[	1	4	4	-4	16	16
[5;7[	2	6	12	-2	4	8
[7;9[	4	8	32	0	0	0
[9;11[	2	10	20	2	4	8
[11;13[	1	12	12	4	16	16
	10		80			48
			8			4,8
						2,19089023

$\mu = \frac{\sum C_i n_i}{N} = 80/10$

$\mu = 8$

$\sigma^2 = \frac{\sum (ci-\mu)^2 n_i}{N}$

$\sigma^2 = 4,8$

$\sigma = \sqrt{4,8} = 2,2$

Interprétation

La plupart des données se situent entre 5,8 et 10,2.

Écart type corrigé noté s (exemple, p.67)

Écart type corrigé dans le cas d'un échantillon

Dans le cadre d'une étude par sondage, l'écart type de l'échantillon est retenu pour estimer celui de la population. Toutefois, les statisticiens ont démontré que l'estimation est meilleure si l'on divise le numérateur de la formule de l'écart type par le nombre de données de l'échantillon, moins 1. On obtient ainsi l'écart type corrigé, qui prend la notation s. En se rappelant que la moyenne de l'échantillon se note $\overline{x}$ et le nombre de données n, on obtient la formule suivante pour calculer l'écart type corrigé :

Écart type corrigé et variance corrigée

$$\text{Écart type corrigé } (s) = \sqrt{\text{variance corrigée } (s^2)} = \sqrt{\frac{\text{somme des carrés des écarts à la moyenne}}{\text{nombre total de données} - 1}}$$

$$\text{Écart type corrigé } (s) = \sqrt{\frac{\sum (x_i - \overline{x})^2 n_i}{n-1}}$$

Travail à faire

Exercices 1.3 : 5, 9 p. 59-60.

Exercices 1. 4 : 2, 4, 5 et 11 p. 72-73.

Remarque importante :

✓ Illustrer votre démarche en précisant vos formules de calcul et en insérant les colonnes nécessaires dans votre tableau de distribution.