Séries regroupées par modalités.

I Moyenne.

Définition 1

Les *modalités* d'une série statistiques sont les valeurs différentes de la série.
À la *série des modalités* sont associées les *séries des effectifs* (ou fréquences) et *effectifs cumulés croissants* (ou fréquences cumulées croissantes).

Lorsque la série est regroupée par modalités et série des effectifs

Modalités	x_1	$\dots$	x_p
Effectifs	n_1	$\dots$	n_p

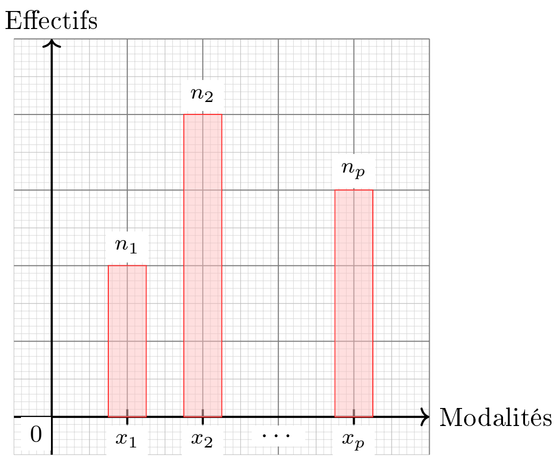
pour calculer la moyenne nous utiliserons la formule de la *moyenne pondérée*

$$\overline{x} = \frac{n_1 \cdot x_1 + \cdots + n_p \cdot x_p}{n_1 + \cdots + n_p}$$

Remarques.

1. Cette formule n'est qu'une autre présentation de la formule de moyenne que vous connaissez déjà.
2. Cette formule est encore valable si à la place des effectifs sont donnés des fréquences (par exemple des pourcentages). Rappel : la fréquence se calcule comme le quotient de l'effectif par l'effectif total.

La représentation graphique d'une série regroupée par modalités et effectifs la plus couramment utilisée est celle du diagramme en barre.
La hauteur d'une barre indique l'effectif correspondant à la modalité.



Lorsque les effectifs sont présentés sous forme de pourcentages le diagramme circulaire est souvent utilisé.

II Écart type.

Définition 2

La *variance*, notée V et donnée par

$$V(x) = \frac{n_1(\overline{x} - x_1)^2 + n_2(\overline{x} - x_2)^2 + \cdots + n_p(\overline{x} - x_p)^2}{n_1 + n_2 + \cdots + n_p}$$

L'*écart type*, noté σ, est une caractéristique de dispersion :

$$\sigma(x) = \sqrt{V(x)}$$

.

Remarques.

1. L'écart-type représente l'éloignement (écart) moyen des valeurs avec la moyenne. Plus l'écart-type est grand plus les valeurs sont dispersées autour de la moyenne.
2. Une grande proportion de données se trouve dans l'intervalle centré sur la moyenne et de rayon 2σ. Ceci permet d'enlever les valeurs extrêmes de la série (très petites et très grandes) souvent qualifiées d'aberrantes.
3. Nous pourrions être tentés de définir l'écart-type comme la moyenne des écarts par rapport à la moyenne (sans mettre au carré donc) mais cela donne toujours 0. Il a donc fallu considérer les carrés des écarts.
4. Pourquoi ne pas se contenter de la variance ? Par ce que la variance est dans une unité étrange. Si la série représente des prix, la variance s'exprime en euro au carré (?). Pour obtenir des euros il faut donc considérer la racine carrée et donc calculer l'écart-type.
5. Pour calculer l'écart-type vous userez de la calculatrice (ou du tableur) et non de la formule.
6. Pour déterminer les indicateurs statistiques d'une série regroupée par modalités nous utiliserons la calculatrice.

   Par exemple pour déterminer la moyenne et l'écart-type de la série

Valeurs x_i	2	4	5	7	8	9
Effectif n_i	4	2	2	2	1	1

Entrez en liste L1 les valeurs x_i et en L2 les effectifs n_i. Puis faites la commande :

On obtient la moyenne : $\overline{x} = 4{,}75$ et l'écart-type $\sigma(x) \approx 2{,}4195$.

III Exercices.

Exercice 1. ♥

Voici les notes obtenues au premier devoir de mathématiques de l'année.

Notes	4	5	6	7	8	9	10	11	12
Effectif	2	1	3	6	2	2	1	1	5

Notes	13	14	15	16	17	18	19	20
Effectif	1	2	1	4	2	0	1	1

1. Calculez la fréquence du mode de cette série. (Le mode est la modalité ayant l'effectif le plus important).
2. Calculez la moyenne de cette série.
3. Calculez l'étendue.
4. Calculez la médiane ainsi que les quartiles Q_1 et Q_3.
5. Déterminez l'écart inter-quartile.
6. Déterminez l'écart type de la série.

Correction exercice 1

1. Calculons la fréquence du mode de la série.

   Le mode de la série est le 7 qui apparait 6 fois. Or l'effectif total est $N = 2 + 1 + 3 + 6 + \cdots + 1 + 1 = 35$, donc sa fréquence est

   $$\boxed{f_{Mod} = \frac{6}{35} \approx 0{,}1714.}$$

2. Calculons la moyenne.

   La série étant regroupée par modalités nous utiliserons la formule de la moyenne pondérée.

$$\begin{aligned}\overline{x} &= \frac{n_1x_1 + n_2x_2 + \cdots + n_rx_r}{n_1 + n_2 + \cdots + n_r}\\ &= \frac{2\times 4 + 1\times 5 + \cdots + 1\times 20}{2 + 1 + \cdots + 1}\\ &= \frac{381}{35}\\ &\approx 10{,}89\end{aligned}$$

La moyenne des notes de la classe est

$$\boxed{\overline{x} \approx 10{,}89 \text{ points.}}$$

3. L'étendue est $e = max - min = 20 - 4 = 16$.

$$\boxed{e = 16 \text{ points.}}$$

4. La série étant regroupée par modalités, nous aurons besoin pour déterminez les médianes et quartiles de la série des effectifs cumulés croissants.

Notes	4	5	6	7	8	9	10	11	12
Effectif	2	1	3	6	2	2	1	1	5
E.C.C.	2	3	6	12	14	16	17	18	23

Notes	13	14	15	16	17	18	19	20
Effectif	1	2	1	4	2	0	1	1
E.C.C.	24	26	27	31	33	33	34	35

(a) Recherchons la médiane.

La série des notes est ordonnée.
$\frac{N}{2} = \frac{35}{2} = 17{,}5$ donc le médiane est la dix-huitième valeur (série impaire).
D'après les E.C.C.

$$\boxed{Me = 11 \text{ points.}}$$

Rappelons que cela signifie que 50 % des élèves ont obtenu une note inférieure à 11.

(b) Déterminons Q_1.

La série est ordonnée.
$\frac{N}{4} = \frac{35}{4} = 8{,}75$. Q_1 est donc la neuvième valeur.
D'après les E.C.C.

$$\boxed{Q_1 = 7 \text{ points.}}$$

Rappelons que cela signifie que 25 % des élèves ont obtenu une note inférieure à 7.

(c) Déterminons Q_3.

La série est ordonnée.
$\frac{3}{4} \times N = \frac{3}{4} \times 35 = 26{,}75$. Q_3 est donc la vingt-septième valeur de la série.
D'après les E.C.C.

$$\boxed{Q_3 = 15 \text{ points.}}$$

Rappelons que cela signifie que 75 % des élèves ont obtenu une note inférieure à 15.

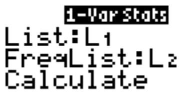
5. Déterminons l'écart inter-quartile.

$Q_3 - Q_1 = 15 - 7$.

$$\boxed{Q_3 - Q_1 = 3 \text{ points.}}$$

Ceci peut s'interpréter en que 50 % des notes de la classe se trouvent autour de la moyenne avec de 3 points maximum entre elles.

6. Avec la calculatrice : il faut entrer les listes L_1 et L_2 (listes de l'exercice).
La fonction calculant les indicateurs statistiques s'obtient en faisant $\boxed{\text{Stats}}$, (CALC), 1 :Stat-1-var.



Exercice 2. Application.
Sur le site internet d'une enseigne de vente, on a relevé le prix (en euros) de 43 casques intra-auriculaires.

Prix, x_i	8	10	13	15	18	20	21	25
Effectif, n_i	2	2	3	15	4	15	1	1

Déterminez la moyenne et l'écart-type de la série statistique obtenue.

Correction exercice 2

1. Calculons $\overline{x}$.

   La série est donnée regroupée par modalités (*i.e.* selon les valeurs identiques et avec les effectifs) donc on la calcule avec la formule de la moyenne pondérée.

$$\begin{aligned}\overline{x} &= \frac{n_1x_1 + n_2x_2 + \cdots + n_px_p}{n_1 + n_2 + \cdots + n_p} \\ &= \frac{2 \times 8 + 2 \times 10 + \cdots + 1 \times 25}{2 + 2 + \cdots + 1} \\ &\approx 16{,}70\end{aligned}$$

   Le prix moyen d'un casque est de 16,70 euros.

2. Calculons l'écart-type.

   Rappelons la formule de la variance puis le lien avec l'écart-type.

$$V(x) = \frac{n_1(\overline{x} - x_1)^2 + n_2(\overline{x} - x_2)^2 + \cdots + n_p(\overline{x} - x_p)^2}{n_1 + n_2 + \cdots + n_p}$$

   Et :

$$\begin{aligned}\sigma(x) &= \sqrt{V} \\ &\approx 3{,}68\end{aligned}$$

   L'écart-type des prix est de 3,68 euros.

Exercice 3. ♥

Julien est animateur dans un centre aéré tous les mercredis. En début d'année, des groupes d'enfants sont constitués et pris en charge par les différents animateurs.

Le directeur sait que Julien préfère s'occuper d'enfants d'âges voisins de 6 ans. Il lui propose de choisir entre deux groupes et lui soumet les informations ci-contre.

	Groupe 1	Groupe 2
Moyenne	6 ans	7 ans
Écart-type	5 ans	1 an

Quel groupe Julien va-t-il choisir ?

Correction exercice 3

Les moyennes des deux groupes sont proches.

Par contre l'écart-type du groupe 1 est beaucoup plus important. Les valeur de cette série sont donc bien plus dispersées. Autrement dit les écarts d'âge dans le groupe 1 sont très importants.

Julien préférera donc travailler avec le groupe 2.

Exercice 4. ♥

1. Un régleur tourneur a reçu comme instruction d'affiner les réglages de son tour s'il observe, dans un échantillon de 20 pièces usinées choisies aléatoirement, l'un des cas suivants.

   - Moins de 68 % des pièces ont un diamètre qui appartient à l'intervalle $[\overline{x} - \sigma ; \overline{x} + \sigma]$.
   - Plus de 60 % des diamètres sont supérieurs au diamètre moyen de l'échantillon.

   À 17 h, il relève les diamètres (en cm) suivant.

   23,5 ; 23,8 ; 24,7 ; 25,1 ; 25,1 ; 25,1 ; 25,2 ;
   25,2 ; 25,3 ; 25,3 ; 25,4 ; 25,4 ; 25,4 ; 25,5 ;
   25,8 ; 25,9 ; 26,0 ; 26,0 ; 26,3 ; 27,1.

   Avec sa calculatrice il obtient

   ```
   x̄=25.355
   Σx=507.1
   Σx²=12869.35
   Sx=0.7890533969
   σx=0.7690741187
   n=20
   ```

   Ce régleur doit-il modifier les réglages de sa machine ?

2. Une erreur de calibrage du à coulisse modifie la série. Chaque mesure doit être amputée de 0,1 cm. Cela modifie-t-il la décision du régleur ? Justifiez.

Exercice 5. Application.

Les tailles sont exprimées en centimètre.

Partie A : à la maternité « Beaux jours ».
Sur la totalité du mois de janvier 2012, il y a eu
57 nouveau-nés à la maternité « Beaux jours ».
Leur taille est donnée dans le tableau ci-dessous.

Taille	46	47,5	48	48,5	49	49,5	50
Effectifs	1	2	3	5	5	7	9

Taille	50,5	51	51,5	52	52,5	53	
Effectifs	8	7	5	2	2	1	

1. Calculer la moyenne puis la médiane des tailles de ces 57 nouveau-nés en précisant la démarche.
2. Calculer le pourcentage de nouveau-nés ayant une taille inférieure ou égale à 49 cm.
   Donner la réponse arrondie à 0,1 %.
3. Parmi toutes ces tailles, déterminer la plus petite taille t telle qu'au moins les trois quarts des nouveau-nés aient une taille inférieure ou égale à t cm.
   Quel paramètre de la série des tailles a été ainsi trouvé ?

Partie B : à la maternité « Bon accueil ».
L'étude statistique de la taille des 64 nouveau-nés
durant le même mois de janvier 2012 à la maternité « Bon accueil » a donné les résultats suivants :

- Minimum : 46
- Maximum : 53
- Moyenne : 49,3
- Médiane : 49
- 1^{er} quartile : 48
- 3^{e} quartile : 50,5

1. Des deux maternités, une seule possède un service pour les naissances prématurées.
   Les résultats précédents permettent-ils de trouver laquelle ? Justifier votre réponse.
2. Les deux maternités sont les seules de la ville.
   (a) Calculer la moyenne des tailles des nouveau-nés, en janvier 2012, dans les maternités de cette ville.
   (b) Les données de l'énoncé permettent-elles de déterminer la médiane des tailles des nouveau-nés des deux maternités réunies ?
   Si oui, la déterminer ; sinon expliquer pourquoi.

Correction exercice 5

1. $\overline{x} \approx 50$.

$Me = 50$.

2. 28,1 %.
3. $Q_3 = 51$.

Exercice 6. Application.

Sébastien, étudiant de 19 ans, veut s'inscrire dans une station balnéaire pour un séjour d'été où il aurait des chances de rencontrer des jeunes femmes de son âge.
Prenant quelques références, les stations lui fournissent la moyenne d'âge des inscrites.
Station A : 19 ans. Station B : 31 ans.
Sans hésiter, il s'inscrit dans la station A !

1. Le choix de Sébastien est-il judicieux ?

Les tableaux ci-dessous indiquent les âges des inscrites dans les deux stations.

Station A

Âge	Effectif
2	3
4	1
5	1
7	1
10	1
11	2
34	1
35	2
50	1
58	1

Station B

Âge	Effectif
18	1
19	5
20	2
45	2
46	1
47	1
48	1
50	1

2. Pour les deux stations :
   (a) donner la fréquence de la valeur 19 ;
   (b) calculer la médiane et les quartiles ;
   (c) calculer l'étendue ;
   (d) déterminer la modalité de la plus grande fréquence.
3. Finalement, le choix de Sébastien est-il judicieux ? Argumenter.
4. Écrire l'algorithme que Sébastien a utilisé pour calculer la fréquence de la valeur 19.

Correction exercice 6

Il ne s'agit pas d'une correction rédigée mais simplement d'éléments de réponses.

1. Le choix de Sébastien n'est pas nécessairement judicieux : la moyenne ne révèle pas les écarts d'âge. Il faudrait prendre en compte des indicateurs de dispersion.
2. (a) $f_A = 0$ et $f_B \approx 0{,}3571$. Effectif total 14

(b) $Q_{1A} = 4$, $Me_A = 10{,}5$, $Q_{3A} = 35$.
$Q_{1B} = 19$, $Me_B = 20$, $Q_{3B} = 46$.

(c) $e_A = 58 - 2 = 56$ et $e_B = 50 - 18 = 32$.

(d)

3. Le choix de Sébastien n'est pas judicieux. Les indicateurs de dispersion sont plus petits pour la station B.
$\overline{x}_A = 19$ et $\sigma_B \approx 18{,}63$
$\overline{x}_B = 31$ et $\sigma_B \approx 13{,}77$

Exercice 7. Application.

En 2016, une entreprise employait deux cadres et six ouvriers.

Catégorie	Cadre	Ouvrier
Salaire (2016)	6 000 €	1 400 €

1. Calculez le salaire moyen d'un employé.
2. Les résultats de l'entreprise étant bons, le chef d'entreprise décide d'augmenter tous les salaires de 10 % et d'engager cinq ouvriers supplémentaires. Un employé affirme : « Malgré nos augmentations, le salaire moyen de l'entreprise a baissé. » S'est-il trompé ? Justifiez.

Correction exercice 7

1. $\overline{x}_1 = \frac{2\times6000+6\times1400}{2+6}$.
2. $\overline{x}_2 = \frac{2\times(1{,}1\times6000)+(6+5)\times1{,}1\times1400}{2+(6+5)}$

IV Ce qu'il faut retenir.

1. Lorsqu'une série est présentée avec ses modalités penser à utilise les effectifs et les effectifs cumulés croissants pour les médianes et quartiles.
2. Déterminer médianes, quartiles et moyenne (formule de la moyenne pondérée).
3. Calculer l'écart type à la calculatrice et savoir ce qu'il représente.
4. Représentation graphiques : camembert (pourcentages), diagramme en barres.