Plaquette 2 – Séries en classes, histogrammes et effectifs cumulés

Statistiques descriptives — Seconde

Corrections détaillées

Boîte à outils

Moyenne d'une série en classes. On remplace chaque classe par son centre c_i :

$$\overline{x} = \frac{n_1 c_1 + n_2 c_2 + \cdots + n_p c_p}{N} \qquad N = n_1 + \cdots + n_p$$

Le résultat est une **approximation** : on suppose les valeurs réparties uniformément dans chaque classe.

Effectifs et fréquences cumulés croissants.

Notation	Définition
N_k	$n_1 + n_2 + \cdots + n_k$ — nombre de valeurs < borne haute de la classe k
F_k	$\dfrac{N_k}{N}$ — même chose en fréquence
f_i	$\dfrac{n_i}{N}$ — fréquence de la classe i

On a toujours $N_p = N$ et $F_p = 1$.

Médiane et quartiles par interpolation linéaire. Si la classe $[a\,;b[$ contient le rang cherché r, avec N_{k-1} valeurs avant et n_k valeurs dans la classe :

$$\text{valeur} \approx a + (b-a) \times \frac{r - N_{k-1}}{n_k}$$

— pour la **médiane**, $r = \dfrac{N}{2}$;

— pour Q_1, $r = \dfrac{N}{4}$; pour Q_3, $r = \dfrac{3N}{4}$.

Histogramme. L'aire de chaque rectangle est proportionnelle à l'effectif. Si les classes n'ont pas la même amplitude, la hauteur est la **densité** :

$$\text{densité} = \frac{\text{effectif}}{\text{amplitude}}$$

Moyenne de sous-groupes. Deux groupes d'effectifs n_1 et n_2, de moyennes $\overline{x_1}$ et $\overline{x_2}$:

$$\overline{x} = \frac{n_1 \overline{x_1} + n_2 \overline{x_2}}{n_1 + n_2}$$

Ce n'est **pas** la moyenne des deux moyennes, sauf si $n_1 = n_2$.

Correction de l'exercice 1 – Moyenne d'une série en classes

Pour une série regroupée en classes, on ne connaît pas les valeurs exactes : on remplace chaque valeur par le **centre** de sa classe, ici 5, 15, 25 et 35. La moyenne est alors une moyenne pondérée par les effectifs :

$$\overline{x} = \frac{8\times 5 + 12\times 15 + 15\times 25 + 5\times 35}{40} = \frac{40 + 180 + 375 + 175}{40} = \frac{770}{40} = 19{,}25 \text{ min}$$

C'est une **approximation**, puisqu'on suppose les durées réparties autour du centre de chaque classe. **Contrôle** : en prenant les bornes basses puis hautes des classes, la moyenne serait comprise entre 14,25 et 24,25 min ; 19,25 est bien entre les deux.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Réponse. $\overline{x} = \dfrac{770}{40} = 19{,}25$ minutes.

Correction de l’exercice 2 – Classe modale et fréquences

La classe modale est celle qui a le plus grand effectif : $[20\,;30[$, avec 15 salariés (les classes ont toutes la même amplitude, la comparaison est donc légitime). Les fréquences s’obtiennent en divisant chaque effectif par 40 :

Durée	Effectif	Fréquence
$[0\,;10[$	8	20 %
$[10\,;20[$	12	30 %
$[20\,;30[$	15	37,5 %
$[30\,;40[$	5	12,5 %

Contrôle : la somme des fréquences vaut 100 %. Les salariés qui mettent moins de 20 minutes sont ceux des deux premières classes, soit $8 + 12 = 20$ salariés :

$$\frac{20}{40} = 50\,\%$$

Réponse. Classe modale $[20\,;30[$; fréquences 20 ; 30 ; 37,5 ; 12,5 % ; 50 % des salariés sous 20 minutes.

Correction de l’exercice 3 – Médiane par interpolation linéaire

On calcule les effectifs cumulés croissants : 6 ; 20 ; 42 ; 50. La médiane partage l’effectif en deux : elle correspond au rang $\dfrac{50}{2} = 25$. Comme $20 < 25 \leq 42$, elle se trouve dans la classe $[10\,;15[$.

Pour l’estimer, on suppose les 22 notes de cette classe **réparties régulièrement** sur ses 5 points. Il faut avancer de $25 - 20 = 5$ notes sur 22 dans la classe, soit une fraction $\dfrac{5}{22}$ de sa largeur :

$$\text{Méd} \approx 10 + 5 \times \frac{5}{22} \approx 11{,}14$$

Contrôle : 40 % des élèves ont moins de 10 et 84 % moins de 15 ; la médiane est donc dans la classe, plutôt près de 10.

Réponse. $\text{Méd} \approx 10 + \dfrac{25}{22} \approx 11{,}14$.

Correction de l’exercice 4 – Quartiles par interpolation

On utilise les mêmes effectifs cumulés, 6 ; 20 ; 42 ; 50, et la même hypothèse de répartition régulière dans chaque classe.

Q_1 correspond au rang $\dfrac{50}{4} = 12{,}5$, dans la classe $[5\,;10[$ puisque $6 < 12{,}5 \leq 20$: on avance de 6,5 notes sur 14.

$$Q_1 \approx 5 + 5 \times \frac{6{,}5}{14} \approx 7{,}32 \qquad Q_3 \approx 10 + 5 \times \frac{17{,}5}{22} \approx 13{,}98$$

Pour Q_3, de rang 37,5, on avance de $37{,}5 - 20 = 17{,}5$ notes sur 22 dans $[10\,;15[$. L’écart interquartile vaut $13{,}98 - 7{,}32 \approx 6{,}66$: la moitié centrale du groupe a une note entre 7,3 et 14 environ. **Contrôle** : $Q_1 < \text{Méd} < Q_3$.

Réponse. $Q_1 \approx 7{,}32$, $Q_3 \approx 13{,}98$, écart interquartile $\approx 6{,}66$.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s’expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Correction de l’exercice 5 – Moyenne et comparaison avec la médiane

On remplace chaque classe par son centre : 2,5 ; 7,5 ; 12,5 ; 17,5.

$$\overline{x} = \frac{6 \times 2{,}5 + 14 \times 7{,}5 + 22 \times 12{,}5 + 8 \times 17{,}5}{50} = \frac{15 + 105 + 275 + 140}{50} = \frac{535}{50} = 10{,}7$$

La moyenne, 10,7, est un peu inférieure à la médiane, 11,14 : les notes faibles de $[0\,;5[$ tirent la moyenne vers le bas, alors qu’elles ne déplacent pas la médiane.

Contrôle : $\frac{22+8}{50} = 60\,\%$ des élèves ont au moins 10, il est donc cohérent que la médiane dépasse 10. Pour décrire l’élève « typique », la médiane est ici plus fidèle.

Réponse. $\overline{x} = \frac{535}{50} = 10{,}7$, légèrement **inférieure** à la médiane.

Correction de l’exercice 6 – Histogramme à classes inégales

Quand les classes n’ont pas la même amplitude, la hauteur d’une barre d’histogramme est la **densité**, effectif divisé par amplitude, afin que l’**aire** de la barre soit proportionnelle à l’effectif.

Âge	Amplitude	Effectif	Densité
$[0\,;10[$	10	20	2
$[10\,;20[$	10	30	3
$[20\,;40[$	20	40	2
$[40\,;80[$	40	40	1

La barre la plus haute est celle de $[10\,;20[$, bien que les deux dernières classes aient un effectif plus grand : elles s’étalent sur des intervalles plus larges. Avec les centres 5, 15, 30, 60 :

$$\overline{x} = \frac{100 + 450 + 1\,200 + 2\,400}{130} = \frac{4\,150}{130} \approx 31{,}9 \text{ ans}$$

Réponse. Densités 2 ; 3 ; 2 ; 1 — la plus haute barre est $[10\,;20[$; $\overline{x} \approx 31{,}92$ ans.

Correction de l’exercice 7 – Retrouver les effectifs depuis les cumuls

L’effectif d’une classe est la différence entre son effectif cumulé et celui de la classe précédente : $n_k = N_k - N_{k-1}$.

Classe	Cumul	Effectif
$[0\,;20[$	12	12
$[20\,;40[$	30	18
$[40\,;60[$	58	28
$[60\,;80[$	70	12
$[80\,;100]$	80	10

Contrôle : 12 + 18 + 28 + 12 + 10 = 80, l’effectif total. Avec les centres 10, 30, 50, 70, 90, la moyenne vaut

$$\overline{x} = \frac{120 + 540 + 1\,400 + 840 + 900}{80} = \frac{3\,800}{80} = 47{,}5$$

Réponse. Effectifs 12 ; 18 ; 28 ; 12 ; 10 ; moyenne 47,5.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s’expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Correction de l’exercice 8 – Fréquences cumulées

La fréquence cumulée d’une classe est son effectif cumulé divisé par l’effectif total, 80 :

Classe	Cumul	Fréquence cumulée
[0 ; 20[	12	15 %
[20 ; 40[	30	37,5 %
[40 ; 60[	58	72,5 %
[60 ; 80[	70	87,5 %
[80 ; 100]	80	100 %

Les mesures inférieures à 60 sont celles des trois premières classes : leur proportion est la fréquence cumulée en 60, soit $\frac{58}{80} = 72{,}5\,\%$. Aucune estimation n’est nécessaire, car 60 est une borne de classe. Les autres, 27,5 %, sont au-delà.

Réponse. F : 15 ; 37,5 ; 72,5 ; 87,5 ; 100 % — donc 72,5 % des mesures sont sous 60.

Correction de l’exercice 9 – Médiane graphique

La médiane correspond au rang $\frac{80}{2} = 40$. Les cumuls valent 12 ; 30 ; 58, donc elle est dans [40 ; 60[. En supposant les 28 mesures de cette classe réparties régulièrement sur 20 unités, on avance de $40 - 30 = 10$ mesures :

$$\text{Méd} \approx 40 + 20 \times \frac{10}{28} \approx 47{,}14$$

De même, Q_1 correspond au rang 20, dans [20 ; 40[($12 < 20 \leq 30$) :
$Q_1 \approx 20 + 20 \times \frac{8}{18} \approx 28{,}89$.

Contrôle : la moyenne vaut 47,5, très proche de la médiane : la série est presque symétrique.

Réponse. Méd $\approx 47{,}14$ et $Q_1 \approx 28{,}89$.

Correction de l’exercice 10 – Effectif manquant connaissant la moyenne

On remplace chaque classe par son centre, 5, 15 et 25, et l’on écrit que la moyenne vaut 17. L’effectif total est $30 + x$:

$$\frac{10 \times 5 + 15x + 20 \times 25}{30 + x} = 17 \Leftrightarrow 550 + 15x = 17(30 + x) = 510 + 17x$$

On obtient $40 = 2x$, donc $x = 20$.

Vérification : avec les effectifs 10, 20, 20, la somme vaut $50 + 300 + 500 = 850$ pour 50 valeurs, et $\frac{850}{50} = 17$. La moyenne 17 dépasse le milieu de l’étendue, 15 : il fallait beaucoup de valeurs dans les classes hautes.

Réponse. $x = 20$ (effectifs 10, 20, 20).

Correction de l’exercice 11 – Effet d’un regroupement de classes

On regroupe les effectifs, $8 + 12 = 20$ et $15 + 5 = 20$, et l’on prend les nouveaux centres 10 et 30 :

$$\overline{x} = \frac{20 \times 10 + 20 \times 30}{40} = \frac{800}{40} = 20 \text{ min}$$

L’estimation passe de 19,25 à 20 minutes, soit 0,75 min d’écart (environ 4 %). Chaque

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s’expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

centre de classe représente maintenant des valeurs étalées sur 20 minutes, et la répartition réelle à l'intérieur des classes est ignorée.

Plus les classes sont larges, moins l'estimation par les centres est précise : on calcule les indicateurs sur les classes les plus fines dont on dispose.

Réponse. La moyenne passe à 20 minutes, soit 0,75 min de plus : le regroupement dégrade la précision.

Correction de l'exercice 12 – Moyenne de deux groupes

La moyenne de la réunion n'est pas la moyenne des deux moyennes : il faut revenir aux **sommes** de notes. Pour la première classe, $28 \times 11{,}5 = 322$; pour la seconde, $32 \times 13 = 416$.

$$\overline{x} = \frac{322 + 416}{28 + 32} = \frac{738}{60} = 12{,}3$$

La moyenne simple des moyennes, 12,25, serait fausse : le groupe le plus nombreux, qui a aussi la meilleure moyenne, pèse davantage et tire le résultat vers le haut.

Contrôle : $11{,}5 < 12{,}3 < 13$. Avec deux groupes de 30 élèves, on retrouverait exactement 12,25.

Réponse. $\overline{x} = \frac{738}{60} = 12{,}3$.

Correction de l'exercice 13 – Somme des valeurs et ajout d'une donnée

On revient à la somme des valeurs : $\overline{x} = \frac{S}{N}$ donne $S = N \times \overline{x} = 24 \times 15{,}5 = 372$. On ajoute la nouvelle valeur et l'on divise par le nouvel effectif :

$$\overline{x'} = \frac{372 + 30}{25} = \frac{402}{25} = 16{,}08$$

La moyenne augmente de 0,58. **Contrôle** : la valeur ajoutée dépasse l'ancienne moyenne de $30 - 15{,}5 = 14{,}5$, et cet excédent se répartit sur 25 valeurs : $\frac{14{,}5}{25} = 0{,}58$. Une valeur égale à 15,5 n'aurait rien changé.

Réponse. $\overline{x'} = \frac{402}{25} = 16{,}08$.

Correction de l'exercice 14 – Résumé en cinq nombres

Les cinq nombres sont le minimum, Q_1, la médiane, Q_3 et le maximum. Pour une série en classes, les extrêmes ne sont connus qu'approximativement : on prend les bornes extrêmes des classes, 0 et 20. Avec les résultats des exercices 3 et 4 :

$$0 \ ; \ 7{,}32 \ ; \ 11{,}14 \ ; \ 13{,}98 \ ; \ 20$$

La boîte va de 7,32 à 13,98, avec un trait à la médiane 11,14, un peu à droite du centre de la boîte. Les moustaches vont de 0 à 7,32 et de 13,98 à 20, de longueurs comparables (7,32 et 6,02). La boîte, de largeur 6,66, ne couvre qu'un tiers de l'étendue : la moitié centrale des notes est assez concentrée.

Réponse. 0 ; 7,32 ; 11,14 ; 13,98 ; 20 — boîte de 7,32 à 13,98.

Correction de l'exercice 15 – Comparer deux villes

On calcule les deux moyennes avec les centres 20 ; 60 ; 100 ; 140 :

$$\overline{x_A} = \frac{80 + 720 + 600 + 280}{24} = 70 \qquad \overline{x_B} = \frac{200 + 240 + 400 + 840}{24} = 70$$

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Les deux villes ont la **même moyenne**, 70 mm. Mais leurs répartitions diffèrent : la moitié des mois de A tombent dans la seule classe $[40\,;80[$, alors que B compte 10 mois très secs et 6 mois très arrosés.

Les médianes le confirment : en supposant les valeurs réparties régulièrement dans les classes, $\text{Méd}_A \approx 40 + 40 \times \dfrac{8}{12} \approx 66{,}7$ mm et $\text{Méd}_B \approx 40 + 40 \times \dfrac{2}{4} = 60$ mm. A a un climat régulier, B un climat contrasté : la moyenne seule ne décrit pas une série.

Réponse. Même moyenne (70 mm), mais A est régulière et B contrastée (profil bimodal).

Correction de l’exercice 16 – Synthèse : salaires d’une entreprise

a) Avec les centres 1 400 ; 1 800 ; 2 200 ; 3 200 ; 7 000 :

$$\overline{x} = \frac{42\,000 + 45\,000 + 33\,000 + 32\,000 + 35\,000}{85} = \frac{187\,000}{85} = 2\,200 \text{ €}$$

b) Les effectifs cumulés valent 30 ; 55 ; 70 ; 80 ; 85. Le rang médian, 42,5, tombe dans $[1\,600\,;2\,000[$; en supposant les 25 salaires répartis régulièrement, on avance de 12,5 sur 25, soit la moitié de la classe : Méd ≈ 1 800 €.

c) La moyenne dépasse la médiane de 400 € : les 5 très hauts salaires représentent $\dfrac{35\,000}{187\,000} \approx 19\,\%$ de la masse salariale pour moins de 6 % des salariés. Plus de 80 % des salariés (70 sur 85) gagnent moins que la moyenne. Le syndicat mettra en avant la médiane, qui décrit le salarié typique ; la direction, la moyenne, plus flatteuse. Le plus honnête est de donner les deux, avec les quartiles.

Réponse. a) 2200 € · b) 1800 € · c) le syndicat la médiane, la direction la moyenne — l’écart vient des très hauts salaires.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s’expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).