Plaquette 3 – Variance, écart-type et mesure de la dispersion

Statistiques descriptives — Seconde

Corrections détaillées

Boîte à outils

Variance et écart-type. Pour une série de N valeurs de moyenne $\overline{x}$:

$$V = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_N - \overline{x})^2}{N} \qquad \sigma = \sqrt{V}$$

Avec un tableau d'effectifs n_i :

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

Formule de la moyenne des carrés (plus rapide à la main) :

$$V = \overline{x^2} - (\overline{x})^2$$

— « moyenne des carrés moins carré de la moyenne ». Elle donne toujours un résultat ≥ 0 ; un résultat négatif signale une erreur de calcul.

Lecture des deux indicateurs.

Indicateur	Ce qu'il mesure	Sensible aux extrêmes ?
étendue	plage totale	très
écart interquartile	étalement de la moitié centrale	non
écart-type	écart moyen à la moyenne	oui

Effet d'une transformation affine. Si $y = ax + b$ pour toutes les valeurs :

$$\overline{y} = a\overline{x} + b \qquad \sigma_y = |a|\,\sigma_x$$

Ajouter une constante **déplace** la série sans la disperser ; multiplier **dilate** la dispersion.

Coefficient de variation. Pour comparer des séries d'unités différentes :

$$CV = \frac{\sigma}{\overline{x}} \quad \text{(souvent en \%)}$$

Cas particulier. $\sigma = 0$ si et seulement si **toutes** les valeurs sont égales.

Correction de l'exercice 1 – Variance par la définition

La variance est la moyenne des carrés des écarts à la moyenne, et l'écart-type sa racine carrée :

$$V = \frac{1}{N}\sum (x_i - \overline{x})^2 \qquad \sigma = \sqrt{V}$$

La moyenne vaut $\frac{40}{5} = 8$. Les écarts à 8 sont -4 ; -2 ; 0 ; 2 ; 4 (leur somme est nulle, comme toujours), et leurs carrés 16 ; 4 ; 0 ; 4 ; 16, de somme 40. Donc $V = \frac{40}{5} = 8$ et $\sigma = \sqrt{8} = 2\sqrt{2} \approx 2{,}83$.

Contrôle de vraisemblance : σ est bien plus petit que l'étendue, $12 - 4 = 8$. Les valeurs s'écartent de la moyenne d'environ 2,8 en moyenne quadratique.

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. $\overline{x} = 8$, $V = 8$, $\sigma = 2\sqrt{2} \approx 2{,}83$.

Correction de l'exercice 2 – Écart-type d'une série irrégulière

La moyenne vaut $\frac{30}{5} = 6$. On calcule les écarts à 6, -4 ; -3 ; 1 ; 2 ; 4, puis leurs carrés, 16 ; 9 ; 1 ; 4 ; 16, de somme 46 :

$$V = \frac{46}{5} = 9{,}2 \qquad \sigma = \sqrt{9{,}2} \approx 3{,}03$$

La série a la même étendue que la précédente ($10 - 2 = 8$), mais un écart-type plus grand (3,03 contre 2,83) : ses valeurs sont regroupées loin de la moyenne, deux en bas et trois en haut. L'étendue, qui ne regarde que les deux extrêmes, ne voit pas cette différence. La médiane, 7, dépasse la moyenne.

Réponse. $\overline{x} = 6$, $V = 9{,}2$, $\sigma \approx 3{,}03$.

Correction de l'exercice 3 – La formule de la moyenne des carrés

La formule donnée demande la moyenne des valeurs, puis la moyenne de leurs carrés. La moyenne vaut $\frac{20}{5} = 4$; les carrés 1 ; 9 ; 16 ; 16 ; 64 ont pour somme 106, donc $\overline{x^2} = \frac{106}{5} = 21{,}2$.

$$V = \overline{x^2} - (\overline{x})^2 = 21{,}2 - 4^2 = 5{,}2 \qquad \sigma = \sqrt{5{,}2} \approx 2{,}28$$

Vérification par la définition : les écarts à 4 sont -3 ; -1 ; 0 ; 0 ; 4, leurs carrés ont pour somme $9 + 1 + 0 + 0 + 16 = 26$, et $\frac{26}{5} = 5{,}2$. Les deux méthodes concordent.

Le piège est de confondre $\overline{x^2} = 21{,}2$, la moyenne des carrés, et $(\overline{x})^2 = 16$, le carré de la moyenne.

Réponse. $V = 21{,}2 - 16 = 5{,}2$ et $\sigma \approx 2{,}28$.

Correction de l'exercice 4 – Écart-type avec un tableau d'effectifs

L'effectif total vaut 20. Chaque valeur compte autant de fois que son effectif : la somme des valeurs est $0 \times 5 + 1 \times 10 + 2 \times 3 + 3 \times 2 = 22$, d'où $\overline{x} = \frac{22}{20} = 1{,}1$ animal par foyer.

Pour la variance, on pondère de même les carrés des écarts à 1,1 :

$$V = \frac{5 \times (-1{,}1)^2 + 10 \times (-0{,}1)^2 + 3 \times 0{,}9^2 + 2 \times 1{,}9^2}{20} = \frac{6{,}05 + 0{,}1 + 2{,}43 + 7{,}22}{20} = 0{,}79$$

Donc $\sigma = \sqrt{0{,}79} \approx 0{,}89$. **Contrôle** : la moyenne des carrés vaut $\frac{0 + 10 + 12 + 18}{20} = 2$, et $2 - 1{,}1^2 = 0{,}79$. Le mode et la médiane valent 1 : un foyer typique a un animal.

Réponse. $\overline{x} = 1{,}1$ et $\sigma = \sqrt{0{,}79} \approx 0{,}89$.

Correction de l'exercice 5 – Même moyenne, dispersions différentes

Les deux séries ont la même moyenne, $\frac{50}{5} = 10$, et la même médiane, 10. On compare leurs dispersions.

— Pour A, les écarts -2 ; -1 ; 0 ; 1 ; 2 donnent des carrés de somme 10 : $V_A = 2$ et $\sigma_A = \sqrt{2} \approx 1{,}41$.
— Pour B, les écarts -5 ; -3 ; 0 ; 3 ; 5 donnent des carrés de somme 68 : $V_B = 13{,}6$ et $\sigma_B \approx 3{,}69$.

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).

$$\frac{\sigma_B}{\sigma_A} \approx \frac{3{,}69}{1{,}41} \approx 2{,}6$$

La série B est environ deux fois et demie plus dispersée que A, ce que confirment les étendues (10 contre 4). Même moyenne et même médiane ne signifient pas séries semblables.

Réponse. Même moyenne 10, mais $\sigma_A = \sqrt{2} \approx 1{,}41$ contre $\sigma_B = \sqrt{13{,}6} \approx 3{,}69$.

Correction de l'exercice 6 – Écart-type nul

On part de la définition de la variance :

$$V = \frac{1}{12}\sum (x_i - \overline{x})^2$$

Si $\sigma = 0$, alors $V = 0$, donc la somme des $(x_i - \overline{x})^2$ est nulle. C'est une somme de carrés, donc de nombres positifs ou nuls ; elle ne peut être nulle que si **chacun** de ses termes l'est. Ainsi $x_i = \overline{x}$ pour tout i : les 12 valeurs sont toutes égales.

Réciproquement, si toutes les valeurs sont égales, tous les écarts à la moyenne sont nuls et $\sigma = 0$. Un écart-type nul caractérise donc les séries constantes, pour lesquelles l'étendue et l'écart interquartile sont nuls aussi.

Réponse. Toutes les valeurs sont égales : $\sigma = 0 \Leftrightarrow$ série constante.

Correction de l'exercice 7 – Ajouter une constante

Ajouter 3 à chaque valeur donne la série 7 ; 9 ; 11 ; 13 ; 15, de moyenne $\frac{55}{5} = 11 = 8 + 3$.

Les écarts à la nouvelle moyenne sont -4 ; -2 ; 0 ; 2 ; 4, exactement les mêmes qu'avant, puisque

$$y_i - \overline{y} = (x_i + 3) - (\overline{x} + 3) = x_i - \overline{x}$$

La variance et l'écart-type sont donc inchangés : $V = 8$ et $\sigma = \sqrt{8} \approx 2{,}83$.

Ajouter une constante déplace toute la série sans changer les écarts entre les valeurs. Ajouter 2 points à tout un contrôle remonte la moyenne de 2 sans réduire les différences entre élèves.

Réponse. $\overline{y} = 11$ et $\sigma_y = \sqrt{8} \approx 2{,}83$ (écart-type **inchangé**).

Correction de l'exercice 8 – Multiplier par une constante

Doubler chaque valeur donne 8 ; 12 ; 16 ; 20 ; 24, de moyenne $16 = 2 \times 8$. Les écarts à la moyenne, -8 ; -4 ; 0 ; 4 ; 8, sont **doublés** ; leurs carrés, de somme 160, sont multipliés par 4 :

$$V = \frac{160}{5} = 32 = 4 \times 8 \qquad \sigma = \sqrt{32} = 4\sqrt{2} \approx 5{,}66 = 2 \times 2\sqrt{2}$$

Multiplier les valeurs par a multiplie l'écart-type par $|a|$ et la variance par a^2. Écrire que la variance double serait faux. L'écart-type s'exprime dans l'unité des données, la variance dans l'unité au carré, ce qui explique cette différence.

Réponse. $\overline{y} = 16$ et $\sigma_y = 4\sqrt{2} \approx 5{,}66$ (écart-type **doublé**).

Correction de l'exercice 9 – Transformation affine : Celsius vers Fahrenheit

La conversion multiplie chaque température par 1,8 puis ajoute 32. La moyenne subit la même transformation :

$$\overline{F} = 1{,}8 \times 20 + 32 = 68\ °\text{F}$$

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).

Pour l'écart-type, seul le facteur 1,8 compte, l'ajout de 32 déplaçant toutes les valeurs sans changer leurs écarts : $\sigma_F = 1{,}8 \times 5 = 9$ °F.

Contrôle sur une valeur : 25 °C, à 5 °C au-dessus de la moyenne, devient $1{,}8 \times 25 + 32 = 77$ °F, à 9 °F au-dessus de 68. Le +32 n'intervient jamais dans un écart.

Réponse. $\overline{F} = 68$ °F et $\sigma_F = 9$ °F.

Correction de l'exercice 10 – Écart-type d'une série en classes

On remplace chaque classe par son centre, 5 ; 15 ; 25 ; 35, et l'on calcule la variance pondérée par les effectifs. Le plus simple est de passer par la moyenne des carrés :

$$\overline{x^2} = \frac{8 \times 25 + 12 \times 225 + 15 \times 625 + 5 \times 1\,225}{40} = \frac{18\,400}{40} = 460$$

Puis $V = 460 - 19{,}25^2 = 460 - 370{,}562\,5 = 89{,}437\,5$ et $\sigma = \sqrt{89{,}437\,5} \approx 9{,}46$ min.

Comme la moyenne, cet écart-type est une approximation liée aux centres de classes. **Contrôle de vraisemblance** : 9,46 min est bien inférieur à l'étendue de 40 min, et l'intervalle $[\overline{x} - \sigma\,;\overline{x} + \sigma] \approx [9{,}8\,;28{,}7]$ couvre le cœur de la série.

Réponse. $V = 460 - 370{,}5625 = 89{,}4375$ et $\sigma \approx 9{,}46$ minutes.

Correction de l'exercice 11 – Intervalle « moyenne plus ou moins un écart-type »

On calcule les bornes de l'intervalle avec $\sigma = 2\sqrt{2} \approx 2{,}83$:

$$[\overline{x} - \sigma\,;\overline{x} + \sigma] \approx [8 - 2{,}83\,;8 + 2{,}83] = [5{,}17\,;10{,}83]$$

Les valeurs 6, 8 et 10 y appartiennent ; 4 et 12 en sont exclues. Cela fait 3 valeurs sur 5, soit 60 %.

L'intervalle à deux écarts-types, $[2{,}34\,;13{,}66]$, contient les 5 valeurs. Sur de grandes séries « en cloche », on trouve souvent environ 68 % des valeurs à moins d'un écart-type de la moyenne et 95 % à moins de deux ; avec 5 valeurs, ce ne sont que des ordres de grandeur.

Réponse. Trois valeurs (6, 8, 10) sur cinq, soit 60 %.

Correction de l'exercice 12 – Sensibilité aux valeurs extrêmes

Pour S_1, la moyenne vaut $\frac{98}{7} = 14$ et l'écart-type $\sqrt{6} \approx 2{,}45$ (la moyenne des carrés vaut $\frac{1\,414}{7} = 202$, et $202 - 14^2 = 6$). La médiane est la 4^{e} valeur, 14 ; Q_1 est la 2^{e} valeur, 12, et Q_3 la 6^{e}, 16.

Pour S_2, seule la plus grande valeur change : la moyenne passe à $\frac{130}{7} \approx 18{,}57$ et l'écart-type à environ 12,96. Mais la médiane et les quartiles occupent les mêmes rangs, avec les mêmes valeurs : 14, 12 et 16.

$$\frac{\sigma_2}{\sigma_1} \approx \frac{12{,}96}{2{,}45} \approx 5{,}3$$

Une seule valeur extrême multiplie l'écart-type par plus de 5 sans toucher médiane ni écart interquartile : ces deux indicateurs sont **robustes**.

Réponse. S_1 : $\overline{x} = 14$, $\sigma \approx 2{,}45$ · S_2 : $\overline{x} \approx 18{,}57$, $\sigma \approx 12{,}96$ — médiane et écart interquartile inchangés.

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 13 – Retrouver une valeur manquante

La somme des 4 valeurs vaut $N \times \overline{x} = 4 \times 12 = 48$. Les trois valeurs connues totalisent $9 + 11 + 15 = 35$, donc la quatrième vaut

$$x_4 = 48 - 35 = 13$$

La série est 9 ; 11 ; 13 ; 15. Les écarts à 12 sont -3 ; -1 ; 1 ; 3, de carrés 9 ; 1 ; 1 ; 9 de somme 20. Donc $V = \dfrac{20}{4} = 5$ et $\sigma = \sqrt{5} \approx 2{,}24$.

Contrôle : la série est régulière (de 2 en 2) et symétrique autour de 12 ; sa médiane, $\dfrac{11 + 13}{2} = 12$, est égale à sa moyenne.

Réponse. $x_4 = 13$ et $\sigma = \sqrt{5} \approx 2{,}24$.

Correction de l’exercice 14 – Coefficient de variation

On ne peut pas comparer directement 5 secondes et 16 grammes : les unités diffèrent, et les valeurs n’ont pas la même taille. On rapporte donc l’écart-type à la moyenne, par le coefficient de variation, qui est sans unité :

$$CV = \frac{\sigma}{\overline{x}} \qquad CV_A = \frac{5}{50} = 10\,\% \qquad CV_B = \frac{16}{200} = 8\,\%$$

La série A est **relativement** la plus dispersée, même si son écart-type est plus de trois fois plus petit en valeur. **Contrôle** : exprimer B en kilogrammes ($\overline{x} = 0{,}2$ et $\sigma = 0{,}016$) donne toujours 8 % : le coefficient de variation ne dépend pas de l’unité choisie.

Réponse. $CV_A = 10\,\%$ contre $CV_B = 8\,\%$: A est relativement plus dispersée.

Correction de l’exercice 15 – Deux façons de remonter une moyenne

Première option, ajouter 1,4 point : la moyenne passe à $9{,}6 + 1{,}4 = 11$, et l’écart-type ne change pas, 2,5, puisque toutes les notes se déplacent de la même quantité.

Seconde option, multiplier par $a = \dfrac{11}{9{,}6} \approx 1{,}146$: la moyenne devient $a \times 9{,}6 = 11$, et l’écart-type est multiplié par a :

$$\sigma = 1{,}146 \times 2{,}5 \approx 2{,}86$$

Les deux options donnent la même moyenne mais pas la même dispersion. Avec la multiplication, le gain est proportionnel à la note : un 5 devient 5,73, un 16 devient 18,33, ce qui creuse les écarts ; avec l’ajout, chacun gagne 1,4 point.

Réponse. Même moyenne 11, mais σ reste 2,5 avec l’ajout et passe à $\approx 2{,}86$ avec la multiplication.

Correction de l’exercice 16 – Synthèse : contrôle de production

a) Les écarts à 20,00 mm ont une somme nulle : le diamètre moyen est exactement 20,00 mm. Leurs carrés, 16 ; 4 ; 0 ; 0 ; 1 ; 4 ; 9 ; 16 ; 1 ; 9, ont pour somme 60, soit une variance de 6 en centièmes de millimètre au carré et un écart-type de $\sqrt{6} \approx 2{,}45$ centièmes :

$$\sigma \approx 2{,}45 \times 0{,}01 \approx 0{,}024\,5 \text{ mm}$$

b) Le plus grand écart vaut 4 centièmes, soit 0,04 mm, inférieur à la tolérance de 0,05 mm : les dix pièces sont conformes, avec une marge de 0,01 mm seulement.

c) La seconde machine est décalée de 0,02 mm, mais son écart-type est plus de deux fois plus petit : l’intervalle $[\overline{x} - 2\sigma\,;\overline{x} + 2\sigma] = [20{,}00\,;20{,}04]$ tient dans la tolérance. On la préfère, car un **décalage se corrige** par un réglage, alors qu’une dispersion trop forte demande de changer de procédé.

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. a) $\overline{x} = 20{,}00$ mm et $\sigma = \sqrt{0{,}0006} \approx 0{,}0245$ mm · b) oui, le pire écart est $0{,}04 < 0{,}05$ mm · c) la machine 2, dont le décalage se règle alors qu'une dispersion ne se règle pas.

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).