

◆ Chapitre 9. — Pourcentages et statistique

I. — Proportions et pourcentages

1) Proportions

Définition 1

Soit A un ensemble non vide ayant n_A éléments et B une partie de A ayant n_B éléments. La proportion des éléments de B dans A est le nombre réel $\frac{n_B}{n_A}$.

Ce nombre peut s'exprimer sous forme d'une fraction, d'un décimal ou d'un pourcentage (ces deux derniers cas n'étant pas toujours possibles).

Exemple 2. Dans une classe de 35 élèves, 12 suivent un cours d'allemand, 23 suivent un cours d'espagnol et 7 suivent un cours de latin.

La proportion d'élèves de la classe suivant un cours d'allemand est $\frac{12}{35}$, la proportion d'élèves de la classe suivant un cours d'espagnol est $\frac{23}{35}$ et la proportion d'élèves suivant un cours de latin est $\frac{7}{35} = \frac{1}{5}$.

Les deux premières proportions ne peuvent pas s'écrire de façon exacte sous forme décimale ou sous forme d'un pourcentage. En revanche, la proportion d'élèves de la classe suivant un cours de latin est égale à 0,2 soit 20%.

Remarque 3. Une proportion est un réel compris entre 0 et 1 car, avec les notations de la définition 1, on a $0 \leq n_B \leq n_A$ et donc, en divisant par $n_A > 0$, $0 \leq \frac{n_B}{n_A} \leq 1$.

Exercice 4. Dans une population, on estime que 5% des individus possèdent un caractère génétique C . Sachant que cette population compte 2500 individus, combien de personnes possèdent le caractère C ?

Solution. — Notons n_C le nombre de personnes ayant le caractère C dans la population. Alors, $\frac{n_C}{2500} = 5\% = \frac{5}{100}$ donc $n_C = \frac{5}{100} \times 2500 = 125$. Il y a donc 125 personnes qui possèdent le caractère C dans cette population.

2) Pourcentage de pourcentage

Propriété 5

On considère un ensemble non vide A . On suppose que B est une partie non vide de A et que C est une partie de B .

Si la proportion des éléments de B dans A est p_1 et la proportion des éléments de C dans B est p_2 alors la proportion des éléments de C dans A est $p_1 \times p_2$.

Démonstration. Notons n_A , n_B et n_C les nombres d'éléments respectifs de A , B et C . Alors, par définition, $p_1 = \frac{n_B}{n_A}$ et $p_2 = \frac{n_C}{n_B}$. On a donc $\frac{n_C}{n_A} = \frac{n_C}{n_B} \times \frac{n_B}{n_A} = p_2 \times p_1$. La proportion des éléments de C dans A est donc bien $p_1 \times p_2$. $\square$

Exemple 6. Sur une année, un musée français a obtenu les chiffres suivants concernant sa fréquentation : 26% des visiteurs du musée ne sont pas français et 32% des visiteurs étrangers ne sont pas des européens.

Ici, A est l'ensemble des visiteurs du musée dans l'année, B est l'ensemble des visiteurs étrangers et C est l'ensemble des visiteurs qui ne sont pas européens.

On en déduit donc que $\frac{26}{100} \times \frac{32}{100} = \frac{53}{625} = 8,32\%$ des visiteurs du musée ne sont pas européens.

Exercice 7. Dans une ville, on distingue deux types de logements : les appartements et les maisons. On sait que 30% des logements sont des appartements chauffés au gaz et 40% des appartements sont chauffés au gaz. Quelle est la proportion de logements qui sont des maisons ?

Solution. Notons p la proportion d'appartements dans la ville. Alors, $p \times \frac{40}{100} = \frac{30}{100}$ donc, en multipliant par 100, $40p = 30$ c'est-à-dire $p = \frac{30}{40} = \frac{3}{4}$. On en déduit que la proportion de logements qui sont des maisons est $1 - \frac{3}{4} = \frac{1}{4}$ soit 25%.

3) Coefficients multiplicateurs, variations successives, variation réciproque

Définition 8

On considère une quantité qui varie d'une valeur initiale V_i à une valeur finale V_f .

1. La variation absolue de cette quantité est $V_f - V_i$.
2. Si $V_i \neq 0$, la variation relative de cette quantité est $\frac{V_f - V_i}{V_i}$.

Exemple 9. Dans une ville, la température le 1er janvier est $5^\circ C$ et la température le 1er février est $1^\circ C$. La variation absolue de la température entre ces deux dates est $1 - 5 = -4^\circ C$ et la variation relative est $\frac{1-5}{5} = -\frac{4}{5} = -80\%$.

Remarque 10. La variation absolue a la même unité que la quantité considérée alors que la variation relative est sans unité.

Propriété 11

On considère une quantité qui varie d'une valeur initiale V_i à une valeur finale V_f . On note t la variation relative associée. Alors, $V_f = (1 + t)V_i$.

Démonstration. Par définition, $t = \frac{V_f - V_i}{V_i}$ donc, en multipliant par V_i , $t \times V_i = V_f - V_i$ puis, en ajoutant V_i aux deux membres, on obtient $V_i + t \times V_i = V_f$. On factorisant le membre de gauche par V_i , on en déduit que $(1 + t)V_i = V_f$ comme annoncé. $\square$

Remarque 12. Ainsi, d'après la propriété 11, si une quantité augmente de $x\%$, sa valeur finale est donc de $(1 + \frac{x}{100})V$ et si elle diminue de $x\%$, sa valeur finale est $(1 - \frac{x}{100})V$.

Exemple 13. Un objet valant 30 euros voit son prix baissé de 40%. Son prix soldé est donc $(1 - \frac{40}{100}) \times 30 = 0,6 \times 30 = 18$ euros.

Définition 14

Avec la notation précédente, le nombre $1 + t$ s'appelle le coefficient multiplicateur associé à la variation relative t .

Exemple 15. Le coefficient multiplicateur associé à une hausse de 5% est $1 + \frac{5}{100} = 1,05$ et le coefficient multiplicateur associé à une baisse de 5% est $1 - \frac{5}{100} = 0,95$.

Exercice 16. Une quantité a été multipliée par 3. Quel est son pourcentage d'augmentation ?

Solution Notons x le pourcentage d'augmentation. Le coefficient multiplicateur associé est donc $1 + \frac{x}{100}$. Or, ce coefficient vaut 3 donc $1 + \frac{x}{100} = 3$ donc $\frac{x}{100} = 2$ et ainsi $x = 200$. Ainsi, une quantité qui est multipliée par 3 subit une augmentation de 200%.

Remarque 17. Pour une quantité qui ne prend que des valeurs positives, le coefficient multiplicateur associé à une hausse est supérieur à 1 et le coefficient multiplicateur associé à une baisse est compris entre 0 et 1.

Propriété 18

Si une quantité subit deux variations relatives successives t_1 puis t_2 alors le coefficient multiplicateur associé à la variation globale est $(1 + t_1)(1 + t_2)$.

Démonstration. Notons V la valeur initiale de la quantité. Après la variation relative t_1 , sa nouvelle valeur est $(1 + t_1)V$. Cette nouvelle valeur subit elle-même une variation relative t_2 donc la valeur finale est $(1 + t_2)[(1 + t_1)V] = (1 + t_1)(1 + t_2)V$. Ainsi, le coefficient multiplicateur de la variation globale est bien $(1 + t_1)(1 + t_2)$. $\square$

Remarque 19. Comme $(1 + t_1)(1 + t_2) = (1 + t_2)(1 + t_1)$, on voit que l'ordre des variations n'a en fait aucune importance.

Exemple 20. Lors des soldes, le prix d'un objet baisse de 30% lors de la première démarque puis de 10% supplémentaires lors de la deuxième démarque. Le coefficient multiplicateur associé à la variation totale est donc $(1 - \frac{30}{100})(1 - \frac{10}{100}) = 0,63$. Or, $0,63 = 1 - 0,37 = 1 - \frac{37}{100}$ donc le prix a subi une baisse totale de 37%

Remarque 21. On voit donc à travers cet exemple que les pourcentages successifs ne s'additionnent pas ! Une baisse de 30% puis de 10% n'est pas équivalente à une baisse de 40% mais de 37% ! Cela vient du fait que la deuxième baisse n'a pas lieu sur le prix initial mais sur un prix moindre. Par exemple, si le prix initial est 100 euros, le prix après la première baisse est 70 euros et la baisse de 10% s'effectue sur ce prix-là : on obtient donc $70 - \frac{10}{100} \times 70 = 63$ euros.

Propriété 22

Si une quantité varie d'une valeur initiale V_i à une valeur finale V_f en subissant une variation relative t , on appelle variation réciproque de t la variation relative qui permet de passer de la valeur V_f à la valeur V_i .

Remarque 23. Ainsi, on peut voir la variation réciproque comme la variation à appliquer à la valeur V_f pour revenir à la valeur initiale V_i .

Propriété 24

Le coefficient multiplicateur associé à la variation réciproque d'une variation t est $\frac{1}{1+t}$.

Démonstration. Supposons que la valeur initiale soit V_i et la valeur finale V_f . Alors, $V_f = (1+t)V_i$ donc $V_i = \frac{1}{1+t}V_f$. Ainsi, le coefficient multiplicateur de la variation t' qui permet de passer de V_f à V_i est bien $\frac{1}{1+t}$. $\square$

Exemple 25. La variation réciproque d'une hausse de 25% correspond à un coefficient multiplicateur de $\frac{1}{1+\frac{25}{100}} = 0,8 = 1 - 0,2 = 1 - \frac{20}{100}$ donc la variation réciproque d'une hausse de 25% est une baisse de 20%.

Exercice 26. Le prix TTC est obtenu en ajoutant la TVA au prix hors-tax. En France, le taux normal de TVA est 20%. Le prix TTC d'un objet est 30 euros. Quel est son prix hors-tax ?

Solution. On passe du prix hors-tax au prix TTC par une augmentation de 20%. Le coefficient multiplicateur associé à la variation réciproque est donc $\frac{1}{1+\frac{20}{100}} = \frac{5}{6}$. Ainsi, le prix hors-tax est $\frac{5}{6} \times 30 = 25$ euros.

II. — Statistique descriptive

Le but de la statistique descriptive est de donner une vision globale d'une série de valeurs à l'aide de quelques paramètres. Il s'agit de résumer tout un ensemble de nombres avec seulement un ou deux nombres. On distingue deux types de mesures :

- les mesures à tendance centrale qui visent à déterminer une valeur « centrale » pour la série ;
- les mesures de dispersion qui déterminent si les valeurs de la série sont plutôt rapprochées ou plutôt écartées les une des autres.

1) Mesures à tendance centrale

Les deux principales mesures à tendance centrale ont déjà été étudiées au collège : il s'agit de la médiane et de la moyenne. Dans ce qui suit, nous rappelons les définitions et nous donnons quelques propriétés de la moyenne.

Dans toute la suite, S désigne une série de N nombres : $x_1, x_2, \dots, x_N$. Le nombre N est appelé l'effectif total de la série.

Par exemple, si on considère la série : $1 ; -5 ; \frac{1}{2} ; 1 ; 1 ; \sqrt{2} ; -3,71$ alors l'effectif total est 7. On remarque qu'un même nombre peut être répété plusieurs fois et on le compte autant de fois qu'il apparaît : ici, 1 apparaît à 3 reprises, on le compte donc 3 fois.

a) Médiane

Définition 27

On définit la médiane Me de la série S de la manière suivante : on commence par ordonner les valeurs de la série dans l'ordre croissant puis

- Si l'effectif total N est impair alors Me est la valeur numéro $\frac{N+1}{2}$ dans la liste ordonnée ;
- si l'effectif total N est pair alors Me est la demi-somme des valeurs numéro $\frac{N}{2}$ et $\frac{N}{2} + 1$ dans la liste ordonnée.

Remarque 28. Intuitivement, la médiane correspond à la valeur située au milieu de la série ordonnée. Cette valeur fait partie de la série initiale si N est impair mais ce n'est pas nécessairement le cas si N est pair.

Exemple 29. On considère les deux séries

$$S_1 : -3 ; 4 ; -8 ; 2 ; 0 ; 4 ; -2 \quad \text{et} \quad S_2 : -3 ; 4 ; -8 ; 2 ; 0 ; 4 ; -2 ; 5.$$

Pour déterminer la médiane, il faut commencer par ordonner la série par ordre croissant.

Pour S_1 , on obtient : $-8 ; -3 ; -2 ; 0 ; 2 ; 4 ; 4$.

Pour S_2 , on obtient : $-8 ; -3 ; -2 ; 0 ; 2 ; 4 ; 4 ; 5$.

L'effectif total de S_1 est $N = 7$ qui est impair donc la médiane de S_1 est le nombre numéro $\frac{7+1}{2} = 4$ dans la liste ordonnée. Ainsi, pour S_1 , la médiane est $Me_1 = 0$.

L'effectif total de S_2 est $N = 8$ qui est pair donc la médiane de S_2 est la demi-somme des nombres numéro $\frac{8}{2} = 4$ et $\frac{8}{2} + 1 = 5$ dans la liste ordonnée. Le nombre numéro 4 est 0 et le nombre numéro 5 est 2 donc, pour S_2 , la médiane est $Me_2 = \frac{0+2}{2} = 1$.

Remarque 30. Pour la série S_2 , la médiane n'est pas une valeur de la série S_2 .

b) Moyenne

Définition 31

La moyenne de S , notée $\bar{x}$, est la somme des valeurs de S divisée par l'effectif total N .

Exemple 32. Reprenons les séries S_1 et S_2 précédentes. Alors, la moyenne de S_1 est

$$\bar{x}_1 = \frac{-3 + 4 + (-8) + 2 + 0 + 4 + (-2)}{7} = -\frac{3}{7}$$

et la moyenne de S_2 est

$$\bar{x}_2 = \frac{-3 + 4 + (-8) + 2 + 0 + 4 + (-2) + 5}{7} = \frac{2}{8} = \frac{1}{4}.$$

Dans la pratique, il arrive souvent qu'on ait beaucoup de valeurs à traiter mais que certaines de ces valeurs se répètent plusieurs fois. Considérons, par exemple, la série S_3 suivante :

2 ; 2 ; 0 ; 1 ; 0 ; 0 ; -1 ; 1 ; 2 ; 1 ; 1 ; 2 ; 0 ; 0

1 ; -1 ; 2 ; 1 ; 0 ; -1 ; -2 ; 0 ; -1 ; 2 ; -1 ; -2 ; -2 ; 0

L'effectif total de cette série est 28 mais il n'y a que 5 nombres différents qui apparaissent : -2, -1, 0, 1 et 2.

Ainsi, il est plus court et plus pratique de présenter cette même série sous la forme du tableau suivant :

valeurs	-2	-1	0	1	2
effectifs	3	5	8	6	6

Ce tableau signifie que dans la série S_3 , le nombre -2 est répété 3 fois, le nombre -1 est répété 5 fois, le nombre 0 est répété 8 fois, le nombre 1 est répété 6 fois et le nombre 2 est répété 6 fois.

L'effectif total est alors $3 + 5 + 8 + 6 + 6 = 28$ comme on l'a dit précédemment.

On peut facilement calculer la somme de ces 28 valeurs grâce au tableau : il s'agit de $(-2) \times 3 + (-1) \times 5 + 0 \times 8 + 1 \times 6 + 2 \times 6 = 7$ donc la moyenne de la série S_3 est $\frac{7}{28} = \frac{1}{4}$.

Le calcul que l'on vient de faire s'appelle une moyenne pondérée. Plutôt que d'additionner les valeurs, on multiplie chaque valeur par son effectif (son « poids ») puis on divise par la somme des effectifs qui est égal à l'effectif total. De manière générale, on a la propriété suivante.

Propriété 33

On considère une série statistique donnée par le tableau suivant :

valeurs	x_1	x_2	$\cdots$	x_k
effectifs	n_1	n_2	$\cdots$	n_k

Alors, la moyenne de cette série est

$$\bar{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + \cdots + n_k \times x_k}{n_1 + n_2 + \cdots + n_k}.$$

Démonstration. Chaque valeur x_i du tableau apparaît n_i fois dans la série donc la somme des valeurs est $n_1 \times x_1 + n_2 \times x_2 + \dots + n_k \times x_k$. De plus, l'effectif total est $n_1 + n_2 + \dots + n_k$ donc la moyenne de la série est bien

$$\bar{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + \dots + n_k \times x_k}{n_1 + n_2 + \dots + n_k}.$$

□

Remarque 34. On utilise également une moyenne pondérée lorsqu'on calcule une moyenne de notes avec coefficients.

Exemple 35. Le tableau suivant donne les notes (sur 20) d'un candidat aux différentes épreuves d'un concours ainsi que les coefficients de chaque épreuve.

épreuves	Math 1	Math 2	Physique 1	Physique 2	Français	Anglais	informatique
notes	13,6	12,8	10,2	11	13	14,5	14,7
effectifs	4	5	3	4	5	3	2

La moyenne de ce candidat est donc

$$\frac{13,6 \times 4 + 12,8 \times 5 + 10,2 \times 3 + 11 \times 4 + 13 \times 5 + 14,5 \times 3 + 14,7 \times 2}{4 + 5 + 3 + 4 + 5 + 3 + 2} = \frac{330,9}{26} \approx 12,73.$$

Propriété 36. — Linéarité de la moyenne

On considère une fonction affine f et une série statistique S . On note $\bar{x}$ la moyenne de S . On note S' la série obtenue à partir de S en appliquant à chaque valeur de S la fonction f . Ainsi, si S est la série $x_1, x_2, \dots, x_N$ alors S' est la série $f(x_1), f(x_2), \dots, f(x_N)$.

Alors, la moyenne de S' est égale à $f(\bar{x})$.

Démonstration. Par définition, il existe deux réels m et p tels que $f : x \mapsto mx + p$. Notons $x_1, x_2, \dots, x_N$ les valeurs de S . Alors, les valeurs de S' sont $mx_1 + p, mx_2 + p, \dots, mx_N + p$. Ainsi, la somme des valeurs de S' est

$$\begin{aligned} (mx_1 + p) + (mx_2 + p) + \dots + (mx_N + p) &= mx_1 + mx_2 + \dots + mx_N + \underbrace{p + p + \dots + p}_{N \text{ fois}} \\ &= m(x_1 + x_2 + \dots + x_N) + N \times p \end{aligned}$$

Dès lors, la moyenne de S' est

$$\begin{aligned} \frac{m(x_1 + x_2 + \dots + x_N) + N \times p}{N} &= \frac{m(x_1 + x_2 + \dots + x_N)}{N} + \frac{N \times p}{N} \\ &= m \times \frac{x_1 + x_2 + \dots + x_N}{N} + p \end{aligned}$$

Mais, par définition, $\frac{x_1 + x_2 + \dots + x_N}{N} = \bar{x}$ donc la moyenne de S' est $m\bar{x} + p$ c'est-à-dire $f(\bar{x})$. □

Exemple 37. La moyenne à un devoir est 8,2. Le professeur trouvant cette moyenne trop faible, il décide d'augmenter toutes les notes d'un point. Autrement dit, il applique à chaque note la fonction affine $x \mapsto x + 1$. Par linéarité de la moyenne, la nouvelle moyenne sera $8,2 + 1 = 9,2$.

2) Mesures de dispersion

Les mesures à tendance centrale donnent une idée du « centre » de la série. Cependant, si on considère les deux séries de notes suivantes : 3, 10, 20 et 10, 10, 13 alors on obtient la même moyenne (11 pour les deux) et la même médiane (10 pour les deux). Pour autant, la répartition des notes est très différentes dans les deux séries. Pour rendre compte de cette répartition, on utilise une mesure de dispersion.

a) Étendue et écart interquartile

Une première mesure de dispersion, déjà vue au collège, est l'étendue.

Définition 38

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

Exemple 39. Si on considère la série S_3 : 1 ; 2 ; 3 ; 2 ; 1 ; 3 ; 2 ; 3 ; 100 alors l'étendue est $100 - 1 = 99$.

On voit à travers l'exemple précédent que l'étendue présente l'inconvénient de dépendre seulement des valeurs extrêmes (le minimum et le maximum de la série) et une seule valeur (comme 100 dans l'exemple précédent) peut avoir un énorme effet sur l'étendue. Remarquons que si on avait eu 1000 valeurs toutes égales à 1 et une valeur égale à 100, l'étendue aurait aussi été égale à 99. Cette mesure n'est donc pas très pertinente.

On lui préférera une autre mesure qui tient compte du nombre de valeurs de la série, à savoir l'écart interquartile.

Définition 40

On considère une série S . On définit :

1. le premier quartile de S , noté Q_1 , comme étant la plus petite des valeurs de S telle qu'au moins $\frac{1}{4}$ des valeurs de S soient inférieures à Q_1 ;
2. le troisième quartile de S , noté Q_3 , comme étant la plus petite des valeurs de S telle qu'au moins $\frac{3}{4}$ des valeurs de S soient inférieures à Q_3 ;
3. l'écart interquartile e par $e = Q_3 - Q_1$.

Remarque 41. Grossièrement, Q_1 partage la série en $\frac{1}{4}/\frac{3}{4}$ et Q_3 partage la série en $\frac{3}{4}/\frac{1}{4}$. On remarquera qu'on n'a pas défini le deuxième quartile car il correspond en fait (plus ou moins) à la médiane.

Méthode 42

Pour calculer les quartiles d'une série S dont l'effectif total est N ,

- on commence par ordonner la série dans l'ordre croissant ;
- on calcule $\frac{1}{4} \times N$ et on note p le plus petit entier supérieur ou égal à ce nombre : Q_1 est alors le nombre numéro p dans la série ordonnée ;
- on calcule $\frac{3}{4} \times N$ et on note q le plus petit entier supérieur ou égal à ce nombre : Q_3 est alors le nombre numéro q dans la série ordonnée.

Exemple 43. Reprenons la série $S_3 : 1 ; 2 ; 3 ; 2 ; 1 ; 3 ; 2 ; 3 ; 100$ vue précédemment.

Commençons par ordonner la série : $1 ; 1 ; 2 ; 2 ; 2 ; 3 ; 3 ; 3 ; 100$.

Ici, $N = 9$ donc $\frac{1}{4} \times 9 = 2,25$ et ainsi $p = 3$ (le premier entier juste après 2,25). Ainsi, Q_1 est le nombre numéro 3 dans la liste ordonnée donc $Q_1 = 2$.

De même, $\frac{3}{4} \times 9 = 6,75$ donc $q = 7$ (le premier entier juste après 6,75). Ainsi, Q_3 est le nombre numéro 7 dans la liste ordonnée donc $Q_3 = 3$.

On conclut que l'écart interquartile est $e = 3 - 2 = 1$.

Remarquons que cet écart est bien plus faible que l'étendue car il n'est pas « faussé » par la valeur extrême 100. D'ailleurs, on pourrait bien remplacer 100 par 1000 ou par 1000000, cela ne changerait pas la valeur de e .

Remarque 44. Dans la méthode précédente, si les calculs $\frac{1}{4} \times N$ et $\frac{3}{4} \times N$ donnent directement des entiers alors on conserve ces valeurs pour déterminer Q_1 et Q_3 . Par exemple, si $N = 12$ alors $\frac{1}{4} \times 12 = 3$ et $\frac{3}{4} \times 12 = 9$ donc Q_1 sera la valeur numéro 3 dans la liste ordonnée et Q_3 sera la valeur numéro 9 dans la série ordonnée.

b) Variance et écart-type

On voit que les notions de quartiles et donc d'écart interquartile sont naturellement liés à celle de médiane.

De la même façon, il existe une mesure de dispersion liée à la moyenne : il s'agit de l'écart-type.

Définition 45

Considérons une série $S : x_1, x_2, \dots, x_N$. On définit :

1. la variance V de S comme la moyenne des carrés des écarts à la moyenne c'est-à-dire

$$V = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N}$$

2. l'écart-type σ de S par $\sigma = \sqrt{V}$.

Avant de donner un exemple, faisons plusieurs remarques.

Remarque 46.

1. Tout d'abord, pourquoi V mesure-t-elle la dispersion de la série ?

On voit que plus les valeurs de la série sont dispersées plus les écarts $x_i - \bar{x}$ vont avoir tendance à être grands (en valeurs absolues). Reprenons les deux séries de notes $3 ; 10 ; 20$ et $10 ; 10 ; 13$ évoquées précédemment. Pour chacune de ces deux séries, la moyenne est 11. Pour la première, les écarts à la moyenne sont $3 - 11 = -8$, $10 - 11 = -1$ et $20 - 11 = 9$ alors que, pour la seconde, ces écarts sont $10 - 11 = -1$, $10 - 11 = -1$ et $13 - 11 = 2$. On voit que dans le premier cas, les écarts sont (en valeurs absolues) plus grands que dans le second.

Ensuite, pour éviter que des écarts positifs se compensent avec des écarts négatifs, on élève au carré.

2. La lettre σ est une lettre grecque : la sigma (minuscule).
3. Le fait d'élever au carré change l'unité des valeurs considérées. Par exemple, si on considère une série de mesures en cm alors l'unité de V sera le cm^2 . Ainsi, pour revenir à une mesure homogène aux valeurs de la série, on prend la racine carrée, ce qui explique la définition de σ .
4. Si la série est donnée par un tableau de la forme

valeurs	x_1	x_2	$\cdots$	x_k
effectifs	n_1	n_2	$\cdots$	n_k

alors la variance s'écrit

$$V = \frac{n_1 \times (x_1 - \bar{x})^2 + n_2 \times (x_2 - \bar{x})^2 + \cdots + n_k \times (x_k - \bar{x})^2}{n_1 + n_2 + \cdots + n_k}$$

Exemple 47. Reprenons la série $S_3 : 1 ; 2 ; 3 ; 2 ; 1 ; 3 ; 2 ; 3 ; 100$ de l'exemple 39. On peut présenter cette série par le tableau suivant :

valeurs	1	2	3	100
effectifs	2	3	3	1

Avant de calculer la variance et l'écart-type, il faut d'abord calculer la moyenne de la série

$$\bar{x} = \frac{2 \times 1 + 3 \times 2 + 3 \times 3 + 1 \times 100}{2 + 3 + 3 + 1} = 13.$$

La variance est donc

$$V = \frac{2 \times (1 - 13)^2 + 3 \times (2 - 13)^2 + 3 \times (3 - 13)^2 + 1 \times (100 - 13)^2}{9} = \frac{2840}{3}$$

et ainsi l'écart-type est $\sigma = \sqrt{\frac{2840}{3}} \approx 30,8$.

Propriété 48. — Formule de König-Huygens

On considère une série statistique $S : x_1, x_2, \dots, x_N$. Alors, la variance est égale à la moyenne des carrés moins le carré de la moyenne c'est-à-dire :

$$V = \frac{x_1^2 + x_2^2 + \cdots + x_N^2}{N} - \bar{x}^2$$

Démonstration. En utilisant les identités remarquables,

$$\begin{aligned} (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_N - \bar{x})^2 &= (x_1^2 - 2x_1\bar{x} + \bar{x}^2) + (x_2^2 - 2x_2\bar{x} + \bar{x}^2) + \cdots + (x_N^2 - 2x_N\bar{x} + \bar{x}^2) \\ &= x_1^2 + x_2^2 + \cdots + x_N^2 - 2(x_1\bar{x} + x_2\bar{x} + \cdots + x_N\bar{x}) + \underbrace{\bar{x}^2 + \bar{x}^2 + \cdots + \bar{x}^2}_{N \text{ fois}} \\ &= x_1^2 + x_2^2 + \cdots + x_N^2 - 2(x_1 + x_2 + \cdots + x_N)\bar{x} + N \times \bar{x}^2 \end{aligned}$$

Or, par définition, $\bar{x} = \frac{x_1 + x_2 + \cdots + x_N}{N}$ donc $N \times \bar{x} = x_1 + x_2 + \cdots + x_N$. Dès lors,

$$\begin{aligned} (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_N - \bar{x})^2 &= x_1^2 + x_2^2 + \cdots + x_N^2 - 2(N \times \bar{x})\bar{x} + N \times \bar{x}^2 \\ &= x_1^2 + x_2^2 + \cdots + x_N^2 - 2N \times \bar{x}^2 + N \times \bar{x}^2 \\ &= x_1^2 + x_2^2 + \cdots + x_N^2 - N \times \bar{x}^2 \end{aligned}$$

En divisant par N , on conclut que

$$V = \frac{x_1^2 + x_2^2 + \cdots + x_N^2 - N \times \bar{x}^2}{N} = \frac{x_1^2 + x_2^2 + \cdots + x_N^2}{N} - \frac{N \times \bar{x}^2}{N} = \frac{x_1^2 + x_2^2 + \cdots + x_N^2}{N} - \bar{x}^2.$$

□

Remarque 49. L'intérêt de la formule de König-Huygens est de n'utiliser la moyenne $\bar{x}$ qu'une seule fois. Ainsi, si on a seulement une valeur approchée de $\bar{x}$, on évitera d'accumuler les approximations en utilisant cette formule.

Exemple 50. Si on reprend le calcul de l'exemple 47, à l'aide de la formule de König-Huygens, on trouve

$$V = \frac{2 \times 1^2 + 3 \times 2^2 + 3 \times 3^2 + 1 \times 100^2}{9} - 13^2 = \frac{2840}{3}$$

ce qui est bien cohérent.

Remarque 51. L'écart-type seul (comme l'écart interquartile seul) n'a pas grande signification. Il devient intéressant pour comparer la dispersion d'une série par rapport à une autre. Plus cet écart est grand, plus la série est dispersée.

Exemple 52. Comparons la dispersion des deux séries suivantes :

$$S : -5 ; 3 ; 6 ; -1 ; 5 \quad \text{et} \quad S' : 20 ; 23 ; 18 ; 17 ; 27 ; 15 ; 19 ; 20$$

Utilisons d'abord l'écart interquartile. Pour cela, on commence par ordonner les deux séries par ordre croissant :

$$S : -5 ; -1 ; 3 ; 5 ; 6 \quad \text{et} \quad S' : 15 ; 17 ; 18 ; 19 ; 20 ; 20 ; 23 ; 27.$$

L'effectif total de S est 5. Or, $\frac{1}{4} \times 5 = 1,25$ donc Q_1 est la deuxième valeur et $\frac{3}{4} \times 5 = 3,75$ donc Q_3 est la quatrième valeur. Ainsi, $Q_1 = -1$ et $Q_3 = 5$ donc $e = 5 - (-1) = 6$.

L'effectif total de S' est 8. Or, $\frac{1}{4} \times 8 = 2$ donc Q'_1 est la deuxième valeur et $\frac{3}{4} \times 8 = 6$ donc Q'_3 est la sixième valeur. Ainsi, $Q'_1 = 17$ et $Q'_3 = 20$ donc $e' = 20 - 17 = 3$.

Comme $e' < e$, on conclut que S' est moins dispersée que S .

Utilisons ensuite l'écart interquartile. Pour cela, on calcule d'abord la moyenne de chaque série. Pour S , on trouve

$$\bar{x} = \frac{-5 + 3 + 6 + (-1) + 5}{5} = \frac{8}{5}$$

et pour S' , on trouve

$$\bar{x}' = \frac{20 + 23 + 18 + 17 + 27 + 15 + 19 + 20}{8} = \frac{159}{8}.$$

Ensuite, on calcule la variance. Pour S , on trouve

$$V = \frac{(-5)^2 + 3^2 + 6^2 + (-1)^2 + 5^2}{5} - \left(\frac{8}{5}\right)^2 = \frac{416}{25}$$

et pour S' , on trouve

$$V' = \frac{20^2 + 23^2 + 18^2 + 17^2 + 27^2 + 15^2 + 19^2 + 20^2}{8} - \left(\frac{159}{8}\right)^2 = \frac{775}{64}$$

Ainsi, l'écart-type de S est $\sigma = \sqrt{\frac{416}{25}} \approx 4,1$ et celui de S' est $\sigma' = \sqrt{\frac{775}{64}} \approx 3,5$. Comme $\sigma' < \sigma$, on conclut de même que S' est moins dispersée que S .

Remarque 53. Pour finir, remarquons que tous les calculs précédents deviennent vite (très) fastidieux quand il y a beaucoup de valeurs à traiter. Dans la pratique, on utilise des calculatrices ou des ordinateurs pour faire le travail à notre place. Vous trouverez à la toute dernière page de votre manuel les détails pour faire faire le travail à votre calculatrice !