

Série statistique à une variable

1 Série statistique

1.1 Vocabulaire

Définition 1 Une série statistique est la donnée des différentes valeurs prises par un certain indicateur numérique (qu'on appelle un caractère quantitatif).

Il existe deux grandes familles d'indicateurs :

- les caractères discrets, qui ne peuvent prendre qu'un nombre fini de valeurs ;
- les caractères continus qui peuvent prendre un infinité de valeurs.

Exemple 1 Les notes obtenues par un élève sont des caractères discrets, car les notes d'un devoir ne peuvent prendre que des valeurs finies (au plus 21 valeurs si on note au point près, et au plus 41 valeurs si l'on note au demi-point près).

La taille d'un individu peut prendre, elle, une infinité de valeurs (le problème est alors de fabriquer des instruments de mesure capables de fournir une mesure absolument exacte ... et donc de rendre réellement compte de la possibilité d'obtenir toutes les valeurs (exactes) de cette taille).

1.2 Présentations d'une série statistique sous forme de tableaux

Il y a deux grandes présentations possibles pour les séries statistiques, suivant leur nature discrète ou continue. Dans le cas d'une série discrète, la présentation habituelle consiste à faire la liste de toutes les valeurs prises par le caractère considéré, y compris en répétant certaines valeurs autant de fois qu'elles sont rencontrées. Dans le cas d'un caractère continu, il n'est pas possible de dresser la liste exhaustive des valeurs prises par ce caractère. On opère alors des regroupements, afin de se ramener à une présentation commode : on opère un regroupement en classes.

Dans le cas discret, on regroupe en général les valeurs qui sont obtenues plusieurs fois dans l'étude du caractère, ce qui conduit à une représentation sous la forme d'un tableau de valeurs en fonction des effectifs correspondants. Dans le cas continu, on dresse un tableau associant à chaque classe son effectif (c'est-à-dire le nombre de valeurs relevées du caractère qui sont comprises dans les classes considérées).

Enfin, à partir d'une présentation de la série selon les effectifs, on peut obtenir une nouvelle présentation de la série selon les fréquences. Parfois, la série est directement donnée selon les fréquences d'apparition des valeurs du caractère observé. On passe d'une présentation à l'autre en se rappelant que la fréquence d'apparition d'une valeur ou d'une classe est donnée par la relation

$$\text{fréquence} = \frac{\text{effectif}}{\text{effectif total}}$$

1.2.1 Présentation en valeurs selon les effectifs

La série se présente sous la forme $X :$

x_1	x_2	$\dots$	x_p
n_1	n_2	$\dots$	n_p

 où les x_i sont les valeurs du caractère observé, et chaque n_i l'effectif de la valeur x_i du caractère.

Exemple 2

Exemple 3

1.2.2 Présentation en valeurs selon les fréquences

La série se présente sous la forme $X : \begin{array}{|c|c|c|c|} \hline x_1 & x_2 & \dots & x_p \\ \hline f_1 & f_2 & \dots & f_p \\ \hline \end{array}$ où les x_i sont les valeurs du caractère observé, et chaque f_i la fréquence d'apparition de la valeur x_i du caractère. Il est essentiel de se rappeler (et de vérifier ...) que les fréquences f_i vérifient :

$$\left\{ \begin{array}{l} \text{Pour tout } i, 0 \leq f_i \leq 1 \\ \text{et} \\ f_1 + f_2 + \dots + f_p = 1 \end{array} \right.$$

Exemple 4

Exemple 5

1.2.3 Présentation en classes selon les effectifs

La série se présente sous la forme $X : \begin{array}{|c|c|c|c|} \hline [a_1; b_1[& [a_2; b_2[& \dots & [a_p; b_p[\\ \hline n_1 & n_2 & \dots & n_p \\ \hline \end{array}$ où les $[a_i; b_i[$ sont les classes de valeurs du caractère observé, et chaque n_i l'effectif de la classe $[a_i; b_i[$ du caractère.

Exemple 6

Exemple 7

1.2.4 Présentation en classes selon les fréquences

La série se présente sous la forme $X : \begin{array}{|c|c|c|c|} \hline [a_1; b_1[& [a_2; b_2[& \dots & [a_p; b_p[\\ \hline f_1 & f_2 & \dots & f_p \\ \hline \end{array}$ où les $[a_i; b_i[$ sont les classes de valeurs du caractère observé, et chaque n_i la fréquence d'apparition de la classe $[a_i; b_i[$ du caractère. Il est essentiel de se rappeler (et de vérifier ...) que les fréquences f_i vérifient ici aussi :

$$\left\{ \begin{array}{l} \text{Pour tout } i, 0 \leq f_i \leq 1 \\ \text{et} \\ f_1 + f_2 + \dots + f_p = 1 \end{array} \right.$$

Exemple 8

Exemple 9

1.3 Présentations d'une série statistique sous forme de graphiques

La variété des représentations graphiques de donnée statistiques est grande, mais on peut néanmoins difficilement éviter de parler de deux représentations graphiques qui se généraliseront lorsque que l'on sera amené à étudier les variables aléatoires. Ces deux grands types de présentation sont en général mélangées, et il convient d'être prudent lorsqu'on les rencontre, les conventions les régissant différant sensiblement.

1.3.1 Diagramme en bâtons des effectifs

Cette présentation n'est adaptée que pour les séries statistiques résultant d'un caractère discret. Le principe est de porter en abscisses les valeurs observées du caractère, et en ordonnées les effectifs correspondants. On trace alors un segment perpendiculaire au premier axe (un "bâton"), joignant le point ainsi déterminé. Il est à noter que les bâtons ainsi formés n'ont, en principe, pas d'épaisseur, même si des considérations pratiques font qu'ils en présentent souvent une.

1.3.2 Diagramme en bâtons des fréquences

Cette présentation n'est adaptée que pour les séries statistiques résultant d'un caractère discret. Le principe est le même que pour les diagrammes en bâtons des effectifs, et est obtenu par une construction en tout point similaire, en substituant aux effectifs les fréquences.

1.3.3 Histogramme des effectifs

Cette présentation ne convient en principe que pour les séries présentées en classes, mais on les retrouve aussi parfois pour les séries résultant d'un caractère continu.

Le principe est de faire figurer sur l'axe des abscisses les différentes classes, en portant leurs extrémités. Au-dessus de chaque classe, on trace un rectangle dont l'aire est proportionnelle à l'effectif de la classe considérée.

Remarque 1 *Ce point est source fréquente de confusions : il n'est pas question, dans ce type de représentation, de porter en ordonnée un rectangle de longueur proportionnelle à l'effectif, du moins en général.*

Cependant, lorsque toutes les classes sont de la même amplitude, choisir une aire des rectangles de l'histogramme proportionnelle à l'effectif revient évidemment à choisir une longueur proportionnelle à l'effectif.

1.3.4 Histogramme des fréquences

Cette présentation ne convient en principe que pour les séries présentées en classes, mais on les retrouve aussi parfois pour les séries résultant d'un caractère continu.

Le principe est le même que pour les diagrammes en bâtons des effectifs, et est obtenu par une construction en tout point similaire, en substituant aux effectifs les fréquences.

2 Tendances centrale d'une série statistique

Lorsque les séries statistiques étudiées deviennent longues, il n'est pas question de porter à la connaissance de tous les destinataires l'intégralité des valeurs prises par le caractère. Dans ce cas, le statisticien est confronté au problème du résumé de la série : il s'agit de rendre compte, le plus fidèlement possible, de la série, tout en la raccourcissant. Ce problème, est en général, extrêmement difficile, car le résumé est par nature réducteur. Nous n'aborderons pas ici les moyens permettant de condenser l'information, sans trop la déformer, et n'aborderons que les critères usuels de résumé de la tendance centrale (dominante) d'une série statistique.

2.1 Moyenne d'une série statistique

2.1.1 Cas d'une série discrète présentée selon les effectifs

Rappelons la formule générale qui permet de calculer la moyenne d'une telle série statistique (cette série est dite coefficientée - que ce soit de manière totalement arbitraire, par les effectifs réellement observés, ou encore par les fréquences comme on verra ci-dessous).

Théorème 1 *Si X est une série statistique, donnée selon ses effectifs X :*

x_1	x_2	$\dots$	x_p
n_1	n_2	$\dots$	n_p

alors la moyenne de la série X , notée $\bar{X}$, vaut :

$$\bar{X} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{n_1 + n_2 + \dots + n_p}.$$

Remarque 2 *Cette formule s'applique bien sûr au cas d'une série statistique sans coefficients : on est ramené à une série pour laquelle tous les coefficients valent 1.*

Exercice 1 *Voici les notes obtenues par un élève au cours du trimestre :*

Notes obtenues	11	5	18	7	9
----------------	----	---	----	---	---

Calculer la moyenne trimestrielle de cet élève.

Exercice 2 *Voici les notes obtenues par un élève au cours du trimestre, ainsi que les coefficients qui leur sont attribués pour le calcul de la moyenne du trimestre :*

Notes obtenues	11	5	18	7	9
Coefficients	1	4	8	3	2

Calculer la moyenne trimestrielle de cet élève.

2.1.2 Cas d'une série discrète présentée selon les fréquences

La formule précédente s'applique aussi au cas d'une série statistique présentée selon les fréquences d'apparition, avec une adaptation due au fait essentiel que la somme des fréquences vaut 1 :

Théorème 2 Si X est une série statistique, donnée selon ses fréquences X :

x_1	x_2	$\dots$	x_p
f_1	f_2	$\dots$	f_p

alors la moyenne de la série X , notée $\bar{X}$, vaut :

$$\bar{X} = f_1x_1 + f_2x_2 + \dots + f_px_p.$$

Exercice 3

Exercice 4

2.1.3 Cas d'une série continue présentée selon les effectifs

Théorème 3 Si X est une série statistique, donnée selon ses effectifs X :

$[a_1; b_1[$	$[a_2; b_2[$	$\dots$	$[a_p; b_p[$
n_1	n_2	$\dots$	n_p

et si on note c_i le centre de chaque classe (donc $c_i = \frac{a_i + b_i}{2}$), alors la moyenne de la série X , notée $\bar{X}$, vaut :

$$\bar{X} = \frac{n_1c_1 + n_2c_2 + \dots + n_pc_p}{n_1 + n_2 + \dots + n_p}.$$

Exercice 5

Exercice 6

2.1.4 Cas d'une série continue présentée selon les fréquences

Théorème 4 Si X est une série statistique, donnée selon ses fréquences X :

$[a_1; b_1[$	$[a_2; b_2[$	$\dots$	$[a_p; b_p[$
f_1	f_2	$\dots$	f_p

et si on note c_i le centre de chaque classe (donc $c_i = \frac{a_i + b_i}{2}$), alors la moyenne de la série X , notée $\bar{X}$, vaut :

$$\bar{X} = f_1c_1 + f_2c_2 + \dots + f_pc_p.$$

Exercice 7

Exercice 8

2.2 Médiane d'une série statistique

On se donne une série statistique X :

Valeurs observées	x_1	x_2	x_3	$\dots$	x_n
Effectifs	k_1	k_2	k_3	$\dots$	k_n

et on suppose que l'on a classé les x_i dans un certain ordre (croissant ou décroissant).

Définition 2 Une médiane de cette série, notée $med(X)$, est une valeur x_i qui partage l'effectif total en deux classes de même effectif.

Une médiane d'une série statistique partage donc cette série en deux groupes de même effectif :

- les valeurs inférieures ou égales à la valeur médiane ;
- les valeurs supérieures ou égales à la valeur médiane.

Exemple 10 Un professeur a classé par ordre croissant les notes des 13 garçons et des 14 filles d'une classe.

Garçons	7	8	9	9	10	10	11	12	13	14	14	15	17
---------	---	---	---	---	----	----	----	----	----	----	----	----	----

Filles	7	7	9	9	10	11	12	13	13	13	14	14	15	15
--------	---	---	---	---	----	----	----	----	----	----	----	----	----	----

La valeur médiane des notes des garçons est de 11, celle des filles est de 13.

Ceci signifie que la moitié des garçons a obtenu 11 ou plus, et que la moitié des filles a obtenu 13 ou plus. Ceci signifie aussi que la moitié des garçons a moins de 11, et que la moitié des filles a moins de 13.

Exercice 9

Remarque 3 Une médiane $\text{med}(X)$ peut être la valeur x_i correspondant au dernier indice i pour lequel

$$k_1 + k_2 + \dots + k_{i-1} \leq \frac{1}{2}(k_1 + k_2 + \dots + k_n).$$

De la même manière, si $f_1, f_2, \dots, f_n$ sont les fréquences de cette série, une médiane $\text{med}(X)$ de cette série peut aussi être la valeur x_i correspondant au dernier indice i pour lequel

$$f_1 + f_2 + \dots + f_{i-1} \leq \frac{1}{2}.$$

Preuve

◀ On a le résultat immédiatement en regardant le tableau :

Valeurs observées	x_1	x_2	x_3	...	x_n
Effectifs	k_1	k_2	k_3	...	k_n
Fréquences	f_1	f_2	f_3	...	f_n

▶

Pour déterminer concrètement une médiane d'une série, on classe les données de la série par ordre croissant, puis :

* si la série est donnée par distribution d'effectifs, on calcule les effectifs cumulés, et on calcule le demi-effectif total $\left[\frac{N+1}{2}\right]$. Une médiane est alors obtenue en relevant la première valeur de la série pour laquelle l'effectif cumulé jusqu'à cette valeur est supérieur ou égal au demi-effectif.

* si la série est donnée par distribution de fréquence, on calcule les fréquences cumulées. Une médiane est obtenue en relevant la première valeur de la série pour laquelle la fréquence cumulée jusqu'à cette valeur est supérieure ou égale à $\frac{1}{2}$.

Exemple 11 A la question " Depuis combien d'années résidez-vous dans la même ville " les cinquante personnes interrogées ont donné les réponses suivantes :

Nombre d'années	1	2	3	4	5	6	Plus de 6
Effectif	2	4	5	10	6	12	11
Effectif cumulé	2	6	11	21	27	39	50

Une médiane de cette série est 5. Ceci signifie que la moitié des personnes interrogées sont restées 5 ou plus dans la même ville.

Exercice 10

Exercice 11

Remarque 4 La méthode la plus visuelle et la plus précise pour obtenir une médiane consiste à tracer, selon les cas, le polygone des effectifs cumulés, ou le polygone des fréquences cumulées, puis à faire une interpolation linéaire. Voir, pour cette méthode, les exercices abordés en TP.

Remarque 5 On prendra bien soin de distinguer moyenne et médiane, qui peuvent être fort différentes.

De la même manière que l'on a défini une notion de médiane, en reprenant la médiane de chacune des deux sous-séries obtenues après le calcul de la médiane, on définit les notions de quartiles. Ce procédé se généralise facilement, et le plus souvent on choisit comme découpage suivant de la série considérée celui qui est issu de la notion de décile ou centile. Ces découpages constituent des informations sur la répartition des valeurs de la série statistique, et on les rencontrera lors de l'étude des

2.3 Mode et classe modale

Définition 3 Le mode d'une série statistique discrète est une valeur de la série d'effectif maximal (ou de fréquence maximale).

Définition 4 La classe modale d'une série statistique est une classe de la série d'effectif maximal (ou de fréquence maximale).

Ces deux instruments sont très rudimentaires, et sont donc rarement utilisés en pratique.

2.4 Propriétés communes de la moyenne, de la médiane et du mode

Proposition 5 *La recherche de la moyenne, de la médiane, ou du mode d'une série statistique sont des opérations stables par les transformations affines.*

Ceci se traduit en disant que si l'on opère une transformation affine de toutes les valeurs de la série, la nouvelle moyenne, médiane, ou le nouveau mode, sont obtenus à partir des anciens par la même transformation affine.

3 Mesures de dispersion d'une série statistique

3.1 Insuffisance des mesures de tendance centrale

L'expérience montre très rapidement que la connaissance d'un ou même plusieurs résumés de tendance centrale d'une série statistique est largement insuffisante, et souffre en règle générale d'une perte d'information considérable par rapport à la série de départ. Cette insuffisance est commune à tous les résumés de tendance centrale, qui, par nature, ne donnent aucune information sur la répartition des valeurs de la série. Quelques exemples simples vont illustrer clairement ce propos.

3.1.1 Insuffisance de la médiane

Les principales raisons qui conduisant à rejeter les notions de médiane, tiennent à des difficultés de nature algorithmique. Il est en effet fastidieux de classer correctement les données, puis d'obtenir des valeurs approchées de ces grandeurs, même si une calculatrice permet de s'en sortir sans trop de mal pour des séries de taille "raisonnable". L'utilisation des ordinateurs corrige en partie seulement cette difficulté majeure.

3.1.2 Insuffisance de la moyenne

Nous commençons par un exemple simple. On trouvera plus bas un exemple qui prouve l'insuffisance simultanée des indicateurs de tendance centrale : moyenne, médiane et mode.

Activité 1 *Voici les résultats de deux élèves au cours d'un trimestre de classe :*

Notes de A	10	10	10	10	10
Notes de B	11	5	18	7	9

1°) *Calculer les moyennes des deux élèves. Que constate-t-on ?*

1°) *Peut-on dire que les résultats de ces deux élèves sont comparables sur ce trimestre ? Pourquoi ?*

Comme on le voit sur cet exemple, les résultats d'un calcul de moyenne sont imparfaits pour rendre compte de la réalité de la série statistique considérée; en particulier, le calcul de la moyenne ne rend pas compte des variations par rapport à la moyenne calculée finalement. Pour remédier à ce problème, il va nous falloir un autre outil. Ce sera, entre autres, la variance et/ou l'écart-type qui sont introduits par la suite, avec d'autres mesures de dispersion, moins perfectionnées.

3.1.3 Insuffisance de l'ensemble des mesures de tendance centrale

Ci-dessous, on trouvera un exemple qui prouve l'insuffisance simultanée des trois types d'indicateurs de tendance centrale : moyenne, médiane et mode.

Exemple 12 *Voici les résultats de deux élèves au cours d'un trimestre de classe :*

Notes de A	7	8	11	12	13	13	13
Notes de B	4	7	9	12	13	13	19

1°) *Déterminer les moyennes, les médianes et les modes des notes des deux élèves. Que constate-t-on ?*

1°) *Peut-on dire que les résultats de ces deux élèves sont comparables sur ce trimestre ? Pourquoi ?*

3.2 Étendue

Il s'agit de la première mesure de la dispersion d'une série statistique. Son principal mérite a longtemps été d'exister, et de fournir une information sur la dispersion très simple à obtenir. Mais elle est très largement insuffisante, précisément à cause de sa "naïveté".

Définition 5 *Soit X une série statistique discrète. On appelle étendue de la série X le réel, noté $Etd(X)$, défini par $Etd(X) = \max(X) - \min(X)$.*

3.3 Quartiles et intervalle interquartile

Définition 6 Soit X une série statistique. On appelle quartiles de X un triplet de réels $(Q_1; Q_2; Q_3)$, tel que les intervalles $[\min(X); Q_1[$, $[Q_1; Q_2[$, $[Q_2; Q_3[$ et $[Q_3; \max(X)[$ déterminent un partage des effectifs (ou des fréquences) de la série X en quatre parties de même importance.

Remarque 6 Par définition, si X est une série statistique, $Q_2 = \text{med}(X)$.

Le calcul des valeurs des quartiles se fait en général à partir des graphiques des effectifs (ou fréquences) cumulés croissants, par interpolation linéaire. La calculatrice donne (en principe) les valeurs de Q_1 et Q_3 , la valeur de Q_2 étant par définition une médiane de la série considérée.

Définition 7 Soit X une série statistique. On appelle intervalle inter-quartiles de X l'intervalle $[Q_1; Q_3[$. L'amplitude de cet intervalle est appelée écart inter-quartiles.

Par construction, l'intervalle inter-quartile est donc un intervalle qui comprend la moitié des valeurs de la série située la plus au centre de celle-ci.

3.4 Déciles et centiles

Définition 8 Soit X une série statistique. On appelle déciles de X un 9-uplet de réels $(Q_1; Q_2; \dots; Q_9)$, tel que les intervalles $[\min(X); Q_1[$, $[Q_1; Q_2[$, ..., $[Q_8; Q_9[$ et $[Q_9; \max(X)[$ déterminent un partage des effectifs (ou des fréquences) de la série X en quatre parties de même importance.

Remarque 7 Par définition, si X est une série statistique, $Q_5 = \text{med}(X)$.

Le calcul des valeurs des déciles se fait en général à partir des graphiques des effectifs (ou fréquences) cumulés croissants, par interpolation linéaire. Peu de calculatrices fournissent les déciles d'une série statistique.

Définition 9 Soit X une série statistique. On appelle centiles de X un 99-uplet de réels $(Q_1; Q_2; \dots; Q_{99})$, tel que les intervalles $[\min(X); Q_1[$, $[Q_1; Q_2[$, ..., $[Q_{98}; Q_{99}[$ et $[Q_{99}; \max(X)[$ déterminent un partage des effectifs (ou des fréquences) de la série X en quatre parties de même importance.

Remarque 8 Par définition, si X est une série statistique, $Q_{50} = \text{med}(X)$.

Le calcul des valeurs des centiles se fait en général à partir des graphiques des effectifs (ou fréquences) cumulés croissants, par interpolation linéaire. Peu de calculatrices fournissent les centiles d'une série statistique.

3.5 Variance d'une série statistique

3.5.1 Construction de la variance d'une série statistique

Afin d'éclairer cette construction, nous présentons celle-ci sur un exemple, qui illustre les raisons de procéder ainsi pour fabriquer un outil de mesure de la dispersion d'une série statistique.

Voici les résultats de deux élèves au cours d'un trimestre de classe :

Notes de A	10	10	10	10	10
Notes de B	11	5	18	7	9

Nous avons déjà observé que les moyennes de ces élèves étaient identiques, mais qu'elles ne traduisaient pas la régularité de l'élève B par rapport à A. Pour pouvoir prendre en compte les écarts par rapport à la moyenne, on peut mesurer les écarts de chaque valeur constituant la série : dans le cas de la série de notes de l'élève B précédent, dont la moyenne était de 10, on aurait donc :

Notes de B	11	5	18	7	9
Écarts à la moyenne	1	-5	8	-3	-1

Un tel tableau est redondant avec la série statistique, et en tout état de cause, il est aussi long ! Donc il nous faut trouver un moyen de condenser l'information. Pour résumer ce tableau à un seul nombre, on peut avoir envie de calculer la moyenne de ces écarts constatés à la moyenne : on va trouver ici $\frac{1 - 5 + 8 - 3 - 1}{5} = 0$! C'est le fait que le calcul de cette moyenne fasse figurer des nombres de signes opposés qui se compensent, qui

nous ont conduit à un résumé de l'information sur les écarts qui est nul. On va donc supprimer les signes, en considérant plutôt les écarts au carré par rapport à la moyenne ¹ :

Notes de B	11	5	18	7	9
Écarts carrés à la moyenne	$1^2 = 1$	$(-5)^2 = 25$	$8^2 = 64$	$(-3)^2 = 9$	$(-1)^2 = 1$

Puis, comme précédemment on va résumer ce long tableau à une valeur : sa moyenne. Pour la série considérée, ce calcul donne :

$$\frac{1^2 + (-5)^2 + 8^2 + (-3)^2 + (-1)^2}{5} \simeq 20$$

On calcule donc la moyenne des carrés des écarts constatés par rapport à la moyenne. Ce nombre est appelé la variance de la série statistique considérée.

Définition 10 La variance d'une série statistique X est le nombre $\mathbb{V}(x)$ obtenu comme moyenne des carrés des écarts constatés par rapport à la moyenne de la série.

Dans le cas d'une série sans coefficients, on a :

Théorème 6 Si X est une série statistique donnée par $X : \begin{array}{|c|c|c|c|} \hline x_1 & x_2 & \dots & x_p \\ \hline \end{array}$ dont la moyenne est notée $\bar{X}$, alors la variance de X , notée $\mathbb{V}(X)$, vaut :

$$\mathbb{V}(X) = \frac{(x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + \dots + (x_p - \bar{X})^2}{p}.$$

3.5.2 Cas d'une série discrète présentée selon les effectifs

Théorème 7 Si X est une série statistique, donnée selon ses effectifs $X : \begin{array}{|c|c|c|c|} \hline x_1 & x_2 & \dots & x_p \\ \hline n_1 & n_2 & \dots & n_p \\ \hline \end{array}$ alors la variance de la série X , notée $\mathbb{V}(X)$, vaut :

$$\mathbb{V}(X) = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{n_1 + n_2 + \dots + n_p}.$$

Remarque 9 Cette formule s'applique bien sûr au cas d'une série statistique sans coefficients : on est ramené à une série pour laquelle tous les coefficients valent 1.

Exercice 12 Voici les notes obtenues par un élève au cours du trimestre :

Notes obtenues	11	5	18	7	9
----------------	----	---	----	---	---

Calculer la variance des notes de cet élève.

Exercice 13 Voici les notes obtenues par un élève au cours du trimestre, ainsi que les coefficients qui leur sont attribués pour le calcul de la moyenne du trimestre :

Notes obtenues	11	5	18	7	9
Coefficients	1	4	8	3	2

Calculer la variance des notes de cet élève.

3.5.3 Cas d'une série discrète présentée selon les fréquences

La formule précédente s'applique aussi au cas d'une série statistique présentée selon les fréquences d'apparition, avec une adaptation due au fait essentiel que la somme des fréquences vaut 1 :

Théorème 8 Si X est une série statistique, donnée selon ses fréquences $X : \begin{array}{|c|c|c|c|} \hline x_1 & x_2 & \dots & x_p \\ \hline f_1 & f_2 & \dots & f_p \\ \hline \end{array}$ alors la variance de la série X , notée $\mathbb{V}(X)$, vaut :

$$\mathbb{V}(X) = f_1(x_1 - \bar{X})^2 + f_2(x_2 - \bar{X})^2 + \dots + f_p(x_p - \bar{X})^2.$$

Exercice 14

Exercice 15

Exercice 16

Exercice 17

¹En considérant la moyenne des valeurs absolues des écarts, on aurait obtenu un autre indicateur : l'écart moyen absolu, qui n'est pas employé, pour d'autres raisons que nous ne détaillons pas ici.

3.5.4 Cas d'une série continue présentée selon les effectifs

Théorème 9 Si X est une série statistique, donnée selon ses effectifs X :

$[a_1; b_1[$	$[a_2; b_2[$	$\dots$	$[a_p; b_p[$
n_1	n_2	$\dots$	n_p

, et si on note c_i le centre de chaque classe (donc $c_i = \frac{a_i + b_i}{2}$), alors la variance X , notée $\mathbb{V}(X)$, vaut :

$$\mathbb{V}(X) = \frac{n_1(c_1 - \bar{X})^2 + n_2(c_2 - \bar{X})^2 + \dots + n_p(c_p - \bar{X})^2}{n_1 + n_2 + \dots + n_p}.$$

Exercice 18

Exercice 19

3.5.5 Cas d'une série continue présentée selon les fréquences

Théorème 10 Si X est une série statistique, donnée selon ses fréquences X :

$[a_1; b_1[$	$[a_2; b_2[$	$\dots$	$[a_p; b_p[$
f_1	f_2	$\dots$	f_p

, et si on note c_i le centre de chaque classe (donc $c_i = \frac{a_i + b_i}{2}$), alors la variance de la série X , notée $\mathbb{V}(X)$, vaut :

$$\mathbb{V}(X) = f_1(c_1 - \bar{X})^2 + f_2(c_2 - \bar{X})^2 + \dots + f_p(c_p - \bar{X})^2.$$

Exercice 20

Exercice 21

3.6 Écart-type d'une série statistique

Définition 11 L'écart-type d'une série statistique X , noté $\sigma(X)$, est la racine carrée de la variance de cette série :

$$\sigma(X) = \sqrt{\mathbb{V}(X)}.$$

Exercice 22

On peut donc observer que :

Proposition 11 La variance et l'écart-type présentent les propriétés suivantes :

- i. La variance est toujours un nombre positif ou nul : l'écart-type est donc toujours calculable à partir de la variance ;
- ii. L'écart-type est un nombre toujours positif ou nul ;
- iii. Une variance nulle, ou un écart-type nul, signifient que toutes les valeurs de la série par rapport son égales à sa moyenne (il n'y a aucune dispersion autour de la moyenne) ;
- iv. Plus la variance (ou l'écart-type) d'une série est grande, plus cette série est dispersée autour de sa moyenne ;
- v. Entre deux séries de même moyenne, la série qui présente la plus grande dispersion est celle dont la variance (ou l'écart-type) est la plus grande ;

Exercice 23 On considère la série statistique suivante, qui est issue des résultats d'un contrôle :

<i>Intervalles de notes</i>	$[2; 5[$	$[5; 9[$	$[9; 13[$	$[13; 16[$	$[16; 18[$	$[18; 20[$
<i>Effectifs</i>	3	7	10	4	13	6

1°) Construire l'histogramme associé à cette série.

2°) Calculer la moyenne, la variance et l'écart-type de cette série.

3°) Tracer les polygones des fréquences cumulées croissantes et décroissantes ; en déduire la meilleure estimation possible de la médiane, et des quartiles de cette série.

Exercice 24

Exercice 25

Exercice 26

Exercice 27

Références