

DE LA MOYENNE A L'ECART-TYPE

Alain Macé
IREM de Rouen

Objectif principal :

Amener les élèves à construire eux-mêmes l'écart-moyen.

Énoncé :

Après quelques exercices préalables pour fixer les notions d'élève régulier et de résultats homogènes :

Dans une classe, un professeur de mathématiques veut examiner avec soin et comparer les résultats de tous ses élèves. Ce professeur, très consciencieux, décide de mettre au point un calcul, dont le résultat serait C , à faire sur les notes de chacun des élèves.

Le nombre C obtenu par le même procédé pour chacun des élèves répondrait simultanément aux deux exigences suivantes :

- C est positif
- Plus C est petit, plus les résultats sont **homogènes**.

Que proposez-vous comme calcul qui répondrait à ces deux conditions ?

Déroulement de l'activité :

Après une phase préliminaire essentielle d'appropriation de la notion de résultats homogènes, les élèves, par un travail de groupe, proposent un calcul qui puisse traduire cette notion. Un débat sur la pertinence des propositions faites, conduit à écarter de nombreuses propositions et à privilégier l'écart-moyen. L'écart-type se trouve défini au cours de la phase d'institutionnalisation.

Prévoir deux séances distinctes.

DE LA MOYENNE A L'ECART-TYPE

Objectifs :

- * Amener les élèves à construire eux-mêmes l'écart-moyen.
- * Introduire l'écart-type.
- * Réfléchir sur la signification des outils statistiques.
- * Utiliser le contre-exemple et le débat.

Situation :

** Pour ce qui est de la gestion de la classe :*

Il est souhaitable que cette activité ait lieu après avoir familiarisé les élèves à un travail de groupe avec débat et recherche de contre-exemples sur des énoncés ouverts. En effet, il s'agit d'«inventer» un nouvel objet, ce qui peut être très loin de la pratique de certains élèves habitués à être des exécutants.

** Pour ce qui est des contenus :*

Cette activité se place :

- n'importe quand dans l'année, même si la valeur absolue n'est pas encore vue.
- après quelques séances destinées à reprendre le vocabulaire statistique, les graphiques et les calculs de moyennes. A partir de quelques exercices (donnés en annexe 1), le professeur a fait réfléchir à la signification des diverses représentations graphiques et en particulier à la signification des «bâtons» comme étant proportionnels au «poids» dont chacune des valeurs est chargée. Il a fait préciser les notions d'«élève régulier» (par rapport à l'ordre chronologique) et de «résultats homogènes» (qui font appel à la notion de dispersion sans tenir compte du temps).

** Attention :* Dans les classes de secondes dites «faibles», cette activité doit être menée avec prudence et chaque étape parfaitement respectée.

Enoncé :(donné aux élèves en trois étapes)

I. Approche. (construction de séries plus ou moins homogènes)

Un professeur, après avoir étudié les notes de 2 élèves, déclare :
«Ces élèves ont exactement 9 de moyenne et l'un a des résultats beaucoup plus homogènes que l'autre».

Quelles pourraient être les notes de ces deux élèves ? (on donnera une dizaine de notes environ par élève).

II. Recherche (propositions de calcul pour C).

Dans une classe, un professeur de mathématiques veut examiner avec soin et comparer les résultats de tous ses élèves. Ce professeur, très consciencieux, décide de mettre au point un calcul, dont le résultat serait **C**, à faire sur les notes de chacun des élèves.

Le nombre **C** obtenu par le même procédé pour chacun des élèves répondrait simultanément aux deux exigences suivantes :

- **C** est positif
- Plus **C** est petit, plus les résultats sont **homogènes**.

Que proposez-vous comme calcul qui répondrait à ces deux conditions ?

III. Validation (fait suite à la restitution).

Sans faire de représentation graphique, mais en utilisant le calcul de **C** qui n'a pas (encore !) été refusé, déterminer lequel de ces deux élèves a les résultats les plus homogènes :

E1 : 05.10.12.16.14.08.14.12.08 ou E2 : 10.05.11.10.12.12.17.12.10.11 ?

Scénario :

Première séance (1 heure).

Approche : construction des séries plus ou moins homogènes.

Travail individuel puis par groupes de 2 ou 4 élèves. Débat interne au groupe sur la validité des exemples proposés. Intervention du professeur si nécessaire. 15 minutes dont 5 minutes en recherche individuelle.

Recherche : propositions de calculs pour C.

Travail individuel puis par groupes. Production d'un document qui sera rendu au professeur à la fin de l'heure. Aucune intervention du professeur.

Deuxième séance (1 heure).

Restitution :

Débat en classe entière : pourquoi accepter telle ou telle proposition ? Si elle a été suggérée, seule reste invalidée la somme des écarts à la moyenne.

Validation :

Le professeur propose à tous l'étude du même exemple pour «valider» la proposition d'écart-moyen.

Troisième séance (1/2 heure).

Institutionnalisation :

Le professeur établit un résumé de cours comportant : l'étendue, l'écart moyen, la variance et l'écart-type.

L'analyse présentée ci-après ne distingue pas les analyse a priori et a posteriori. Sous des formes proches, cette activité a été faite de nombreuses fois ; le scénario et l'analyse ici proposés en sont l'aboutissement.

Analyse :

L'approche :

Cette phase fait suite aux exercices (voir annexe 1) ayant permis de réfléchir à la différence entre «élève régulier» qui joue avec la relation d'ordre fixée par le temps et «résultats homogènes» où il est fait abstraction de tout ordre.

Cette phase permet à l'enseignant de s'assurer que l'idée d'homogénéité est acquise pour tous, il intervient donc si tel n'est pas le cas. De plus, elle donne aux élèves l'occasion de produire leurs propres exemples de séries de notes en réponse à des exigences données simples ; ils sont donc plus à même de produire des contre exemples par la suite.

Avant la séance suivante, le professeur regroupe les propositions des élèves, en préparant des contre-exemples. (cf. analyse).

La recherche :

Le travail de recherche ici demandé est une alternance de conjectures et de validations. Il s'agit pour l'élève, tout d'abord d'**imaginer** des réponses à un type de problème jamais rencontré. Certains élèves peuvent se trouver dérouterés par ce type de travail qui fait largement appel à l'intuition. La plupart des groupes formulent des propositions, mais peu opèrent un contrôle de validation des réponses apportées.

La restitution :

Au cours d'une deuxième séance, les différentes productions, avec le nom des auteurs, sont présentées à la classe. Le professeur peut, par exemple, avoir préalablement recopié celles-ci au tableau, mais d'autres formes de restitution sont possibles. Les élèves doivent alors débattre des propositions faites. Pour l'élève, il s'agit d'arriver à une critique de son travail et de celui des autres : justification ou invalidation.

Les différentes propositions sont, suivant la classe concernée :

* L'étendue.

* Le pourcentage de note non autour de la moyenne (non comprises entre $x - 1$ et $x + 1$).

- * Le rapport entre le nombre de notes supérieures à la moyenne de l'élève et le nombre de notes inférieures à sa moyenne.
- * L'addition de tous les écarts successifs entre les notes placées dans l'ordre chronologique.
- * La différence entre la moyenne des notes au-dessus de la moyenne et le moyenne de celles situées en-dessous (ou de même par rapport à 10).
- * L'addition de tous les écarts positifs de chaque note à 10, ou à la moyenne.

...

Dans cette phase de restitution-débat, il est important que la validation vienne des autres élèves et non du professeur. Celui-ci n'est là que pour diriger les interventions et éviter les débordements possibles. Il faut donc prévoir pour les propositions qui ne sont pas immédiatement rejetables, des temps de recherche sur la validité des solutions proposées.

Cependant, si pour une proposition, les élèves ne trouvent pas de contre-exemples (voir ceux cités en annexe 2), le professeur amorce le débat en donnant à réfléchir sur des séries de notes judicieusement choisies. Il est donc très vivement conseillé à l'enseignant d'avoir préalablement prévu ses propres contre-exemples pour ne pas alourdir cette phase.

La «validation» :

Le proposition de la moyenne des écarts à la moyenne ne peut être invalidée. Le calcul fait sur chacune des séries de notes E1/E2 a force de validation pour beaucoup d'élèves. Cependant les questions suivantes doivent apparaître :

«Est-on sûr que cela convient dans tous les cas ?»

«N'y aurait-il pas une autre possibilité ?»

Le résultat est sujet à caution car il n'est pas issu d'une démonstration : la validation repose en effet uniquement sur l'absence de contre-exemples. Ici, nous aboutissons en fait à une définition de ce terme. C'est donc l'enseignant qui affirme autoritairement que l'objet construit répond bien aux contraintes de départ dans tous les cas. Toutefois, il peut être intéressant de signaler aux élèves que la démarche scientifique adoptée ici est différente de celle qui est habituelle, à savoir *validation = démonstration*.

L'institutionnalisation :

La communauté scientifique appelle l'objet découvert, ou proposé par l'enseignant, «l'écart-moyen», mais il existe une autre possibilité très voisine : l'écart-type. La justification proposée pour l'écart-type, à savoir prendre la racine carrée des écarts pour éviter les valeurs absolues, est la seule possible et naturelle pour les élèves. Elle leur convient d'ailleurs parfaitement.

Signalons que, d'un point de vue historique, ces deux paramètres de dispersion sont apparus de manière concomitante, l'écart moyen ayant eu, dans un premier temps, les faveurs de Laplace et l'écart-type, celles de Gauss (voir annexe 4).

Une fiche de synthèse tient lieu à la fois de cours et de trace de l'activité. Les calculs sont effectués par les élèves et la conclusion est élaborée collectivement. (voir annexe 3).

Certains élèves proposent de poursuivre la réflexion :

- Pourquoi n'a-t-on pas écart-type égal écart moyen ?
- A-t-on toujours l'écart-type supérieur à l'écart moyen ?
- Pourquoi ne fait-on pas le calcul en prenant la racine carrée de la somme des carrés des écarts, puis, en divisant par n .

Réinvestissement :

Cette activité d'introduction de l'écart-type à partir des notes d'élèves se doit, bien évidemment, d'être relayé par quelques situations menant à une réflexion sur les divers caractères de position et de dispersion, en imposant un transfert par rapport à la situation initiale.

Annexe 1 : Exercices préparatoires

Élève régulier et résultats homogènes.

Exercice 1 : Comparaison des notes de trois élèves.

Trois lycéens ont eu pour notes en mathématiques l'an passé :

L1	4	6	3	9	10	8	12	10	19	12	20	12	18
L2	18	12	20	12	19	10	12	8	10	9	3	6	4
L3	8	12	10	10	16	12	10	12	9	15	9	8	12

- 1) Représenter graphiquement les notes de chaque lycéen en fonction de l'ordre des contrôles (3 couleurs sur le même graphique)
- 2) Représenter par des diagrammes en bâtons les effectifs (3 diagrammes différents).
- 3) Calculer la moyenne de chacun de ces lycéens.
- 4) En utilisant chacune des informations mises en évidence par les trois questions précédentes, comparer les résultats de ces lycéens.
- 5) Imaginez les commentaires de l'enseignant sur les livrets scolaires de chacun.

Exercice 2 : Comparaison des notes de deux classes.

Voici les notes obtenues au même contrôle par les élèves de deux classes C₁ et C₂:

11	12	9	15	1	18	7
6	10	7	2	14	2	15
4	7	14	13	6	17	16
13	9	15	5	10	4	5
15	7	5	9	11	14	14

7	17	8	11	12	9
0	12	10	2	8	19
7	5	11	1	9	10
3	12	8	11	13	17
12	11	15	10	12	9

- 1) Comparer les deux classes à l'aide des diagrammes de chaque série de notes.
- 2) Calculer les moyennes de chaque classe.
- 3) Quelle est la meilleure classe ?

Exercice 3 : Comparaison des notes de deux élèves.

Dans une classe, le professeur de mathématiques veut examiner avec soin et comparer les résultats de deux élèves. Voici leurs notes :

Élève E₁: 13 11 10 09 09 11 08 13 11

Élève E₂: 08 11 13 07 12 12 06 13 07 18

- 1) Calculer la moyenne de chacune des séries.
- 2) Représenter chacune de ces deux séries par un diagramme en bâtons (on placera ceux-ci l'un en dessous de l'autre).
- 3) Que peut-on en conclure ?

Annexe 2 : Productions d'élèves.

Quelques propositions et corrections établies par une classe.

1

Pour trouver C, il faut faire la moyenne de tous les élèves puis enlever toutes les notes autour de la moyenne puis diviser les autres par le nombre d'élèves total.

Correction : FAUX, car enlever toutes les notes autour de la moyenne c'est trop vague. De plus, il ne s'agit pas d'un seul élève.

On prend : la note la plus élevée – la note la plus basse.

Correction : FAUX.

On peut avoir $E_1 : 14.10.13.10.14.09$ et on aurait $c = 6$

Or, E_2 a des résultats plus homogènes que E_1 , le calcul proposé ne convient pas.

c = moyenne des notes les plus fortes – moyenne des notes les plus faibles

(c'est-à-dire en dessus et en dessous de la moyenne).

Pour A : moyenne des notes au dessus de 12 : 13

moyenne des notes au dessous de 12 : 11, d'où l'écart $c = 2$.

Pour B : moyenne des notes au dessus de 11,4 : 14

moyenne des notes au dessous de 11,4 : 11. D'où l'écart $c = 3$.

Ceci est contraire à ce que l'on cherche car B a des notes plus homogènes que A..

La moyenne $\times \frac{\text{la note la plus haute}}{\text{la note la plus basse}}$

Exemple 1 : $10,5 \times \frac{14}{7} = 21$ Le résultat est petit car les notes sont homogènes (7 notes sur 14).

Exemple 2 : $11 \times \frac{20}{3} = 73,3$ Le résultat est très grand car les notes sont hétérogènes. (3 notes sur 20).

Correction : FAUX.

Avec $E_1 : 20.19.20.20.19.02$ on a $m = 16,7$ et $c = 13,7 \times (20/2) = 167$.

$E_2 : 08.20.03.14.10.09$ on a m = 10.6 et c = $10.6 \times (20/3) \approx 71$

Selon ce calcul, on aurait E_2 plus homogène que E_1 . Or, nous constatons l'inverse.

5

En additionnant les écarts entre chaque note et en divisant par le nombre d'écarts, nous obtenons C.

$$C = \frac{A \quad B \quad C}{N}$$

$$C_1 \quad \begin{array}{cccccccccccc} 14 & 11 & 10 & 09 & 09 & 10 & 11 & 07 & 13 & 11 & : & \text{notes} \\ & \backslash & / & \backslash & / & \backslash & / & \backslash & / & \backslash & & \\ & 3 & 1 & 1 & 0 & 1 & 1 & 4 & 6 & 2 & = 19/9 & C = 2,1 \end{array}$$

$$C_2 \quad \begin{array}{cccccccccccc} 08 & 11 & 13 & 07 & 12 & 12 & 06 & 13 & 07 & 16 & : & \text{notes} \\ & \backslash & / & \backslash & / & \backslash & / & \backslash & / & \backslash & & \\ & 3 & 2 & 6 & 5 & 0 & 6 & 7 & 6 & 9 & = 35/6 & C = 5 \end{array}$$

Conclusion : C est plus petit car C₁ est plus homogène que C₂.

Correction : FAUX car l'ordre des notes intervient dans la solution proposée. Pour un élève qui aurait les mêmes notes que C₁, mais dans un autre ordre, on aurait un nombre C différent et des résultats tout aussi homogènes.

Par exemple pour $\begin{array}{cccccccccccc} 14 & 13 & 11 & 11 & 11 & 10 & 10 & 09 & 09 & 07 \\ & \backslash & / & \backslash & / & \backslash & / & \backslash & / & \backslash \\ & 1 & 2 & 0 & 0 & 1 & 0 & 1 & 0 & 2 \end{array}$ on a $c = 8$.

Annexe 3

Fiche de cours.

Les caractéristiques de dispersion.

Exemple :

Pour comparer les notes de tous les élèves d'une classe, on pourrait ...

Comment exprimer l'homogénéité des notes ?

Etude sur les notes de 2 élèves :

A : 05.10.12.16.14.08.14.12.08 B : 10.05.11.10.12.12.17.12.10.11

L'étendue

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

étendue 1^{re} série : 11

étendue 2^{me} série : 12

L'écart moyen.

L'écart moyen est la moyenne des écarts entre la valeur d'un caractère et la moyenne de ces valeurs.

1ère série : moyenne des valeurs $\bar{x} = 11$

valeurs A	effectifs	écarts à la moyenne	effectifs x écarts
5	1	6	6
8	2	3	6
10	1	1	1
12	2	1	2
14	2	3	6
16	1	5	5
total	9	XXX	26

écart moyen :

2,9

2ème série : moyenne des valeurs $\bar{x} = 11$

valeurs B	effectifs	écarts à la moyenne	effectifs x écarts
5	1	6	6
10	3	1	3
11	2	0	0
12	3	1	3
17	1	6	6
total	10	XXX	18

écart moyen :

1,8

Pour écrire la formule, il faut différencier les cas où $\bar{x} > m$ et $\bar{x} < m$, et donc utiliser la valeur absolue, outil peu pratique. Les mathématiciens définissent donc un autre calcul qui donne le même type d'information :

L'écart type.

L'écart type d'une série statistique est la racine carrée de la moyenne des carrés des écarts à la moyenne des valeurs du caractère, il est noté σ .

1ère série :

valeurs B	effectifs	écarts à la moyenne	carré des effectifs	effectifs x carrés
5	1	6	36	36
8	2	3	9	18
10	1	1	1	1
12	2	1	1	2
14	2	3	9	6
16	1	5	25	5
total	9	XXX	XXX	100

moyenne des valeurs $\bar{x} = 11$

moyenne des carrés des écarts : 11,1

écart type : $\sigma = \boxed{2,9}$

2ème série :

valeurs B	effectifs	écarts à la moyenne	carré des effectifs	effectifs x carrés
5	1	6	36	36
10	3	1	1	3
11	2	0	0	0
12	3	1	1	3
17	1	6	36	36
total	10	XXX	XXX	78

moyenne des valeurs $\bar{x} = 11$

moyenne des carrés des écarts : 7,8

écart type : $\sigma = \boxed{2,8}$

Conclusion de l'exemple étudié :

Les élèves A et B ont la même moyenne. Les notes de l'élève B ont une étendue supérieure à celles de A, mais elles sont plus regroupées autour de la moyenne : les écarts-moyens et écarts-types de B sont moindres que ceux de A. Les résultats de B sont, en effet, plus homogènes que ceux de A.

Annexe 4

Brèves données historiques

Paramètres de position et de dispersion

L'apparition des paramètres de position et dispersion est essentiellement liée à l'ASTRONOMIE, la position obtenue pour un objet céleste s'établissant à partir de séries d'observations.

Pour établir ces positions, le **mode** et «le milieu de l'intervalle défini par les valeurs extrêmes» semblent s'être imposés naturellement dans un premier temps. La moyenne et la médiane n'étaient pas encore un choix systématique pour le grec Ptolémée.

C'est au XVI^e siècle que Tycho Brahé a recours à la **moyenne arithmétique**. La **médiane** voit naître son intérêt en 1757 avec Roger-Joseph Boscovich. Au XVIII^e siècle, les astronomes étudient la «distribution des erreurs» pour déterminer la position des objets célestes à partir des premiers travaux dus à Galilée en 1632.

La **variance** naît au XIX^e siècle avec les moindres carrés, l'**écart-type** suit : Carl Friedrich Gauss propose en 1816, pour mesurer les variations des

erreurs en astronomie, l'usage de $\varepsilon_k = \left(\frac{1}{n} \sum_{i=1}^n x_i^k \right)^{\frac{1}{k}}$ (x_i désignant la valeur

absolue de l'écart de la i -ème observation à la moyenne arithmétique). L'écart-type ε_2 aura la préférence de Gauss aux autres mesures alors que l'écart-moyen ε_1 avait, dans un premier temps, les faveurs de Laplace. Au XIX^e siècle on utilisera plusieurs mesures de dispersions apparentées à l'écart-type. Ce n'est qu'en 1893 que Carl Pearson introduit le terme de «standart deviation» noté σ .

D'après *Histoire de la Statistique*
Jean-Jacques Droesbeke et Philippe Tassi
«Que sais-je?» n° 2527, P.U.F. 1990.