REPERES - IREM . n°6 - janvier 1992

MODE, MOYENNE, MEDIANE AU COLLEGE : POURQUOI ?

Jean-Claude DUPERRET
Irem de Reims

Préambule

Qu'on ne se méprenne pas sur le titre : je n'ai pas l'intention, en quelques lignes, de tenter de répondre aux questions : "Pourquoi enseigner les notions de moyenne et médiane en 3ème ? Pourquoi, de façon plus générale, enseigner les statistiques en premier cycle ?" (et pourtant, ce sont bien des questions de fond !).

Comme beaucoup de collègues certainement, j'ai découvert les statistiques et leur enseignement à l'occasion des nouveaux programmes de premier cycle. Mon premier souci a alors été : "Comment enseigner ces notions à mes élèves ?". En particulier, la *médiane* (qui, contrairement à la notion de *moyenne*, n'est pas une "connaissance exigible" du programme de 3ème !) a surtout été un prétexte à réinvestir d'autres notions (exemple : le théorème de Thalès) (voir [1] et [2]).

Mais je ne m'attachais pas à montrer à mes élèves pouquoi cette notion de médiane donnait à une série statistique un éclairage complémentaire à celui de la moyenne.

Tout ceci jusqu'à l'an dernier (1990-1991) où ...

Partie I : Méfiez-vous des informations

Un matin, sur une radio périphérique, j'entends le commentateur donner l'information suivante :

"La moyenne d'âge des gens qui suivent des cures est de 60 ans."

Et d'ajouter :

"Si la majorité des gens qui suivent une cure ont la soixantaine, il faut cependant remarquer que 50 % des curistes ont moins de soixante ans."

Sic ! ...

Dans le commentaire qui a suivi, j'ai cru comprendre que l'objet de cette information était le souci des villes d'eaux de s'équiper pour un public potentiellement jeune. Pour moi, quelle chance : j'avais l'intention de commencer les statistiques avec mes élèves de 3ème.

J'arrive donc dans ma classe et rapporte aux élèves l'information précédente en leur demandant s'ils sont d'accord avec la conclusion du commentateur. La plupart des élèves répondent spontanément *"oui"*. Mais quelques-uns (qui ont appris à se méfier de mes questions !) se contentent d'un *"peut-être"* ou d'un *"faut voir"* ...

Façon très polie de me renvoyer mon problème !

Je leur propose alors les deux situations correspondant aux encadrés 1 et 2 qui suivent. La première séquence est fondée sur l'étude d'une série de notes ; le seconde sur la répartition d'une population en fonction de l'âge.

Encadré 1 : *séquence 1.*

Situation 1
Histoire de notes.

Un professeur rend un devoir aux 31 élèves de sa classe.

Voici la liste des notes :

08	16	09	16	19	12	07	09	17
08	09	11	09	15	09	08	08	20
07	08	14	18	06	19	07	15	18
16	17	10	18					

L'étude de la première situation, débutant par la :

Question 1 : *Comment clarifier cette série de notes ? Quelle représentation graphique choisir ?*

les élèves tombent rapidement d'accord sur les deux réponses suivantes :

a) Rangement des données dans un *tableau,*

b) Construction d'un *diagramme en bâtons* (page ci-contre).

Vient ensuite la :

Question 2 : *A partir des informations précédentes, répondre aux questions suivantes :*

— Quelle(s) note(s) représente(nt) le plus grand nombre de copies ?

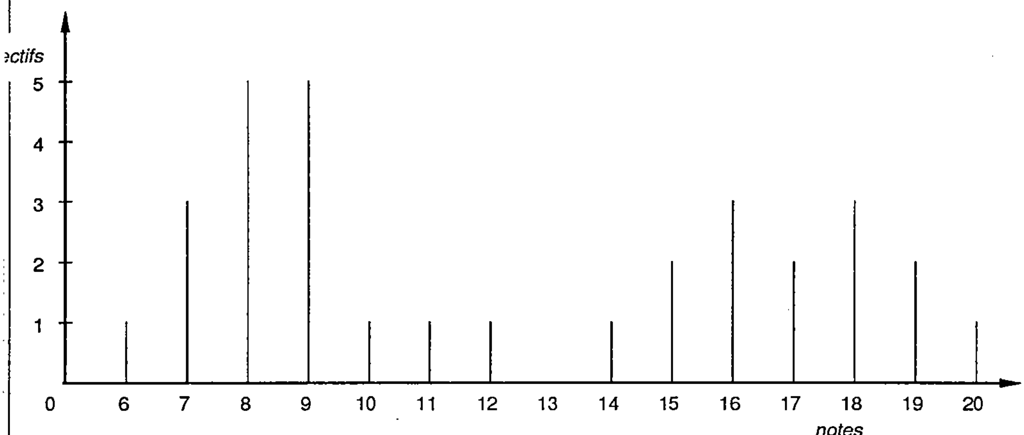
— Quelle est la moyenne de la classe à ce devoir ?

A la première question, la réponse vient rapidement : " 8 et 9 ".

Rangement des données dans un tableau.

Notes	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	Total
Effectifs	1	3	5	5	1	1	1	0	1	2	3	2	3	2	1	31

Diagramme en bâtons.



Pour la moyenne, il n'y a pas de difficulté non plus. C'est cependant l'occasion d'introduire la *moyenne pondérée*.

On trouve donc : $\bar{x} \approx 12,35$.

Nous revenons alors aux questions soulevées par le commentaire radiophonique.

Questions :

— La *moyenne* est environ 12,35 ; est-ce que les notes de la majorité des élèves sont proches de 12 ?

— Y a-t-il 50 % d'élèves qui ont moins de 12,35 ?

A la première question, les élèves constatent que 12 et 13 encadrant 12,35 sont les notes les moins "fréquentées".

A la deuxième question, ils arrivent au total de 17 élèves sur 31 qui ont moins de la "moyenne", soit 55 % des élèves.

Question 3 : *Existe-t-il une note telle qu'il y ait autant d'élèves ayant moins que cette note que d'élèves ayant plus ?*

Vu qu'il y a 31 élèves, il vient rapidement que la note cherchée doit être celle du $16^{ème}$ élève car il y en aura 15 "avant" et 15 "après".

Le tableau déjà fait suffit alors pour donner cette note :

$$m = 11 .$$

Bien que cela ne soit pas nécessaire (les élèves ayant suffisamment d'éléments pour répondre à la question posée), j'en profite pour rappeler des notions vues en 4ème sur les *effectifs cumulés croissants* (car je compte réinvestir cette notion dans la situation 2).

On établit donc le tableau des effectifs cumulés donné ci-dessous ; puis on peut résumer la situation à partir du diagramme de répartition des notes ...

Suit alors une première mise en forme de toutes ces notions suivie d'un premier débat avec les élèves sur la situation d'origine (celle des curistes).

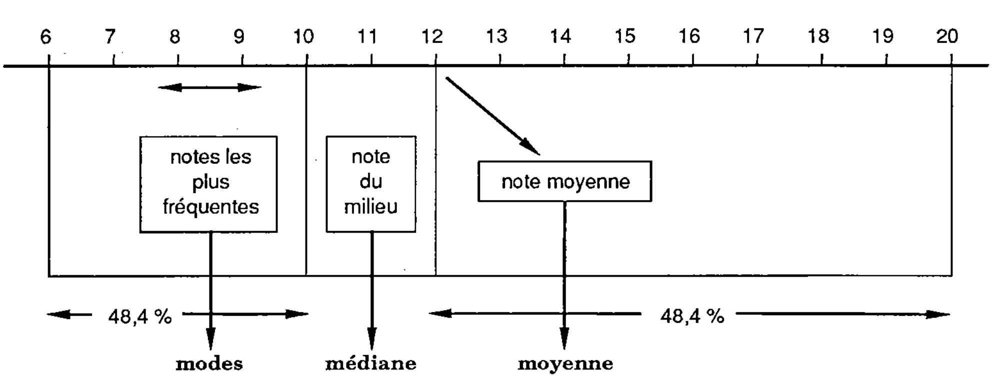
Pour avoir ensuite une situation plus proche de celle du départ, je propose aux

Tableau des effectifs cumulés croissants.

Notes	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
Effectifs cumulés	1	4	9	14	15	16	17	17	18	20	23	25	28	30	31

d'où la solution déjà proposée.

Résumé de la situation.



élèves le graphique de l'encadré ci-dessous représentant la répartition de la population française en "classes d'âges" au premier janvier 1986. Je leur précise que les centenaires ont été comptés dans la classe des 90-100 ans.

Comme précédemment la séquence est guidée par des questions.

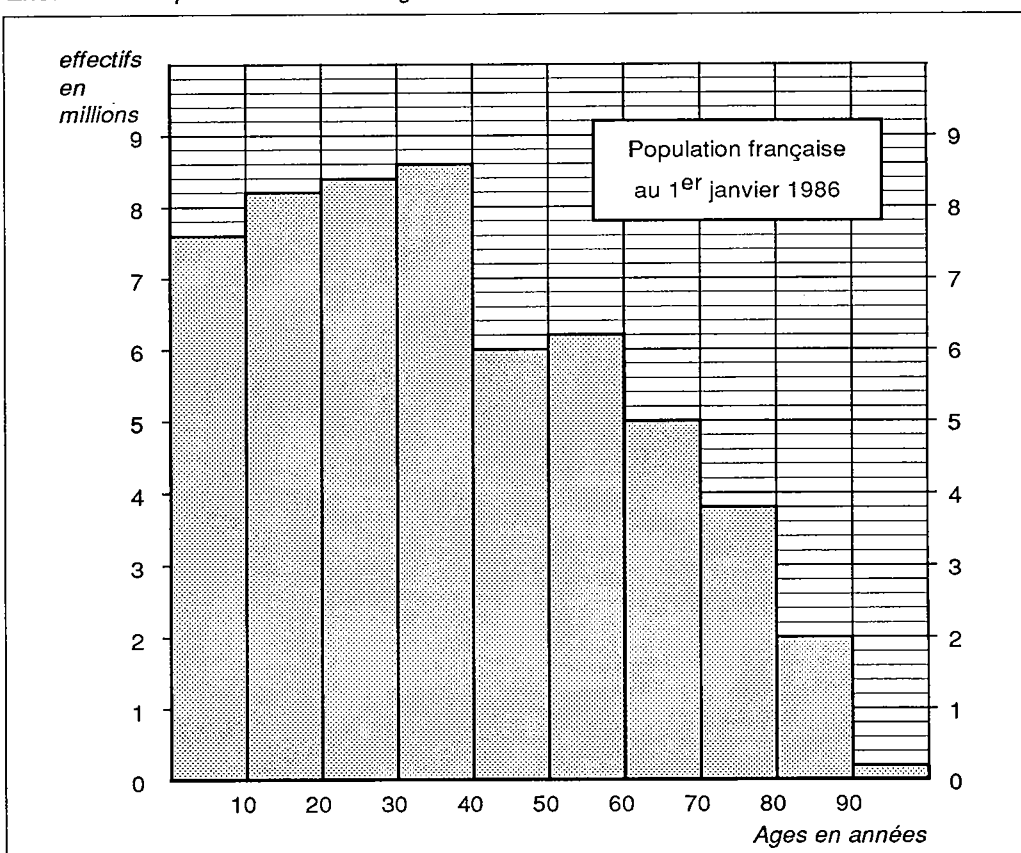
Question 1 : *Quelle est la classe d'âge la plus nombreuse ?*

Avec le graphique proposé, la réponse est immédiate pour les élèves : ce sont les personnes ayant entre 30 et 40 ans.

Question 2 : *Je cherche l'âge d'un Français "partageant" cette population en deux parties égales : ceux qui sont plus jeunes que lui et ceux qui sont plus vieux.*

La recherche par groupes qui suit permet de dégager deux courants de réponses :

Encadré 2 : *séquence 2. Histoire d'âges.*



Les premiers privilégient l'aspect numérique :

L'effectif total est de 56 millions.
Il y a 24,2 millions de Français qui ont moins de 30 ans.
Il y a 23,2 millions de Français qui ont plus de 40 ans.
Il faut donc chercher entre 30 et 40 ans et plus près de 30 que de 40.

Les seconds privilégient l'aspect géométrique :

Il faut partager le diagramme "verticalement" en deux diagrammes d'aires égales.

Je leur propose alors de les aider à affiner leurs résultats pour chacun des deux aspects.

Aspect numérique :

Question : *Combien de Français ont moins de 35 ans ?*

D'une manière naturelle, les élèves partagent la population des 30-40 ans en deux parties égales, attribuant la moitié de 8,6 millions au 30-35 ans. (L'hypothèse d'une répartition uniforme dans chaque classe semble un implicite admis en l'absence d'autres informations ! Cela m'a évité de me lancer dans des explications douteuses sur le choix d'une telle répartition.)

Ils obtiennent donc 28,5 millions (c'est-à-dire 24,2 + 4,3), ce qui leur permet de dire que le "Français du milieu" a entre 30 et 35 ans.

Je leur propose alors de réitérer le procédé : couper les 30-35 ans en deux parties égales, ce qui les amène à la conclusion que l'âge cherché est entre 32,5 et 35 ans.

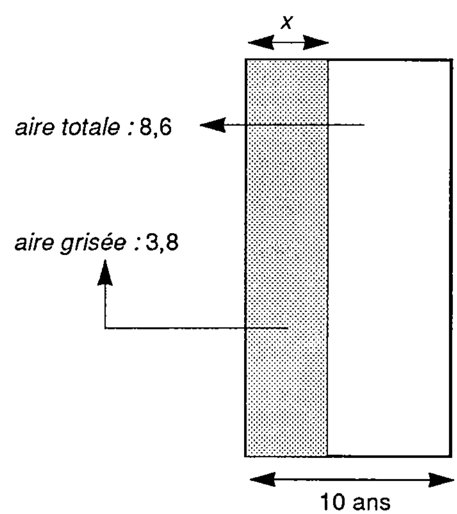
Pour préciser davantage, les élèves sont d'accord sur le fait qu'il suffit de réitérer le procédé.

Aspect géométrique :

Question : *Quelle est l'aire totale du diagramme ?*

La réponse 56 apparaît rapidement, mais bien entendu sans aucune précision sur l'unité d'aire !

On s'intéresse alors au rectangle dans lequel devra s'effectuer le partage : celui des 30-40 ans.



Je les aide à modéliser le problème (voir figure la ci-dessus) après leur avoir fait découvrir qu'il fallait trouver un rectangle d'aire 3,8 contenu dans un rectangle d'aire 8,6 ; ces deux rectangles ayant la même "hauteur" ...

Question : Calculer x ...

L'utilisation de la proportionnalité les conduit à trouver : $x \approx 4{,}4$.

D'où l'âge cherché : 34,4 ans.

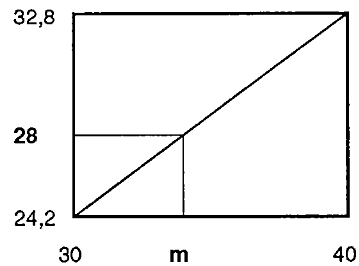
Des méthodes classiques (séquence 3).

Lors de la séquence suivante, je leur propose alors les "méthodes classiques" de recherche de la *médiane*.

1 — Méthode graphique (cf. figure ci-dessous) :

On utilise les polygones des effectifs cumulés croissants et décroissants. On trouve m ≈ 34,5 ans (avec implicitement la même hypothèse d'uniformité).

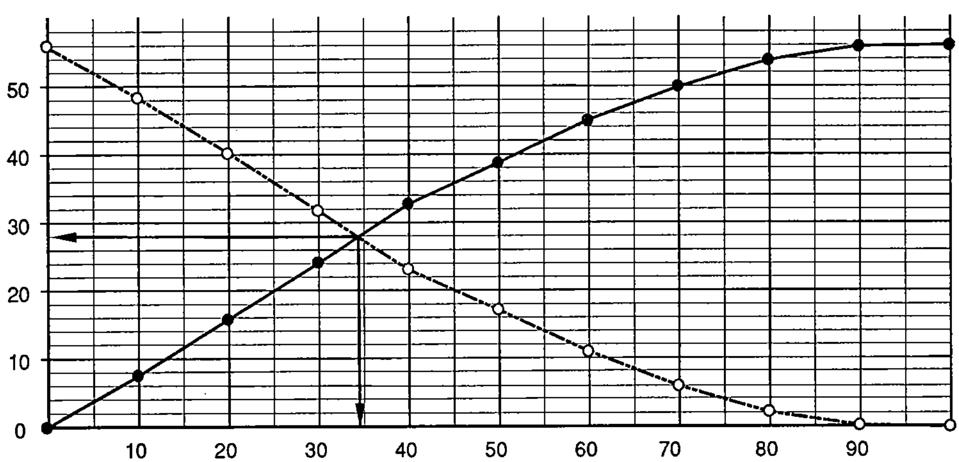
2 — Méthode calcul :



On utilise le théorème de Thalès ou la proportionnalité pour établir :

$$\frac{m - 30}{28 - 24{,}2} = \frac{40 - 30}{32{,}8 - 24{,}2}.$$

Pour la description de ces méthodes, voir [1] et [2].



Polygones des effectifs cumulés

Remarque : Dans cette séquence, j'avais pour objectif de réinvestir des notions de 4ème (effectifs cumulés) et peut-être, pour raison non avouée, de refaire "comme les années précédentes " (amortissement et sécurité !).

Mais avec le recul, cette séquence me paraît tout à fait inutile car elle n'apporte rien de plus aux élèves. Elle a même pour effet une certaine perte de sens de la notion de médiane qui est alors occultée par la présence d'outils mathématiques (effectifs cumulés, polygone, théorème de Thalès) qui ne sont pas forcément maîtrisés par les élèves.

En particulier, la notion d'*effectifs cumulés* me semble une notion difficile pour les élèves de premier cycle, difficulté due à la mauvaise maîtrise d'expressions telles que : "...ceux qui ont plus de...", "...ceux qui ont moins de...".

Et la moyenne ? (séquence 4)

Cette notion paraît *a priori* comme la plus facile et certainement la mieux maîtrisée par nos élèves ; mais, s'il est relativement aisé de donner un sens à la médiane (voir ce qui précède), comment définir en premier cycle la moyenne ? sinon comme le résultat d'un calcul (voire une touche sur la calculatrice) !

Pour justifier ce calcul auprès de mes élèves, je leur pose une première question :

Question : *En dehors de toute indication, quelle moyenne d'âge prendriez-vous pour chacune des classes ?*

La notion de "centre" est alors généralement proposée, sous le terme de "milieu". (Encore une fois, l'aspect "naturel" de la répartition uniforme évite bien des explications difficiles.)

Je leur demande alors :

Question : *En vous aidant de la situation 1, calculer l'âge moyen des Français.*

Pour les aider, je leur donne à compléter le tableau suivant, pour les amener à calculer la moyenne pondérée de :

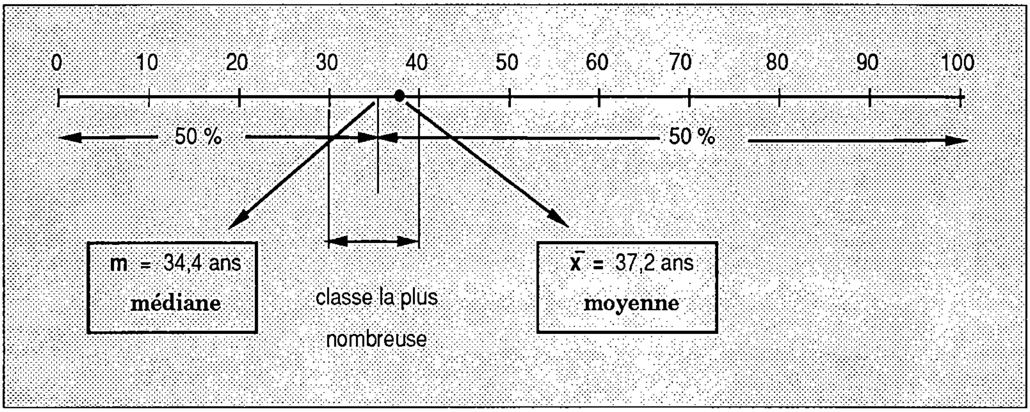
5 (7,6) ; 15 (8,2) ; ... ; **9,5** (0,2) .

Classe d'âge	Effectif en millions	Centre
$0 \leq x < 10$	7,6	5
$10 \leq x < 20$		
$20 \leq x < 30$		
$30 \leq x < 40$		
$40 \leq x < 50$		
$50 \leq x < 60$		
$60 \leq x < 70$		
$70 \leq x < 80$		
$80 \leq x < 90$		
$90 \leq x < 100$		

On obtient : $x \approx 37{,}2$ ans. On peut alors résumer la situation par le schéma de la page suivante.

En revenant alors à la situation d'origine (les curistes), j'ai pu constater dans le débat qui a suivi que tous les élèves avaient compris pourquoi les affirmations du commentateur de radio étaient *a priori* erronées.

Résumé des séquences 2 et 3.



Partie II :

"Les statistiques,
c'est comme le bikini,
ça donne une idée,
mais ça cache l'essentiel"

Louis Armand
(de l'Académie Française)

J'ai proposé, plus tard dans l'année et à des dates différentes, deux autres séquences de statistiques :

La première (séquence A. cf. encadré 3 p. 14) avait pour objectif de faire résoudre des problèmes originaux, et difficiles, pour préparer mes élèves à la participation au Rallye Mathématique Champagne Ardennes organisé par l'Irem de Reims.

La seconde (séquence B. cf. encadré 4) avait pour objectif de ne pas les laisser partir en seconde avec le sentiment que la connaissance de la moyenne et de la médiane suffisait pour décrire parfaitement une série statistique ...

Remarques :

Seules les expressions "moyenne" et "moyenne pondérée" apparaissent dans les programmes de premier cycle (et de seconde), avec tous les dangers d'interprétation de ces mots.

A travers les problèmes précédents, je voulais confronter mes élèves à d'autres types de moyennes.

Les problèmes 1 et 2 avaient pour objectif de leur apprendre à se méfier de l'utilisation du mot "moyenne" et de leur faire comprendre que la moyenne n'était pas *a priori* un élément de la série de départ. (Une famille ayant 2,8 enfants !)

Les problèmes 3 et 4 font intervenir la notion "d'inverse proportionnalité" donc la *moyenne harmonique* (inverse de la moyenne arithmétique des inverses).

Les problèmes 5 et 6 font intervenir la notion d'augmentation ou de diminution suivant une fonction du type $x \longmapsto (1 + t)x$ (ce qui est une compétence exigible du pro-

Encadré 3. Séquence A : *Vous avez dit "moyenne" !*

Problème 1 : C.G.T. - C.N.P.F. [3]

Dans une entreprise familiale, les salaires sont répartis comme suit :

Direction : le patron : 60 000 F
son frère : 30 000 F
6 parents : 10 000 F

Personnel : 5 cont. m^{ts} : 8 000 F
10 ouvriers : 4 000 F

La C.G.T. déclare que le salaire "moyen" est de 4 000 F.
Les contremaîtres s'estiment, eux, dans la "moyenne".
Quant au patron, il prétend que ses parents sont dans la "moyenne".

Chacun a raison, mais comment justifier leurs différences de position ?

Problème 2 : La photo [3]

Pour illustrer un article sur la population d'une ville, un rédacteur en chef demande à un photographe la photo de la famille moyenne de cette ville.

Le photographe prétend qu'il ne peut pas.

Pourquoi ?

Le rédacteur en chef demande alors la photographie d'une famille typique (la plus fréquente).

Nouveau refus.

Pourquoi ?

Problème 3 : Trajet automobile

Ayant eu à effectuer un trajet en voiture, j'ai parcouru la moitié de ce trajet sur des routes encombrées à une vitesse moyenne de 30 km/h. Pour la deuxième moitié de mon trajet, j'ai pris l'autoroute, et j'ai roulé à la vitesse moyenne de 90 km/h.

Quelle a été ma vitesse moyenne sur l'ensemble du trajet ?

Problème 4 : Equipement automobile

Deux villages ont le même nombre d'habitants.

Dans l'un, il y a une voiture pour quatre habitants.

Dans l'autre, il y en a une pour douze habitants.

Quel est le taux d'équipement moyen en voiture des deux villages réunis ?

Problème 5 : La population

Supposons que la population d'une ville augmente avec une vitesse proportionnelle au nombre d'habitants.

Sa population était de 250 000 habitants au premier janvier 1980, et de 490 000 au premier janvier 1990.

Quelle était sa population au premier janvier 1985 ?

Problème 6 : Crédit, prêt et inflation

On place une somme d'argent

S = 64 000 F

au premier janvier 1982, à un taux annuel t.

Le nouveau capital obtenu à partir de ce placement est de 81 000 F au premier janvier 1988.

Quel était ce capital au premier janvier 1985 ?

gramme de 3ème) donc la *moyenne géométrique de deux nombres* (racine carrée du produit de ces deux nombres, notion qu'on peut généraliser à n nombres avec la racine $n^{ième}$ du produit de ces n nombres).

Mais le traitement de ces problèmes ne demande en aucun cas l'introduction de ces moyennes, mais simplement un peu de réflexion et une bonne gestion d'outils du premier cycle (vitesse, fonctions, intérêt, ...).

Ils découvriront en seconde (sans que le mot ne leur soit prononcé) la *moyenne quadratique* (racine carrée de la moyenne arithmétique des carrés) avec l'*écart-type*. Mais je suis déjà dans la séquence B ! ...

Analyse succincte de la séquence A.

La classe devait trouver en 1 H 45 le maximum des quinze problèmes que je leur avais proposés (résultat numérique ou explication sommaire, comme dans le rallye R.M.C.A.). Ces six problèmes en faisaient partie.

Le *problème 1* a été résolu.

Le *problème 2* n'a été résolu qu'à moitié (personnes n'ayant pensé qu'il pouvait y avoir "deux" familles typiques (ou plus de deux !)).

Les *problèmes 3 et 4* ont été résolus (je n'ai pas de traces de la recherche !).

Les *problèmes 5 et 6* n'ont pas été résolus.

Séquence B : pour ne pas arriver trop sûr en seconde !

Cette deuxième séquence repose sur l'étude de deux situations (cf. encadré 4).

A travers la première, je voulais faire prendre conscience à mes élèves de deux notions que les tendances centrales ne

Encadré 4. Séquence B :
"Pour ne pas arriver trop sûr en seconde !"

Situation 1 :
Elève régulier ou dispersé ?

Voici les notes de quatre élèves de 3ème (en mathématiques !) :

Franck : 11, 11, 10, 9, 11, 12, 13, 11, 9, 13, 11.
Manuel : 9, 19, 10, 5, 11, 13, 17, 3, 15, 7, 12.
Lætitia : 1, 2, 6, 7, 10, 11, 12, 15, 16, 18, 19.
Cécile : 20, 19, 17, 15, 13, 11, 9, 7, 5, 3, 2.

Voici les quatre appréciations du professeur de mathématiques concernant ces élèves :

A : *Elève qui a régulièrement progressé grâce à un travail très sérieux.*

B : *Elève moyen qui s'en sort grâce à un travail très régulier.*

C : *Que de capacités gâchées ! Après un très bon début, les résultats se sont "écroulés" par absence totale de travail.*

D : *Elève fantaisiste mais capable. Doit progresser s'il accepte de fournir un travail régulier !*

Question : Redonnez à chaque élève son appréciation.

Et pourtant ! En regardant la colonne des résultats, je n'arrive pas à distinguer ces quatre élèves. Pourquoi ?

Et pourtant ! Fort de mes connaissances en statistiques, j'établis la "médiane" de chacun des élèves. Je n'arrive toujours pas à les distinguer. Pourquoi ?

Question : Est-ce que la connaissance des caractéristiques de position suffit à analyser une série statistique ? Comment qualifieriez-vous les notions supplémentaires qu'il faut ajouter pour mieux comprendre une série ?

Séquence B. *(suite)*

Situation 2 :
Maths ou français :
qu'est-ce qui rapporte le plus ?

(Travail proposé par groupes de deux : dans un premier temps l'un s'occupait des maths, l'autre du français.)

Voici la répartition des notes de deux cents candidats au brevet des collèges en maths et en français :

En maths :

Classe de notes	Effectif	Fréquence	Fréquence %	Effectif cumulé croissant	Effectif cumulé décroissant
$0 \leq x < 10$	2				
$10 \leq x < 20$	6				
$20 \leq x < 30$	8				
$30 \leq x < 40$	28				
$40 \leq x < 50$	38				
$50 \leq x < 60$	46				
$60 \leq x < 70$	42				
$70 \leq x < 80$	16				
$80 \leq x < 90$	10				
$90 \leq x < 100$	4				

En français :

Classe de notes	Effectif	Fréquence	Fréquence %	Effectif cumulé croissant	Effectif cumulé décroissant
$0 \leq x < 10$	0				
$10 \leq x < 20$	8				
$20 \leq x < 30$	25				
$30 \leq x < 40$	40				
$40 \leq x < 50$	55				
$50 \leq x < 60$	37				
$60 \leq x < 70$	26				
$70 \leq x < 80$	7				
$80 \leq x < 90$	2				
$90 \leq x < 100$	0				

peuvent pas mettre en évidence :

— *la notion du temps* : la chronologie des résultats n'apparaît plus.

— *la notion de dispersion* : à tendances centrales égales, l'étendue de deux séries peut être très différente.

Analyse de la deuxième situation.

La deuxième situation à étudier lors de la séquence B est donnée sous la forme de deux tableaux de résultats à comparer. La première question est traitée par groupes de deux élèves :

Question : *Pour la série que vous avez à étudier, remplissez le tableau, donnez la classe la plus fréquente, la moyenne et la médiane.*

On trouve :

En maths : 10 - 12 *En français :* 8 - 10

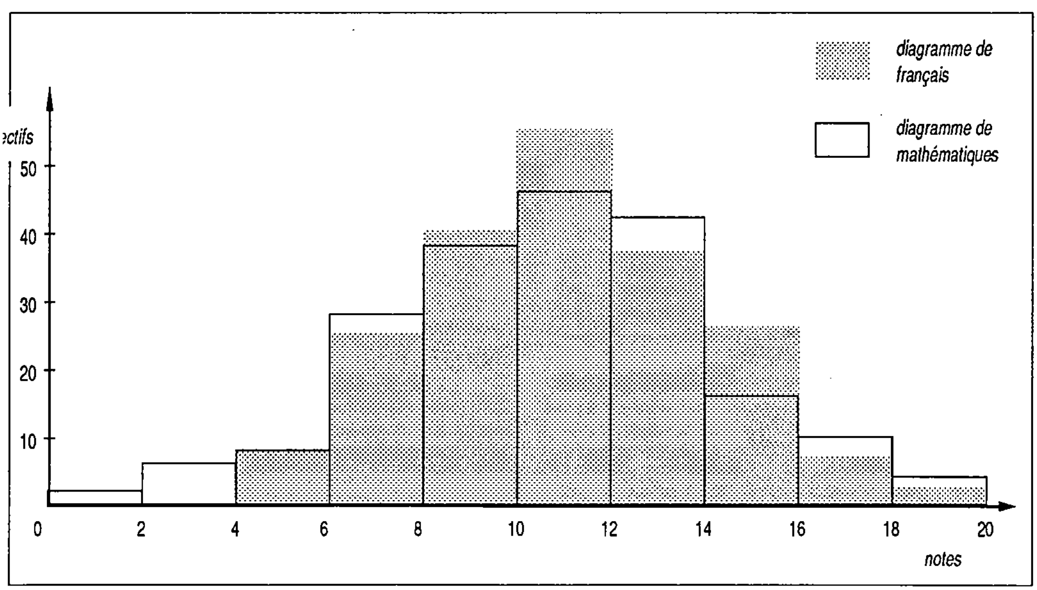
$\overline{x} \approx 10{,}68$ $\overline{x} \approx 9{,}04$

$m \approx 10{,}78$ $m \approx 8{,}98$

Après avoir fait constater aux élèves que pour chacune des séries, il y a une "bonne tendance centrale", je leur propose de les comparer en *relevant de deux points les notes de français*, ce qui correspond environ à l'écart des notes entre les deux matières.

Question : *Après avoir relevé de deux points les notes de français, représentez les séries sur un même graphique en bandes.*

(Ce graphique (voir ci-dessous) était à faire par les élèves en utilisant deux couleurs différentes.)



Question : *Même en faisant coïncider les caractéristiques centrales, on obtient deux graphiques différents. Comparer alors ces deux séries en complétant le tableau suivant :*

	Maths	Français
entre 10 et 12	23 %	27,5 %
entre 8 et 14	63 %	66 %
entre 6 et 16	85 %	91,5 %
entre 4 et 18	94 %	99 %
entre 2 et 20	99 %	100 %

(Ce tableau, complété par les élèves pour les pourcentages, avait pour objectif de mettre en évidence la différence de dispersion. Mais les nombres choisis ne font pas apparaître nettement cette différence. Je n'ai pas parlé d'écart-type, laissant cela pour la seconde !)

En guise de conclusion

— *Aux collègues de 3ème :* Bien que le collège (Albert Camus, voir aussi l'article de Pierre Bissey, Repères n° 3) où je travaille soit situé dans une Z.U.P. et appartienne à une Z.E.P., la classe de 3ème 1 (le numéro ne saurait tromper) dans laquelle j'ai fait ce travail en 1990-91 était ce qu'on appelle par tradition une bonne classe (latiniste). J'aurais certainement hésité à proposer certains problèmes (en particulier dans la partie II) dans une autre 3ème.

— *Aux didacticiens :* Ce que j'ai proposé ne peut pas vraiment s'appeler "activité" (au moins au sens où l'entend Régine Douady) ; j'ai en effet été relativement directif à la fin des temps de recherche et le cadre de travail des élèves était assez fermé (tableau, graphique, programme).

D'autre part, n'ayant pas lors de la préparation de ces séquences pensé qu'elles pourraient faire l'objet d'un article, je n'ai pas prévu d'évaluations des différentes séquences. Plus de six mois s'étant écoulés, je ne peux donc que parler globalement des élèves, en idéalisant certainement quelque peu leurs réactions.

— *Aux statisticiens :* Cet article peut leur apparaître comme simpliste et réducteur par rapport aux notions qui devraient être mises en place.

Je pense cependant que les situations et problèmes proposés sont déjà difficiles pour beaucoup d'élèves de 3ème. Je renvoie d'autre part à un article que j'avais écrit dans [2] et où je défendais une certaine conception de l'enseignement des statistiques.

En premier cycle, il peut avoir un rôle de formation sociale et civique pour mieux appréhender l'environnement :

— en sachant utiliser et choisir les indicateurs statistiques adaptés à la situation à étudier, ...

— en sachant interpréter et critiquer les éléments statistiques proposés pour présenter un phénomène dans un manuel scolaire, dans les journaux, à la radio, à la télévision, ...

A vous, lecteurs de la revue "Repères" : Il m'apparaît très important qu'à côté d'articles de fond sur l'enseignement ou l'histoire des mathématiques, ou d'articles présentant de très beaux résultats mathématiques, la revue "Repères" propose des articles plus "quotidiens", issus directement de notre pratique d'enseignants. C'est pour cette raison que j'ai bien voulu écrire celui-ci, en espérant que d'autres voudront bien continuer. C'est donc un appel : n'hésitez pas à nous faire parvenir vos propositions d'article sur un thème vous tenant à cœur !

Bibliographie

[1] Bulletin Inter-I.R.E.M. Premier Cycle : *"Suivi scientifique 3ème"*.

[2] Bulletin Inter-I.R.E.M. Premier Cycle - Niveaux d'Approfondissement : *"Liaison Collège-Seconde"*.

[3] *La Magie des Paradoxes* - Martin Gardner.