



HAL
open science

**Former à... enseigner la statistique au lycée,
perspectives en formation de formateurs et en formation
d'enseignants**

Christophe Hache, Jacqueline Mac Aleese, Aline Robert

► **To cite this version:**

Christophe Hache, Jacqueline Mac Aleese, Aline Robert. Former à... enseigner la statistique au lycée, perspectives en formation de formateurs et en formation d'enseignants. 2005. hal-03554138

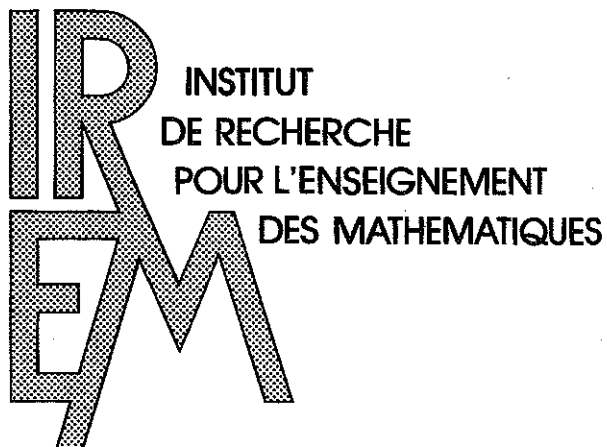
HAL Id: hal-03554138

<https://hal.science/hal-03554138>

Submitted on 3 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



7
Juin 2005

**Former à...enseigner la statistique au lycée
perspectives en formation de formateurs et en formation
d'enseignants**

C. Hache, J. Mac Aleese, A. Robert

DOCUMENT POUR LA FORMATION DES ENSEIGNANTS

UNIVERSITE PARIS 7-DENIS DIDEROT

7
Juin 2005

**Former à...enseigner la statistique au lycée
perspectives en formation de formateurs et en formation
d'enseignants**

C. Hache, J. Mac Aleese, A. Robert

DOCUMENT POUR LA FORMATION DES ENSEIGNANTS

UNIVERSITE PARIS 7-DENIS DIDEROT

Former à ... enseigner la statistique au lycée
Perspectives en formation de formateurs et en formation d'enseignants

(C. Hache, J. Mac Aleese, A. Robert)

INTRODUCTION

L'introduction relativement récente de chapitres de statistique dans le programme du lycée (et de collège) a amené à proposer aux enseignants des stages de formation sur cet enseignement.

Plusieurs problèmes se conjuguent, qui vont influencer les différentes formations.

D'abord ceux liés au **domaine mathématique**, qui est particulier notamment vu la place qu'y prennent *les applications pour d'autres secteurs*, très variés. Ainsi modélisation et interprétation sont des activités très liées à celles du statisticien, voire en partie intégrées à ses activités mathématiques, ce qui n'est pas habituel pour les autres champs mathématiques enseignés au lycée.

De plus la statistique amène à *traiter des données numériques importantes*. De ce fait calculatrices ou ordinateurs sont quasiment indispensables, ce qui est une autre spécificité de ce domaine. Les simulations notamment qui permettent, au lycée, de fabriquer les indispensables données « théoriques » issues de lois de probabilité hors de portée des élèves, sont explicitement au programme et ne peuvent se faire sans ces instruments.

Par ailleurs, comme nous le développerons par la suite, la statistique, y compris la statistique descriptive, utilise *implicitement* des notions mathématiques de façons détournées ou inhabituelles (notion d'ordre, de fonction réciproque...).

Enfin, et ce n'est pas le moindre problème, la statistique fait plus ou moins *appel aux probabilités*, sans qu'il soit possible d'enseigner au lycée tous ces préalables, souvent difficiles. C'est ce qui motive par exemple les simulations. Mais elle a aussi ses propres modes de raisonnement, qui jouent sur l'inconnu, l'incertain et la décision avec risque.

Or les programmes ont fait des choix drastiques en matière de transposition didactique, accentuant le décalage statistique/probabilités et ne contribuant pas à éclaircir les spécificités des raisonnements statistiques. Cela ne rend facile ni la justification des méthodes préconisées ni la détermination du degré de rigueur exigible dans les raisonnements, qui peuvent se transformer en applications de recettes, fautes de moyen d'entrer dans leur spécificité. Un seul exemple : les tests d'adéquation en terminale. Ils ne font travailler qu'à partir d'une seule loi théorique (mais ce qui pourrait varier n'est pas dit) et font comparer systématiquement les

neuvièmes déciles observés et théorique (i.e. simulé), enlevant ainsi aux élèves l'idée de décision du choix et de la valeur de l'écart...

On pourrait caricaturer ces caractéristiques par le fait que pour donner du sens aux activités statistiques, il peut être utile de sortir du cadre strictement mathématique habituel, à la fois pour être confronté à des problèmes sur des données effectives, issues d'expériences ou d'observations, pour faire des calculs importants, qui rendent nécessaires l'utilisation des instruments et pour raisonner de manière différente, toutes choses très difficiles à mettre au point en classe.

Ensuite il s'agit d'aborder précisément les questions liées à la **gestion des séances** de statistique : ce qui vient d'être souligné a des conséquences évidentes sur le déroulement des séances en classe. Comment organiser un travail sur des données ? Comment faire expérimenter et/ou observer ? Quelles sont les principales difficultés des élèves ? Comment, in fine, évaluer des activités aussi « floues » pour le mathématicien ?

Deux attitudes bien compréhensibles peuvent facilement apparaître : minimiser le temps consacré aux travaux statistiques, ou réintroduire des « vraies » mathématiques très vite.

Pour les formateurs, se pose de plus le **problème des ressources** : un certain nombre de publications sont disponibles, il s'agit d'en mettre au point la critique et d'en faciliter, le cas échéant, l'accès.

Enfin la question des **modalités des formations** se pose, là comme ailleurs !

En formation continue, nous adoptons un certain nombre d'hypothèses¹ qui doivent être adaptées à ce cadre précis et dont certaines mettent en jeu des variables indépendantes de nous.

La question du *temps long*, que nous jugeons nécessaire ne dépend pas toujours de nous.

En revanche, il appartient au formateur d'organiser un travail sur les pratiques, et pas seulement sur les connaissances, notamment en mettant en place si possible un travail simultané sur les contenus mathématiques enseignés et les déroulements des séances.

¹ Cf. Robert et Pouyanne 2004

Comment concevoir pour les enseignants des activités réelles, proches de l'expérience et des besoins ?

Mais il s'agit en même temps de tenir compte explicitement à la fois de tous les facteurs contraignants que les pratiques individuelles font intervenir (implicitement) et de la tendance de ces pratiques à se stabiliser sur le plan individuel en référence à des cohérences personnelles. La mise en évidence progressive des contraintes et la réflexion sur les marges de manœuvre et les alternatives est ainsi un moment important des formations.

Enfin il est important à nos yeux de tenir compte explicitement du fait qu'on forme des adultes, enseignants déjà en exercice : cela engage à mettre au point des modalités spécifiques, s'appuyant sur le bagage et les questions des enseignants présents, incluant des moments véritablement collectifs.

Dans ce cahier, nous allons montrer des choix effectivement faits en formations de formateurs et d'enseignants en **décrivant trois scénarios** de formation de formateurs et un stage de formation continue inscrit au PAF. Nous présenterons d'abord quelques réflexions sur la statistique et son enseignement au lycée, puis nous indiquerons un certain nombre de modalités assez générales mises en œuvre dans les formations ainsi que des activités précises proposées aux participants. Nous terminerons en indiquant les chroniques des formations concernées. Des annexes permettent de présenter certains des documents utilisés.

SOMMAIRE

Introduction	2
Sommaire	5
-I- Réflexions sur la statistique et son enseignement : glissements et savoirs cachés.....	6
1) Les objectifs déclarés des programmes :.....	6
2) Glissements et sources d'erreurs pour les élèves	10
3) Les savoirs cachés : trois exemples.....	13
-II- Modalités globales pour les formations et exemples d'activités.	31
1) Des modalités globales.....	31
2) Des activités choisies	32
-III- Exemples de scénarios en formation de formateurs : quatre chroniques.....	35
1) Master Professionnel, Option statistique/Probabilité, 2003-2004	35
2) Master Professionnel, Option statistique/Probabilité, 2004-2005, Groupe "Statistique descriptive"	37
3) Master Professionnel, Option statistique/Probabilité, 2004-2005, Groupe "hasard"	40
4) Déroulement des 3 jours du PAF 2004-2005 : "Statistique et probabilités de la seconde aux classes de terminales"	44
Conclusion.....	46
Bibliographie.....	47
Annexes	48

-I- REFLEXIONS SUR LA STATISTIQUE ET SON ENSEIGNEMENT : GLISSEMENTS ET SAVOIRS CACHES.

Dans ce paragraphe, nous allons rappeler quelques éléments sur les programmes puis indiquer des sources de difficultés spécifiques à l'apprentissage de la statistique que nous avons identifiées (glissements et sources d'erreurs pour les élèves) ; nous compléterons ce diagnostic par des exemples illustrant ce que nous appelons les savoirs cachés en statistique et la manière dont nous les abordons en formation.

1) Les objectifs déclarés des programmes :

L'analyse des programmes et des documents d'accompagnement laisse deviner des objectifs communs à toutes les sections du lycée, avec cependant des différences liées aux filières.

Dans toutes les sections

Ce qui est visé :

La phrase suivante, tirée des documents d'accompagnement de section ES, définit et justifie l'introduction actuelle de la statistique au lycée, dans les voies générales :

"La statistique, utilisée dans de nombreux domaines, n'est pas une mode mais une culture et un mode de pensée dont la diffusion est rendue possible par les développements de la théorie mathématique et de la technologie informatique".

Voici les objectifs essentiels : dans un premier temps, il s'agit de faire émerger l'intuition statistique grâce à de nombreuses simulations, observations et constatations (en seconde). Puis, en première, après en avoir fait sentir la nécessité, il faudra modéliser des situations aléatoires dans des cas simples et dans ce cas développer des calculs simples, en même temps que le vocabulaire classique de la théorie des probabilités. L'analogie entre probabilité et statistique sera fortement marquée. Enfin, en terminale, les problèmes de validation de modèles seront abordés, en utilisant la méthodologie de la simulation et de l'expérimentation vue en seconde, sans s'appuyer sur des théorèmes de probabilité inaccessibles à ce niveau.

Les programmes :

Tout au long des trois années de lycée, explicitement ou non, le théorème omniprésent est la loi des grands nombres, théorème qui énonce de manière tout à fait générale la

convergence de la distribution empirique vers la distribution théorique, dont un énoncé intuitif peut être donné aux élèves mais dont il est hors de question de donner la moindre preuve. C'est ce théorème qui justifie le passage de la notion de "distribution de fréquences" à celle de "loi de probabilité".

L'essentiel de ce qui se trouve dans les programmes correspond donc à des définitions et à des propriétés élémentaires. Il s'agira de développer des techniques de description statistique de séries numériques essentiellement : résumés numériques ou résumés graphiques, en initiant éventuellement des idées de comparaison de ces divers résumés.

Les probabilités apparaissent à partir de la première. Y sont développés les outils élémentaires classiques de cette théorie : arbres, dénombrements et combinatoire, conditionnement et indépendance, variables aléatoires.

Statistique et probabilité :

Les commentaires des programmes insistent bien sur l'opposition entre ce qui est empirique, sujet à fluctuation (objet de la statistique dans ces programmes) et ce qui est théorique, invariant (objet surtout des probabilités).

On trouve page 126 de ces commentaires les phrases suivantes :

"On garde à l'esprit que les distributions de fréquences fluctuent, la loi de probabilité restant fixe, d'où l'émergence de nouvelles questions liées à la reconnaissance d'une loi de probabilité à partir de données fréquentielles... On insiste sur le lien entre concepts probabilistes et données empiriques... entre propriétés des distributions de fréquence et propriétés des lois de probabilité, à partir de données provenant d'expériences de référence".

Toujours dans ces commentaires (pages 70 et 72), à propos de modélisation :

"Modéliser, c'est choisir un ensemble E et une loi de probabilité sur E "

"La modélisation ne relève pas d'une logique du vrai ou du faux"

"Modéliser consiste à associer un modèle à des données expérimentales et simuler consiste à produire des données à partir d'un modèle prédéfini".

Et à propos de simulation, on trouve :

"On indiquera que la simulation permet d'une part d'avoir des estimations de résultats impossibles à calculer explicitement, d'autre part de valider des modèles en comparant les résultats simulés et les résultats expérimentaux".

On ne s'intéresse ici qu'à la partie explicitement statistique des programmes.

Lecture des programmes selon les sections :

En seconde générale, les perspectives de cette partie du programme sont d'une part l'acquisition d'une expérience de l'aléatoire et l'ouverture sur le questionnement statistique, d'autre part l'abord de ce que peut être un modèle probabiliste (sans formalisation) dans des cas simples. Un des objectifs du programme est

"d'aller de l'observation à la conceptualisation, d'observer des phénomènes appelant une explication dans le champ des mathématiques".

Dans la voie économique et sociale ES, les domaines spécifiques nécessitent la maîtrise du maniement des pourcentages, la connaissance des séries chronologiques, l'utilisation et l'interprétation de tableaux croisés, l'utilisation et la manipulation de séries numériques doubles (représentations graphique et numérique). Par contre, la collecte des données, complexe, ne fait pas l'objet de réflexion ni de pratique.

En terminale, sont abordés les tests d'hypothèse, puisqu' on cherche à valider une modélisation, en utilisant des simulations et des caractéristiques numériques d'une distribution de fréquence.

La partie théorique classique de probabilité est, de fait, réduite (par exemple on ne parle que de lois binomiale avec $n \leq 5$). *En option, est introduite une autre thématique : celle des graphes probabilistes, liés au concept de processus aléatoires et de chaînes de Markov (sans le dire ...)*

Dans la voie scientifique S, l'accent est mis davantage sur l'aspect théorique puisque dès la première la notion de variable aléatoire et de loi de variable aléatoire est introduit (ce qui permet de s'écarter de l'hypothèse de loi uniforme, très présente avant) et en terminale la notion de variable aléatoire de loi continue est introduite, en lien avec les techniques d'intégration vues dans d'autres parties du programme. Comme en ES, est abordée la notion difficile de test d'hypothèse ou de validation de modèle.

D'autre part, les calculs classiques de dénombrement et de combinatoire, sans restriction de taille, sont développés.

Dans la voie littéraire L, y compris les options : les trois objectifs affichés de première sont d'une part

"de voir sur des exemples qu'une question peut trouver une réponse dans le champ de la statistique sous réserve de transformer la question initiale, de réfléchir sur les données à recueillir et la façon de le faire et leur traitement",

d'autre part de

"s'approprier la notion d'écart type et de moyenne, associées à la notion de plages de normalité dans la cas gaussien et enfin d'appréhender la notion d'indépendance par l'intermédiaire de l'étude de tableaux croisés".

Les options ont des objectifs plus théoriques puisqu' y sont introduites en première les notions de combinaisons (formule du binôme, arbre...) et en terminale, celles de loi de probabilité, de modélisation, de conditionnement et indépendance et de lois binomiales comme dans les séries scientifiques.

Lecture des programmes selon les grands thèmes statistiques abordés :

Statistique descriptive :

Ce thème a déjà été abordé au collège (moyenne d'une série discrète, représentations graphiques de séries statistiques).

On ajoute au lycée des définitions concernant des mesures de position et de dispersion: moyenne, écart type, médiane, étendue, quartiles, déciles, des représentations graphiques utilisant ces résumés numériques (diagrammes en boîtes). D'autre part, on travaille sur des données en classe, les hypothèses théoriques nécessaires pour faire les calculs de moyenne, médiane, quantile....n'étant pas toujours clairement explicitées.

Fluctuation d'échantillonnage :

Ce thème est introduit en seconde, c'est même un des objectifs principaux du programme de statistique. La fluctuation est facilement mise en évidence par l'usage de la simulation. Il faut bien sûr admettre que les outils utilisés pour obtenir de telles données sont fiables par rapport à l'aléatoire qu'ils fournissent.

Cette méthodologie est reprise en terminale S et ES, pour approcher une distribution théorique que les élèves ne peuvent pas connaître.

Tests d'hypothèses :

En terminale S et ES, on introduit ce concept et ce mode de pensée très spécifiques en mathématique : il n'y a pas une réponse vraie mais des décisions à prendre avec des risques d'erreur.

Une première réflexion sur les difficultés de ces programmes

A priori on peut déjà penser qu'on encouragera une utilisation de la technologie informatique avec toutes les difficultés à prévoir, notamment de gestion de la classe et du matériel. Ce sont les commentaires des programmes qui y engagent et l'inscription dans les programmes de la simulation qui peut difficilement se faire autrement. Notons que cette réflexion ne signifie pas que nous critiquons cette introduction de l'informatique mais vise à souligner la nécessité de prendre explicitement en compte les nouvelles difficultés.

Le manque de théorèmes peut gêner les enseignants du lycée, peu habitués à se limiter à un enseignement sans démonstration.

Par ailleurs il y a un risque de limitation de l'enseignement à la transmission de techniques et de règles, de fait très nombreuses. Ce risque se cumule avec la difficulté de replacer le travail dans un cadre de réflexion et d'interprétation statistique, peu familier dans la culture actuelle des enseignants de mathématiques.

2) Glissements et sources d'erreurs pour les élèves

Un de nos objectifs est de travailler sur les tâches proposées aux élèves en statistique : peut-on par exemple repérer des types de tâches spécifiques ? Y a-t-il des mises en fonctionnement des mathématiques particulières à ce domaine ?

Il s'agit aussi de comprendre si l'ensemble des exercices proposés aux élèves sur le chapitre « statistiques » ne conduit pas à privilégier un aspect du travail au détriment des autres, ce que nous appelons « glissement ».

a) Les glissements.

Dans un travail en collaboration avec Pablo Carranza², nous avons trouvé, notamment en analysant les exercices des manuels, que certains axes du travail sont privilégiés par rapport à d'autres sur l'ensemble des exercices proposés, avec des variantes. Ainsi avons nous repéré

² Dans le cadre de son mémoire de DEA soutenu en septembre 2004 qui se poursuit actuellement en un travail de thèse, voir Carranza 2004 et annexe page 2.

- des « glissements » algébriques :

Un nombre important d'exercices sur la moyenne conduit à des activités exclusivement algébriques : c'est souvent le cas des exercices faisant fonctionner les propriétés affines de la moyenne ou encore de ceux qui proposent le calcul de la moyenne dans une série sur des données en classes utilisant des interpolations affines qui masquent le travail statistique,

- des glissements informatiques

De nombreux exercices de simulation entraînent une utilisation de la calculatrice qui cache les activités statistiques (génération de nombres aléatoire, adéquation à une loi équirépartie)

- des glissements vers une statistique « arithmétique », exclusivement technique

C'est le cas en seconde de beaucoup d'exercices qui ne présentent aucun intérêt statistique, où celles-ci sont complètement artificielles.

b) Les sources d'erreurs

i) Les définitions sont particulières en statistique : elles sont en effet souvent arbitraires, variables d'un ouvrage à l'autre, peu reliées à une problématique statistique

Par exemple, on trouve plusieurs définitions de la médiane, notamment lorsque les données sont des classes... Du coup, on ne donne pas l'interprétation de la moyenne et de la médiane en terme de « minimum d'une certaine somme », et on ne précise par la différence deuxième quartile/médiane.

Différentes notions ne sont pas bien différenciées et surtout on ne présente pas toujours leurs intérêts respectifs : par exemple moyenne et médiane, écart type...

ii) Disponibilité implicite d'une utilisation non simple de l'ordre sur les réels

Tout se passe comme si les utilisations demandées aux élèves de l'ordre étaient parfaitement usuelles et disponibles, alors que souvent des changements de points de vue doivent intervenir : notamment pour les quantiles et même la médiane !

iii) Une utilisation non habituelle de relations fonctionnelles

La correspondance essentielle entre valeurs et effectifs est traitée de manière inverse de celle des x et des y habituelle. De même il y a une utilisation implicite de la « fonction réciproque » de la fonction de répartition. Enfin l'utilisation uniquement graphique de la fonction de répartition pourtant non définie peut s'avérer gênante.

iv) des activités complexes pour des tâches simples et isolées

En statistique même des applications immédiates de définitions n'engendrent pas des activités simples et isolées. Il y a presque toujours des étapes à reconstituer, par exemple ordonner les données et calculer les effectifs cumulés pour calculer une simple médiane : ce sont des activités mathématiques qui mettent en jeu des connaissances sur les réels, souvent sur l'ordre et qui amènent à faire des adaptations par rapport aux utilisations directes de ces propriétés dans leur domaine initial (travail dans \mathbb{R}).

Cette difficulté semble liée aux liens entre les savoirs statistiques et les autres savoirs mathématiques. La médiane, par exemple, est d'un certain point de vue une notion formalisatrice qui nécessite pour son utilisation la mobilisation de mathématiques qui peuvent s'avérer non triviales pour les élèves alors que, dans le champ de la statistique, c'est une notion élémentaire. On retrouve peut-être là une caractéristique des notions qui ne sont pas de simples extensions de notions déjà connues.

De plus dans les programmes on essaie de faire faire des activités algorithmiques pour des raisonnements qui ne le sont pas... On peut citer l'exemple du test d'adéquation où la démarche semble relever du simple algorithme de calcul tant que l'on ne souligne pas que la loi utilisée a fait l'objet d'un choix, que le seuil de "risque" aussi, et que, de ces choix et d'un calcul complexe, résulte un seuil pour un nombre "noté nd^2 " qui sert d'instrument de mesure.

Enfin, il y a peu de situations où les notions statistiques doivent être utilisées comme outils : cela tient en particulier au rôle des "vraies" données dans le travail du statisticien. Le travail du statisticien consiste en effet à choisir et à adapter à chaque nouvelle famille de données les outils qu'il va mettre en fonctionnement pour les traiter : ce choix tient à la fois sa connaissance de ce que chaque outil permet de renseigner et aux besoins supposés accrochés aux données. Il est difficile de mettre les élèves artificiellement dans un tel type d'utilisation, relative.

3) Les savoirs cachés : trois exemples

Ci-dessous trois exemples d'explicitation des savoirs mathématiques en jeu dans certains objets de statistique :

- Comment résumer une série statistique numérique ? 13
- Echantillonnage, fluctuation, fourchette 15
- Tests d'hypothèses, tests d'adéquation 20

Ces trois exemples correspondent à trois des interventions de Jacqueline Mac Aleese au Master Professionnel ou au stage PAF (voir aussi la partie III page 35).

Premier exemple

Comment résumer une série statistique numérique ? (ou exemple de glissement en statistique !)

Introduction

Face à une longue série d'observations numériques, que dire ? Une idée est de la remplacer par un seul nombre. Quelles propriétés attend-on de celui-ci ? qu'il soit à peu près égal à tous ceux de la série ou encore qu'il soit le plus proche de tous. Mais en quel sens ?

On va donner une version mathématiquement correcte, homogène des indicateurs connus que sont la moyenne et la médiane par exemple.

I Traduction mathématique :

- 1) Deux représentations "géométriques" de la série observée :
 - On représente les n nombres observés par des tirets sur une droite représentant \mathbf{R}
 - On représente la série observée (x_1, x_2, \dots, x_n) par un point x de \mathbf{R}^n et on cherche un point t dont toutes les coordonnées sont égales et le plus proche de x .
- 2) Quel critère de proximité entre deux points x et y se donner ?

Il y en a beaucoup !

Classiquement, on choisit un des trois suivants :

$$d_1(x, y) = \sum_{i=1}^n |x_i - y_i| \quad \text{ou} \quad d_2(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad \text{ou} \quad d_\infty(x, y) = \max_{i \in [1, n]} |x_i - y_i|$$

On cherche donc un nombre a tel que $d((x_1, x_2, \dots, x_n), (a, a, \dots, a))$ soit le plus petit possible : si ce a existe, a définira une position centrale et la valeur $d((x_1, x_2, \dots, x_n), (a, a, \dots, a))$ définira la dispersion autour de cette valeur centrale.

- 3) Exemple d'une série de deux nombres (ce qui n' a pas de sens statistique !):
par quel nombre remplacer le couple (8.9, 9.2) et quelles sont les différentes distances entre ce couple et le nombre proposé ?

II Calculs avec les choix précédents : des math !

1) Avec d_2 :

On cherche la (ou les) valeur a rendant $\sqrt{\sum_{i=1}^n (x_i - a)^2}$ minimal.

Le nombre le plus proche de tous les x_i est la moyenne et la dispersion est proportionnelle à l'écart type.

2) Avec d_∞ :

On cherche la (ou les) valeur a rendant $\max_{i \in [1, n]} |x_i - a|$ minimal.

Le nombre le plus proche de tous les x_i est la moyenne des extrêmes et la dispersion est proportionnelle à l'étendue.

3) Avec d_1 :

On cherche la (ou les) valeur a rendant $\sum_{i=1}^n |x_i - a|$ minimal.

Dans les bons cas, le nombre le plus proche de tous les x_i est la médiane et la dispersion est proportionnelle à la somme des valeurs absolues des écarts à la médiane. Sinon, il y a un intervalle solution du problème.

III Réflexions mathématiques et statistiques sur les résultats précédents

1) Les calculs sont plus ou moins faciles !

Avec d_2 , la valeur centrale est facile à obtenir, la dispersion moins.

Avec d_∞ , il faut commencer par trouver les valeurs minimale et maximale. Ensuite, c'est facile.

Avec d_1 , il faut commencer par ranger par ordre croissant toutes les valeurs observées, partager l'effectif en deux ... c'est moins facile.

2) L'utilisation de d_2 est fréquent : d_2 correspond à un produit scalaire sur \mathbf{R}^n , chercher a revient à faire une projection orthogonale d'un point sur un hyperplan.

Comme souvent en stat, on se ramène à un problème de géométrie euclidienne (régression linéaire, analyse de la variance, analyse factorielle)

3) Linéarité de la moyenne, de la médiane et de la moyenne des extrêmes. MAIS attention : la moyenne des moyennes est la moyenne, la médiane des médianes n'est pas la médiane ...

4) Robustesse ou sensibilité des valeurs centrales à des erreurs sur les valeurs observées : la médiane n'est pas sensible aux valeurs aberrantes, contrairement à la moyenne ou à la moyenne des extrêmes.

5) Quel couple préférer pour résumer ? Les trois ont un intérêt :

couple (moyenne, écart type) : linéarité, géométrie

couple (moyenne des extrêmes, étendue) : on peut remplacer ce couple par le couple (minimum, maximum), qui dit à peu près la même chose

couple (médiane, somme...) : la valeur centrale indique une répartition grossière de l'ensemble des nombres observés.

IV Description plus précise de la répartition de la série : diagrammes en boîte (Tukey)

On ne se contente plus d'une valeur pour résumer. L'objectif est de donner une idée graphique de la répartition sur \mathbf{R} des n nombres observés.

- 1) Définition des quantiles : premier et neuvième déciles, premier et troisième quartiles, médiane.

On peut remarquer que ces définitions font intervenir de manière très forte l'ordre dans \mathbf{R} , en utilisant simultanément les expressions "le plus petit tel que", "valeur observée inférieure à", "nombre au moins égal à".

- 2) Calculs de ces paramètres : c'est technique.

Il faut calculer les effectifs correspondants à "au moins un quart", "au moins un dixième"...et en déduire le rang de l'observation à retenir. Il faut donc ranger les observations par ordre croissant. On peut utiliser une représentation en branche et feuille, on peut aussi utiliser la représentation graphique des effectifs cumulés (fonction de répartition)

- 3) Représentation graphique : les boîtes.

- 4) Lecture statistique de ces boîtes :

largeur, symétrie, taille des "moustaches", comparaison de deux boîtes.

Deuxième exemple

Echantillonnage, fluctuation, fourchette : savoirs savants, savoirs enseignés, savoirs cachés

Introduction

On s'intéresse à un caractère d'une population : être bagué ou non dans une population d'oiseaux, changer ou non de voiture pour une classe de ménages, voter oui ou non à un référendum, gagner ou perdre à un jeu de pile ou face ...

On ne gagne pas à tous les coups, on ne peut interroger tous les ménages,... Comment faire pour "évaluer" ses chances ou la proportion de ménages ayant l'intention de changer de véhicule ?

Que veut dire "évaluer" ? il y a nécessairement un passage à une formalisation. Laquelle ?

A partir de quoi "évaluer" : sondage ? échantillonnage ? recensement ?

Quel est l'intérêt d'échantillonner quand on connaît la population ou on connaît la pièce qui sert à jouer ? Quel est l'intérêt d'étudier un modèle théorique ou d'approfondir la connaissance d'un modèle théorique ?

Une telle étude est-elle nécessaire à la connaissance de l'enseignant même s'il ne peut la transmettre à ses élèves ? Savoir enseigné, savoir savant, savoir caché, savoir accessible...

Dans ce type de situation, un des objectifs du programme de seconde est l'observation, la compréhension et l'utilisation du théorème suivant : "lors de n répétitions de lancer de

pièce, dans la plupart des cas, la fréquence observée de pile est une valeur approchée à $\frac{1}{\sqrt{n}}$ de la probabilité p de tomber sur pile". Quelles mathématiques cache cet énoncé très naïf ?

I Un modèle mathématique dans le cas simple d'une situation aléatoire à deux issues possibles :

- 1) On s'intéresse à un caractère prenant deux valeurs, que par convention on note 0 et 1. Lors d'un tirage avec remise dans la population donnée ou lors de répétition de lancers de la pièce, on observe une fréquence de 1, F_n : quel est le lien entre F_n et la vraie proportion p dans la population ou la vraie probabilité p de tomber sur pile.

A chaque tirage ou à chaque lancer on associe un résultat X_i , prenant la valeur 0 ou 1. **On sait que chaque X_i a une probabilité p de prendre la valeur 1.**

Dans la modélisation que je propose ici, je fais l'**hypothèse très forte que les répétitions ou les tirages n'ont pas d'influence les uns sur les autres** : il y a indépendance de la famille (X_1, X_2, \dots, X_n) .

En explorant mathématiquement ce modèle, on prouve des inégalités, on montre des théorèmes. En voici quelques uns, utilisés sans le dire, en seconde.

- 2) Théorèmes limites : LFGN, TCL

- a) **Loi faible des grands nombres** : c'est un théorème de convergence, en un certain sens, d'une suite de variables aléatoires.

Énoncés naïfs :

- la plupart du temps, la fréquence converge vers la probabilité quand le nombre d'observations augmente.
- la probabilité que l'écart entre la fréquence et la probabilité soit grand tend vers 0.

Énoncé mathématique :

Si (X_n) est une suite de variables indépendantes de Bernoulli de paramètre p et si la moyenne des variables $\frac{1}{n} \sum_{i=1}^n X_i$ est notée \bar{X}_n , alors pour tout a , la probabilité que l'écart entre \bar{X}_n et p dépasse a (en valeur absolue) tend vers 0 quand n tend vers l'infini:

$$\lim_{n \rightarrow \infty} \mathbf{P}(|\bar{X}_n - p| > a) = 0$$

- b) **Théorème central limite** : c'est aussi un théorème de convergence d'une suite de variables aléatoires.

Énoncés naïfs :

- quand n est grand, la moyenne empirique \bar{X}_n suit à peu près une loi normale de moyenne p et de variance $\frac{p(1-p)}{n}$.
- quand n est grand, la moyenne empirique centrée et réduite suit à peu près la loi normale centrée réduite.

Énoncé mathématique :

Si (X_n) est une suite de variables indépendantes de Bernoulli de paramètre p et si la moyenne des variables $\frac{1}{n} \sum_{i=1}^n X_i$ est notée \bar{X}_n , alors, pour tout a et tout b , la probabilité que la variable centrée réduite $Y_n = \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}}$ soit dans $[a, b]$ tend

vers la probabilité qu'une variable de loi normale centrée réduite soit dans $[a, b]$:

$$\lim_{n \rightarrow \infty} \mathbf{P}\left(\sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \in [a, b]\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

3) Math utilisées pour démontrer la LFGN :

- a) Inégalité de Markov : si X est une variable aléatoire positive de moyenne m , alors pour tout a la probabilité que X dépasse a est majorée par $\frac{m}{a}$:

$$\mathbf{P}(X > a) \leq \frac{m}{a}$$

- b) Inégalité de Bienaymé Tchébychev : si X est une variable aléatoire de moyenne m et de variance σ^2 , alors pour tout a la probabilité que l'écart entre X et m dépasse a est majorée par $\frac{\sigma^2}{a^2}$:

$$\mathbf{P}(|X - m| > a) \leq \frac{\sigma^2}{a^2}$$

- c) Moyenne et variance de la somme de n variables aléatoires indépendantes : c'est la somme des moyennes ou des variances des variables.
d) On applique BT à la variable $F_n = \bar{X}_n$, de moyenne p et de variance $\frac{p(1-p)}{n}$.

4) Math utilisées pour démontrer le TCL :

Cela peut se montrer à la main, avec des outils simples d'analyse : somme de Riemann, formule de Stirling, développements limités, continuité uniforme, intégrales de Wallis... (cf le petit livre de Emmanuel Lesigne : "Pile ou face : une introduction aux théorèmes limites du calcul des probabilités" chez Ellipses)

Sinon, cela utilise des outils de théorie de la mesure et la notion de fonction caractéristique d'une loi de probabilité.

5) Cas où on n'a pas besoin du passage à la limite :

Ces théorèmes sont utilisés, dans le cadre de notre problématique, pour évaluer la probabilité que l'écart entre \bar{X}_n et p soit plus grand qu'un nombre donné a .

Il y a des modélisations où le passage à la limite est inutile car on sait calculer exactement la probabilité que l'écart entre la moyenne empirique et la vraie moyenne dépasse un seuil fixé. C'est le cas quand on observe une série de mesures, modélisées chacune par une variable de loi normale de moyenne m et variance σ^2 . Alors la moyenne empirique suit la loi normale de moyenne m et de variance $\frac{\sigma^2}{n}$ et la moyenne empirique centrée réduite suit exactement la loi normale centrée réduite !

La loi normale centrée réduite est tabulée.

II Des proba vers les stat

alors que la démarche en seconde est l'inverse : des stat observées, non montrées, vers les proba !

1) Raisonnement statistique :

si p est la proba régissant l'expérience, alors on sait que F_n a la probabilité $1 - \alpha$ au moins de se trouver dans l'intervalle $[p - a, p + a]$ (ou on peut trouver un triplet (n, a, α) , taille d'échantillon, largeur de l'intervalle et probabilité, tel que ...)

On "retourne" l'inégalité pour dire : si p était la vraie valeur, avec proba $1 - \alpha$, on aurait p dans l'intervalle aléatoire $[F_n - a, F_n + a]$. On observe F_n et on conclut qu'une plage vraisemblable de valeurs de p est cet intervalle.

2) Liens entre n , a et α :

Comment jouer sur ces différentes "variables" ?

a) **Liens exacts** : recherche des bornes de l'intervalle de manière exacte à l'aide d'abaques. Lire l'article "Avant le référendum", publié sur le site "culturemath" à destination des enseignants de lycée (<http://dma.ens.fr/culturemath>)

b) **Liens approchés grossiers**, avec la majoration de BT (améliorables si on connaît d'autres majorations plus élaborées ou la loi exacte)

exemple 1 : $p = 0.3, a = 0.1, \mathbf{P}(|F_n - 0.3| > 0.1) \leq \frac{21}{n}$

Cette inégalité n'apprend rien si $n = 20$, elle dit qu'il y a une chance sur 2 d'observer une fréquence en dehors de $[0.2, 0.4]$ si $n = 42$, moins d'une chance sur 10 d'observer une fréquence en dehors de $[0.2, 0.4]$ si $n = 210$

exemple 2 : $p = 0.3, n = 100, \mathbf{P}(|F_{100} - 0.3| > a) \leq \frac{0.21}{100a^2}$

Cette inégalité dit qu'il y a moins d'une chance sur 5 pour observer une fréquence en dehors de $[0.2, 0.4]$ ($a = 0.1$) ; elle n'apprend rien sur la probabilité d'observer une fréquence en dehors de $[0.29, 0.31]$ ($a = 0.01$, le majorant est 21 !).

Cette inégalité n'est pas assez précise, elle permet de montrer une convergence mais elle ne contrôle pas bien la vitesse à laquelle celle-ci a lieu.

Cela permet de réaliser le lien entre ces trois nombres : taille de l'échantillon, largeur de la plage, probabilité que F_n soit dans l'intervalle $[p - a, p + a]$. Cela permet aussi de poser les questions : quelle précision a choisir, quelle majoration (α) de probabilité choisir, combien de répétitions n faire ?

c) Liens approchés plus fins mais asymptotiques avec le TCL :

Ce théorème permet de remplacer la majoration par une valeur approchée de la probabilité :

$$\mathbf{P}\left(\sqrt{n} \frac{|F_n - p|}{\sqrt{p(1-p)}} > a\right) \simeq \mathbf{P}(|Z| > a)$$

où Z suit la loi normale centrée réduite.

Si $a = 1.96$, $\mathbf{P}(|Z| > a) = 5\%$; si $a = 1.65$, $\mathbf{P}(|Z| > a) = 10\%$; si $a = 2.57$, $\mathbf{P}(|Z| > a) = 1\%$.

On en déduit les plages pour F_n en fonction de p , puis en résolvant les inégalités en p les plages pour p en fonction de F_n .

On peut encore simplifier ces inégalités en remplaçant p par F_n dans la variance.

3) Que peut-on dire aux élèves de seconde ?

Il a donc été obtenu une connaissance sur la vraie proportion (ou la vraie proba) à partir d'un échantillon en faisant des hypothèses de modélisation, en démontrant des théorèmes. La connaissance sur un modèle de paramètre inconnu vient après l'étude du même modèle avec paramètre connu.

Comment les élèves peuvent-ils accéder à cette connaissance sur le modèle à paramètre connu ? La théorie est trop complexe et le programme suggère donc de l'aborder expérimentalement :

- à p connu, on expérimente une simulation (d'après une modélisation faite par l'enseignant !) en faisant varier la taille de l'échantillon (n) et on fait des constats : variabilité de F_n à n fixé, convergence de F_n pour n croissant, fréquence du nombre de F_n dans un intervalle fixé à l'avance...

- à p inconnu de l'élève, mais connu du prof, l'élève reproduit la procédure précédente et utilise l'argumentation théorique vue plus haut : il obtient une fourchette de sondage.

- à p inconnu de tous (la vraie vie !), on ne peut pas faire de simulation ! Donc, soit on fait un recensement (quelquefois ou souvent impossible à réaliser), soit on fait des sondages répétés dans la population étudiée pour obtenir une plage expérimentale de valeurs possibles (cela demande du temps et des moyens - voir le traitement de l'adéquation en terminale), soit on fait un seul sondage et on utilise la théorie précédente.

III Textes de référence et quelques pistes de travail:

- le CD des programmes de mathématiques, il y a tout dedans : les programmes de seconde à la terminale, les onze fiches de stat, les commentaires des programmes, des sites internet...

- le site [http:// dma.ens.fr/culturemath](http://dma.ens.fr/culturemath) : on y trouve des dossiers sur des thèmes mathématiques à l'intention des profs

- un site SEL (statistique en ligne) : www.inrialpes.fr/sel

- Article "jeu de pile ou face", de l'Irem de Lille sur "culturemath".

Il s'agit d'évaluer la probabilité de gagner dans un jeu à deux à la règle un peu complexe. La théorie permet de calculer effectivement la valeur mais sous une forme non close (somme d'une série !)

Que peut-on faire explorer aux élèves ?

- Article "Equirépartition d'une suite de nombres" de T. Chomette et F. Boucekkine sur "culturemath" :

quelle est la loi du premier chiffre de 2^n dans son écriture décimale, est-ce une loi uniforme sur $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$?

Les mathématiques sous-jacentes sont complexes, que peut-on faire explorer aux élèves de seconde ?

IV Travail sur quelques exemples :

L'objectif là est d'appliquer ce qui vient d'être dit et d'essayer d'imaginer quelques activités avec des élèves.

1) Un des problèmes classiques du Chevalier de Méré, joueur invétéré, ami de Pascal :

Est-il préférable de parier sur l'apparition d'au moins un 6 en lançant quatre dés (cubiques parfaits) ou sur l'apparition d'un double six en lançant deux dés vingt quatre fois de suite ?

Les enseignants peuvent résoudre le problème exactement. Les élèves de seconde n'ont pas les outils pour le faire. L'outil qui leur est proposé est l'étude expérimentale, réelle ou simulée, qui permet de faire des conjectures.

(exercice inspiré du n°44 page 248 du Déclic de 1ère S de 2001)

Résoudre l'exercice et concevoir une activité pour les élèves.

2) Evaluation de l'aire d'une ellipse :

Utiliser les stat pour résoudre ce problème repose sur le principe suivant : l'ellipse est inscrite dans un rectangle, si on prend "au hasard" un point dans ce rectangle la probabilité qu'il a de se trouver dans l'ellipse est égale au rapport entre l'aire de l'ellipse et l'aire du rectangle. (Ceci est une modélisation probabiliste à faire accepter ou admettre, comme prendre un nombre au hasard dans l'intervalle $[0, 1]$).

A partir de ce principe, on peut mettre en place une procédure expérimentale statistique.

Concevoir une activité pour les élèves.

(exercice inspiré du n°43 page 248 du Déclic de 1ère S de 2001)

Troisième exemple

Tests d'hypothèses, tests d'adéquation : savoirs cachés du programme de terminales S et ES

Introduction

Dans quelles situations utilise-t-on ce mot "test" ?

Que trouve-t-on dans les livres ?

a) Définitions du dictionnaire Larousse en 5 volumes :

- essai d'un produit, d'un appareil pour vérifier son action, son fonctionnement,
- toute circonstance qui permet d'éprouver, de mesurer quelque chose
- manoeuvre, épreuve, réaction biologique, sérologique ou chimique pratiquée à des fins diagnostiques
- test-match

- test statistique ou test d'hypothèse : méthode permettant, à partir d'observations d'un ou de plusieurs échantillons d'une population, d'accepter ou de rejeter, avec un certain risque d'erreur, une hypothèse portant sur la population ou sur la loi de probabilité choisie pour représenter celle-ci.

b) Dans un livre d'introduction à la statistique (Morgenthaler), en début de chapitre sur les tests avant la formalisation :

"Les tests statistiques sont utilisés pour répondre aux questions liées aux paramètres, par exemple pour savoir si oui ou non une espérance est plus petite qu'une certaine constante. Ces réponses doivent être apportées sur la base de mesures soumises à des erreurs aléatoires."

c) Dans un livre de la collection "statistique et probabilités appliquées", dont l'objectif est de rendre accessibles les fondements théoriques de la statistique à un public de niveau deug, ingénieur, chercheur appliqué (Lejeune), en introduction au chapitre "Tests d'hypothèses paramétriques" :

"Les tests constituent une approche décisionnelle de la statistique inférentielle. Un tel test a pour objet de décider sur la base d'un échantillon si une caractéristique de la population répond ou non à une certaine spécification que l'on appelle hypothèse. Ces spécifications peuvent avoir diverses provenances : normes imposées, affirmations faites par un tiers, valeurs cruciales de paramètres de modèles..."

C'étaient trois citations allant du plus commun au plus mathématique.

Prenons trois exemples de situations :

- exemple 1 : dans une étude de pollution atmosphérique, on s'intéresse à la quantité de particules de diamètre inférieur à 2.5 microns, dans la mesure où ces particules peuvent pénétrer nos poumons et provoquer une intoxication si la quantité respirée est supérieure à un certain seuil.

- exemple 2 : le responsable d'une chaîne de production de compteurs électriques veut savoir si les appareils produits sont bien calibrés, c'est à dire si la vitesse de rotation est la bonne (ceci est du contrôle de fabrication).

- exemple 3 : avant d'entrer dans un jeu de hasard, un joueur souhaite savoir si la pièce de monnaie utilisée est équilibrée.

I Elaboration d'une démarche mathématique pour proposer une réponse

Dans ces 3 situations, il s'agit donc de savoir si une caractéristique, une spécification est satisfaite.

Dans chacune, une observation (quantité de particules de diamètre inférieur à 2.5 microns mesurée dans un volume d'air donné à un endroit donné, vitesse de rotation du compteur électrique, côté de la pièce obtenu) a un résultat aléatoire.

On peut donc penser utiliser une modélisation probabiliste et les techniques en découlant pour trouver des réponses argumentées aux questions posées.

Il faut alors faire un travail de modélisation probabiliste de la situation, proposer un modèle où la probabilité dépend d'un paramètre inconnu et où l'hypothèse faite, la question posée, se traduit sur ce paramètre :

-exemple 1 : la quantité de particules mesurée est une variable aléatoire de moyenne m . La question est : a-t-on $m \leq Norme$?

-exemple 2 : la vitesse de rotation du compteur est une variable aléatoire de moyenne m . La question est : a-t-on $m = m_0$, où m_0 est la vitesse calibrée sur la chaîne de production ?

-exemple 3 : le résultat d'un lancer de pièce est pile ou face et la probabilité qu'elle a de tomber sur pile est inconnue valant p . L'hypothèse se traduit par : a-t-on $p = \frac{1}{2}$?

Comment prendre une décision ?

L'idée est de faire plusieurs observations indépendantes du même phénomène et de trouver une méthode pour juger si l'écart entre les données récoltées et ce qui est attendu quand l'hypothèse est satisfaite est significatif. Il faut donc **choisir** une **mesure de cet écart** : plus l'écart sera grand, moins l'hypothèse paraîtra vraisemblable. Un écart "grand" (cette appréciation est à préciser) conduira donc à la décision de rejeter l'hypothèse.

Mais il y a un bémol : il se pourrait que l'hypothèse soit satisfaite et que l'écart soit quand même grand. Cela arrive-t-il souvent ?

Il faut répondre à la question plus précise de **signification d'une grande valeur de l'écart**, et pour cela, évaluer la chance d'obtenir un grand écart quand cette hypothèse est satisfaite : il faut connaître la façon dont cet écart varie sous l'hypothèse, c'est à dire connaître la loi de probabilité de cet écart sous l'hypothèse testée.

Traisons l'exemple de la pièce :

- l'observation du lancer de la pièce est modélisée par une variable aléatoire X de loi de Bernoulli, de paramètre p : $P_p(X = 1) = p$, $P_p(X = 0) = 1 - p$. C'est le modèle probabiliste, ou plutôt statistique choisi.

- l'hypothèse : "la pièce est équilibrée" se ramène donc à l'hypothèse sur la valeur p de la probabilité qu'a la pièce de tomber sur pile : $p = \frac{1}{2}$.

- **on choisit** comme **observation** 8 lancers successifs indépendants de la pièce et on ne retient de l'observation que le nombre de fois où la pièce est tombée sur pile, soit S . A quelle valeur de S s'attend-on ? plutôt autour de 4.

On choisit pour mesurer l'écart entre l'observation et ce qui est attendu quand l'hypothèse est vraie le nombre $|S - 4|$ et **on choisit** de dire que l'écart est grand s'il est supérieur ou égal à 3, c'est à dire si S vaut 0, 1, 7 ou 8.

Par exemple, avec ces choix, l'observation de la valeur 1 pour S conduit à la décision : la pièce n'est pas équilibrée ! et l'observation de la valeur 4 pour S conduit à la décision : la pièce est équilibrée !

- mais ces décisions peuvent être erronées : quels sont les risques qu'elles le soient ?

La probabilité d'observer ces valeurs extrêmes de S sous l'hypothèse " $p = \frac{1}{2}$ " se calcule facilement :

en effet, sous cette hypothèse, la variable aléatoire S suit la loi binomiale de paramètre 8 et $\frac{1}{2}$ et donc la probabilité qu'elle prenne ces 4 valeurs est $\frac{1}{2^8}(C_8^0 + C_8^1 + C_8^7 + C_8^8) = \frac{9}{128} = 0.07$.

L'observation des valeurs 0, 1, 7, 8 est peu probable si l'hypothèse " $p = \frac{1}{2}$ " est vraie : il y a 7 chances sur 100 d'observer une de ces valeurs, alors qu'il y a 93 chances sur 100 d'observer une valeur comprise entre 2 et 6.

On peut se tromper en prenant la décision d'accepter l'équilibre après l'observation d'une valeur de S comprise entre 2 et 6 car la pièce peut ne pas être équilibrée ; si p est la probabilité qu'à la pièce de tomber sur pile, la probabilité de cette erreur de décision est $\sum_{k=2}^6 C_8^k p^k (1-p)^{8-k}$, qu'on ne peut évaluer que si on se donne la valeur de p .

Par exemple, si $p = \frac{3}{4}$, cette probabilité d'erreur de décision vaut environ 0.63.

Bilan du travail fait :

- identification de la chose certaine inconnue, sur laquelle on pose une question.
- modélisation de la situation par un modèle probabiliste connu mais de paramètres inconnus.
- traduction de la question posée en a) dans le modèle.
- choix d'un écart
- choix des valeurs extrêmes de cet écart
- calcul de l'écart pour l'expérience faite et prise de décision
- calcul de la loi de probabilité de l'écart, évaluation des risques d'erreur des décisions prises et retour sur le choix des valeurs extrêmes

Il faut remarquer qu'avant l'expérience, il y a un inconnu aléatoire : c'est le résultat de l'expérience et il y a un inconnu certain : c'est la loi qui régit l'expérience (loi qui peut appartenir à une famille paramétrée simple comme la famille de lois de Bernoulli de paramètre p dans $[0, 1]$). Cet inconnu certain est toujours inconnu après l'expérience et on cherche à le déterminer à partir du résultat de l'expérience.

Qu'est-ce qui aurait changé si on avait observé quatre lancers seulement ?

S est toujours le nombre de pile obtenus et l'écart est $|S - 2|$, considéré anormalement grand s'il est supérieur à 2.

L'erreur consistant à décider que la pièce n'est pas équilibrée alors qu'elle l'est a pour probabilité

$$\frac{1}{2^4} (C_4^0 + C_4^4) = 0.125 ;$$

L'erreur consistant à décider que la pièce est équilibrée alors qu'elle a la probabilité $\frac{3}{4}$ de tomber sur pile a pour probabilité

$$C_4^1 \left(\frac{3}{4}\right)^1 \left(\frac{1}{4}\right)^3 + C_4^2 \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^2 + C_4^3 \left(\frac{3}{4}\right)^3 \left(\frac{1}{4}\right)^1 = 0.68$$

Qu'est-ce qui aurait changé si on avait observé seize lancers ?

S est toujours le nombre de pile obtenus et l'écart est $|S - 8|$, considéré anormalement grand s'il est supérieur ou égal à 5.

Le premier risque vaut alors

$$\frac{1}{2^{16}} (C_{16}^0 + C_{16}^1 + C_{16}^2 + C_{16}^3 + C_{16}^{13} + C_{16}^{14} + C_{16}^{15} + C_{16}^{16}) = 0.021 ;$$

Le deuxième risque, si la probabilité de tomber sur pile est $\frac{3}{4}$, vaut

$$\sum_{k=4}^{12} C_{16}^k \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{16-k} = 0.595.$$

On peut remarquer qu'avec les règles de décision choisies les risques de première et de deuxième espèce sont plus petits quand on a observé un plus grand nombre de lancers et que le risque de deuxième espèce est grand par rapport au risque de première espèce.

Que devient cette démarche dans les exemples 1 et 2?

Les étapes a), b) et c) sont déjà précisées. Pour l'étape d), on peut proposer comme écart la valeur absolue de la différence entre la moyenne empirique des observations (si on fait n mesures) et la moyenne théorique attendue, qui est *Norme* ou m_0 . L'utilisateur, connaissant bien le champ d'application, peut donner des valeurs extrêmes liées à sa pratique ; sinon il faut réaliser le point g) pour les définir. Pour cette étape g), il est nécessaire de faire des hypothèses supplémentaires sur la loi de la variable observée pour aller plus loin dans les calculs !

II Exemple du graphologue

La formulation de cet exercice est celle qui est habituelle dans les cours de statistique. Où retrouve-t-on les points mis en évidence dans la partie précédente ?

Pour embaucher un graphologue, le chef du personnel d'une grosse entreprise envisage de faire un test. Il propose à chaque candidat d'identifier l'écriture d'un médecin et celle d'un avocat présentées par paire et cela pour 12 paires d'écritures. Le candidat sera embauché s'il identifie correctement l'écriture du médecin et celle de l'avocat pour au moins 9 paires.

1) Déterminer la probabilité d'embaucher un incompetent en supposant que c'est un candidat qui répond au hasard. Cette hypothèse se traduit quantitativement par la phrase : la probabilité qu'il reconnaisse correctement une paire d'écritures donnée est $p = \frac{1}{2}$.

2) Pensant que cette probabilité est trop forte, le chef du personnel change son critère et exige 10 bonnes réponses pour pouvoir embaucher un candidat.

Déterminer maintenant la probabilité d'embaucher un incompetent.

3) Cette probabilité est trop faible. Le chef du personnel envisage donc de proposer 13 paires d'écritures et non plus 12.

Déterminer dans cette nouvelle situation la nombre minimal de bonnes réponses à exiger pour que la probabilité d'embaucher un incompetent soit inférieure à 0.05.

4) Après s'être placé du point de vue du chef de personnel qui embauche, plaçons-nous du point de vue du chômeur qui cherche à être embauché : cette personne estime savoir reconnaître une paire d'écritures dans 85% des cas. Il sait qu'il sera embauché s'il fournit au moins 10 bonnes réponses sur les 13 paires proposées.

Déterminer la probabilité qu'il ne soit pas embauché, c'est à dire qu'il fournisse moins de 10 bonnes réponses.

III Un peu de formalisation et de vocabulaire, sur des savoirs cachés de Terminale

Un test de validité d'hypothèse est un mécanisme permettant de confirmer ou d'infirmer une hypothèse par l'observation d'un échantillon. C'est une fonction définie de l'ensemble des valeurs possibles de l'observation vers l'ensemble des décisions qui sont $d_0 =$ "accepter l'hypothèse" et $d_1 =$ "refuser l'hypothèse". C'est une variable aléatoire.

1) Formulation des hypothèses :

- L'hypothèse testée, ou hypothèse nulle (notée H_0) est celle que l'observateur croit vraie et qu'il ne rejettera que si elle est contradictoire avec les résultats de

l'expérience. En général, elle se traduit par une égalité ou des inégalités portant sur un paramètre du modèle de la situation étudiée.

Elle correspond à une hypothèse de prudence . Par exemples :

pour le chef de personnel, embaucher quelqu'un qui répond au hasard n'est pas pertinent et cela ne doit pas arriver ou alors extrêmement rarement : H_0 correspond donc à "le candidat est incompetent" ;

pour un chef de production, affirmer que sa production est mauvaise si elle est bonne le conduit à engager des frais inutiles : H_0 correspond donc à "la chaîne de montage est bien réglée" ;

pour le joueur qui s'engage dans des paris, affirmer que la pièce a plus de chance de tomber sur pile que sur face est très risqué et peut le conduire à la ruine : H_0 correspond donc à "la pièce est équilibrée",...

- L'hypothèse alternative (notée H_1) est une hypothèse estimée a priori peu probable mais possible. On parle d'alternative simple si elle s'exprime par une égalité sur un paramètre, hypothèse multiple sinon.

Quand H_1 n'est pas le contraire de H_0 , le test est dit non contradictoire.

On parle de test unilatéral si H_0 est de la forme $\theta \leq \theta_0$ et H_1 de la forme $\theta > \theta_0$, de test bilatéral si H_0 est de la forme $\theta = \theta_0$ et H_1 de la forme $\theta \neq \theta_0$.

2) Statistique du test :

C'est une variable aléatoire (fonction de l'observation ou de l'échantillon observé) à partir de laquelle est définie la règle de décision. Cela correspond à l'écart de la partie I, la règle de décision étant d'accepter H_0 si l'écart est inférieur à un seuil et de la rejeter sinon.

Dans l'exemple 3 traité, c'est la variable S , nombre de pile observé dans la série de 8 lancers.

Dès qu'on connaît le paramètre du modèle, on connaît la loi de cette statistique : dans l'exemple, S suit la loi binomiale de paramètres 8 et p si p est la probabilité qu'a la pièce de tomber sur pile.

3) Règle de décision, région de rejet de H_0 (ou région critique du test)

Une règle de décision associée à une statistique de test T consiste à rejeter H_0 si T est supérieure à un seuil s : l'ensemble des échantillons tels que T vérifie cette condition est appelée région de rejet ou région critique.

Dans l'exemple du graphologue, H_0 est " $p = \frac{1}{2}$ " où p représente la probabilité du candidat à reconnaître une paire d'écritures et où cette valeur $\frac{1}{2}$ correspond à une incompétence ; la statistique du test est le nombre N de paires reconnues et la région de rejet est l'ensemble des observations telles que N soit supérieur ou égal à 9.

4) Deux types d'erreur :

Toute décision peut être erronée :

- prendre la décision d_1 , c'est à dire rejeter H_0 alors que H_0 est vraie : c'est l'erreur de première espèce

- prendre la décision d_0 , c'est à dire accepter H_0 alors que H_1 est vraie : c'est l'erreur de deuxième espèce.

5) Risques et seuil de confiance :

Ces erreurs sont aléatoires et on veut mesurer la probabilité qu'elles arrivent. Ce sont les risques d'erreurs :

- risque de première espèce, qui mesure la probabilité de commettre l'erreur de première espèce, et qui vaut dans le cas où H_0 est une hypothèse simple $E_{H_0}(d_1)$, ou dans l'exemple du graphologue $P_{0.5}(N \geq 9)$. On appelle seuil de confiance ou niveau du test la valeur de ce risque.

- risque de deuxième espèce, qui mesure la probabilité de commettre l'erreur de deuxième espèce. Dans le cas du graphologue et du candidat qui a une probabilité 0.85 de reconnaître les écritures, accepter H_0 alors que H_0 est fausse a pour probabilité $P_{0.85}(N < 9)$.

6) Détermination de la région de rejet à partir d'un seuil de confiance fixé :

L'idéal est de minimiser ces deux risques simultanément, ce qui est impossible. On choisit donc de contrôler le risque de première espèce. On se donne une valeur α "petite".

On décide de construire une règle de décision à partir d'un écart T de la façon suivante : si T est supérieur à s_α , on rejette H_0 . Le seuil s_α est déterminé par la contrainte

$$P_{H_0}(T > s_\alpha) = \alpha.$$

Après l'expérience, on calcule la valeur de T , soit T_{obs} et on la compare à ce seuil s_α : si T_{obs} est supérieur à s_α , on rejette l'hypothèse et si T_{obs} est inférieur à s_α , on l'accepte.

Le niveau du test α , comme la statistique T , est au choix de l'utilisateur. Le problème du statisticien est alors de proposer des règles de décision donnant un risque de seconde espèce le moins élevé possible pour un risque de première espèce fixé.

7) Tableau récapitulatif des risques :

Le risque de première espèce est noté α , le risque de deuxième espèce est noté $1 - \beta$. On cherchera, le premier fixé, à minimiser le second ou encore à maximiser son complément à 1 (qui représente la probabilité de rejeter H_0 quand elle est fausse), soit maximiser β : ce nombre s'appelle la puissance du test.

	H_0 est vraie	H_0 est fausse
acceptation de H_0	$1 - \alpha$	β
rejet de H_0	α	$1 - \beta$

IV Tests d'adéquation, tests du χ^2

1) Problématique spécifique de l'adéquation :

Face à une situation expérimentale où les résultats issus de répétitions indépendantes d'une même épreuve sont aléatoires, on pose la question suivante : les observations faites sont-elles compatibles avec l'hypothèse que la loi régissant l'épreuve vérifie une caractéristique donnée.

Par exemple, dans le cas du joueur et de la pièce, l'observation de 8 lancers, donnant des fréquences empiriques d'obtention de pile et de face, correspond-elle à la probabilité théorique d'obtention de pile et de face d'une pièce équilibrée ?

Exemple encore plus pratique : il y a eu 1 pile et 7 faces sur 8 lancers ; peut-on dire qu'avec cette pièce il y a une chance sur deux d'avoir pile et une chance sur deux d'avoir face ? Il faut comparer les deux lois sur $\{0, 1\}$: la loi empirique $\{\frac{7}{8}, \frac{1}{8}\}$ et la loi théorique $\{\frac{1}{2}, \frac{1}{2}\}$.

Dans la liste des étapes mises en évidence en partie I, les points a,b et c sont déjà mis en place et en particulier la modélisation probabiliste est faite implicitement.

Il faut par contre choisir un écart : on choisit classiquement la distance dite du χ^2 , que j'expliciterai un peu plus loin de manière générale.

Sur l'exemple de la pièce, cela donne l'écart suivant : $T = \frac{(S-4)^2}{4} + \frac{(8-S-4)^2}{4}$, qui, en gros, somme les carrés des écarts entre les effectifs observés et les effectifs théoriques si l'hypothèse est vraie.

Il faut ensuite décider quand les valeurs de cet écart sont grandes. Soit on le décide arbitrairement, soit on essaie de limiter le risque de première espèce.

On remarque que $T = \frac{(S-4)^2}{2}$ est le carré de l'écart choisi en partie I.

Tester l'adéquation de la loi observée à la loi $\{\frac{1}{2}, \frac{1}{2}\}$ ou tester l'égalité à $\frac{1}{2}$ du paramètre du modèle choisi en I revient au même. C'est un changement de vocabulaire seulement.

2) Exemple du dé pipé ou non :

Un joueur utilise un dé et se demande s'il est pipé ou non. Pour répondre à cette question, il fait un grand nombre de lancers et au vu des résultats obtenus, il doit prendre une décision : le dé est pipé ou le dé est non pipé.

Suivons la démarche vue en I :

-l'étape a) est dite : "le dé est-il équilibré ?".

-l'étape b) consiste à modéliser le résultat d'un lancer de dé par une variable aléatoire prenant ses valeurs dans $\{1, 2, \dots, 6\}$ et de loi (p_1, p_2, \dots, p_6) inconnue (le paramètre (p_1, p_2, \dots, p_6) est un vecteur de \mathbf{R}^6 tel que les coordonnées sont positives de somme 1).

-l'étape c) consiste à interpréter le fait que le dé soit non pipé en l'égalité à $(\frac{1}{6}, \frac{1}{6}, \dots, \frac{1}{6})$ du paramètre.

- l'étape d) utilise la répartition empirique observée après n lancers de ce dé. Si on note n_j le nombre de fois où la face j est sortie, cette répartition empirique est $(\frac{n_1}{n}, \frac{n_2}{n}, \dots, \frac{n_6}{n}) = (f_1, f_2, \dots, f_6)$.

On choisit comme écart :

$$ecart_n = \sum_{j=1}^6 \frac{(n_j - \frac{1}{6}n)^2}{\frac{1}{6}n} = 6n \sum_{j=1}^6 (f_j - \frac{1}{6})^2$$

Reste à préciser quand cet écart est grand : on décide de lier ce seuil au risque de première espèce et donc de le déterminer à l'aide de la loi de l'écart quand

l'hypothèse est vraie. Pour connaître cette loi, soit on procède comme ce qui est proposé en terminale, soit on connaît un peu plus de statistique théorique.

3) Ce que la théorie dit :

C'est compliqué et utilise la notion de convergence en loi (un peu comme le théorème central limite qui affirme que la suite de variables aléatoires $Y_n = \sqrt{n} \frac{\bar{X}_n - m}{\sigma}$ converge vers une variable aléatoire normale centrée réduite) :

La variable $ecart_n$ dépend du nombre n de lancers de dés faits. La suite de variables aléatoires $ecart_n$ converge en loi vers une variable aléatoire dont la loi est la loi du χ^2 à 5 degrés de liberté dans le cas du dé. C'est à dire que dans de bonnes conditions, on peut approcher la loi de $ecart_n$ par celle du χ^2 à 5 degrés de liberté, loi qui est tabulée. La lecture de la table permet de déterminer le seuil s_α tel que la probabilité que l'écart dépasse ce seuil soit α . Par exemple, si $\alpha = 10\%$, on lit $s_{0,10} = 9.236$ et si $\alpha = 5\%$, $s_{0,05} = 11.070$.

4) Ce qu'on propose en terminale :

Ce qui précède est hors de portée des élèves de terminale.

Par contre, ils savent utiliser les méthodes de simulation.

Le programme propose donc de remplacer la loi théorique de l'écart par la loi empirique obtenue à partir de la simulation d'un grand nombre N d'écarts. Chaque écart est calculé à partir de la simulation de n lancers de dés équilibrés.

Cette dernière série de N valeurs observées a une distribution de fréquence dont on peut caractériser les paramètres de tendance centrale et de dispersion, en particulier les quantiles. On peut déterminer ainsi le nombre réel $s_{0,10}$, 9-ème décile tel que 90% de la série soit inférieure à ce seuil $s_{0,10}$ ou encore le réel $s_{0,05}$, 95-ème centile, tel que 95% de la série soit inférieure à ce seuil.

Et on poursuit le raisonnement : si le dé du joueur est équilibré, la valeur de l'écart peut être considérée comme un élément supplémentaire à la série simulée et il y a une probabilité de 90% qu'elle soit inférieure au seuil $s_{0,10}$ que l'on vient de déterminer.

La règle de décision que l'on adoptera sera la suivante : si la valeur observée de l'écart est supérieure au seuil $s_{0,10}$, on affirme que le dé est pipé ; sinon, on accepte le fait qu'il ne le soit pas.

Ces deux décisions peuvent être entachées d'erreur : la probabilité d'affirmer que le dé est pipé alors qu'il ne l'est pas est de 10% si le seuil est $s_{0,10}$, 5% si le seuil est $s_{0,05}$. C'est le risque de première espèce. Le risque de deuxième espèce n'est pas calculé.

On ne parle pas clairement en terminale des risques pris quand on agit de la sorte.

5) Exemple du test d'indépendance (adéquation avec estimation de paramètres):

A partir d'un exemple : lors d'une enquête sociologique portant sur les étudiants de l'université XXX, on se demande s'il existe un lien entre le mode de logement et le sexe d'un étudiant. On veut tester l'indépendance entre ces deux caractères décrivant un étudiant.

L'enquête n'est pas exhaustive ; on considère d'une part que chaque étudiant interrogé fournit une observation du couple "(mode de logement,sexe)" = (X, Y) , observation au résultat aléatoire et d'autre part que ces étudiants ont des comportements indépendants entre eux. La loi de probabilité de ce couple est inconnue et on traduit la question posée dans cette modélisation probabiliste : la loi du couple (X, Y) est-elle le produit des lois marginales de X et de Y ? c'est cela l'hypothèse H_0 .

ou encore, y-a-t-il adéquation entre la répartition empirique observée du couple (X, Y) et la répartition théorique s'il y avait indépendance ?

L'idée est d'utiliser la démarche du paragraphe précédent. Mais se pose le problème du calcul des effectifs théoriques, puisqu'on ne connaît pas les lois de X et de Y sous l'hypothèse nulle.

On va alors les estimer par les lois marginales empiriques et choisir comme écart entre la distribution observée du couple (X, Y) et sa distribution théorique sous H_0 la variable suivante :

$$T = \sum_{i \in \{1,2\}, j \in \{1,..,4\}} \frac{(n_{ij} - n \frac{n_{i.}}{n} \frac{n_{.j}}{n})^2}{n \frac{n_{i.}}{n} \frac{n_{.j}}{n}}$$

où sur les n étudiants interrogés, n_{ij} est le nombre d'étudiants ayant répondu la modalité i à la variable sexe et la modalité j à la variable "mode de logement", $n_{i.}$ le nombre d'étudiants ayant répondu la modalité i à "sexe" et $n_{.j}$ celui des étudiants ayant répondu j à "mode de logement".

On montre que, si l'hypothèse H_0 est vraie, cette variable aléatoire T converge en loi quand n tend vers l'infini vers une variable aléatoire de loi du χ^2 à $(2-1)(4-1) = 3$ degrés de liberté. On peut alors définir une règle de décision dont le risque de première espèce ait une valeur donnée.

Exemple numérique :

Les modes de logement repérés sont : seul, en famille, en couple, autres et le sexe a pour modalité : masculin, féminin.

	seul	en famille	en couple	autres	tout mode
féminin	12	11	14	12	49
masculin	15	6	9	14	44
tout sexe	27	17	23	26	93

Tableau des effectifs théoriques estimés sous l'hypothèse d'indépendance:

	seul	en famille	en couple	autres	tout mode
féminin	14.2	9	12.1	13.7	49
masculin	12.8	8	10.9	12.3	44
tout sexe	27	17	23	26	93

Calcul de l'écart :

$$T_{obs} = \frac{2.2^2}{14.2} + \frac{2^2}{9} + \frac{1.9^2}{12.1} + \frac{1.7^2}{13.7} + \frac{2.2^2}{12.8} + \frac{2^2}{8} + \frac{1.9^2}{10.9} + \frac{1.7^2}{12.3}$$

On trouve $T_{obs} = 2.74$.

D'après la théorie, ce T suit la loi du χ^2 à 3 ddl, la probabilité que T dépasse la valeur 7.8 est de 5%, celle qu'il dépasse 6.2 est de 10%.

T_{obs} est inférieur à ces deux seuils. L'observation permet de prendre la décision suivante : le mode de logement et le sexe sont indépendants, avec le risque de moins de 10% de se tromper.

Conclusion

Mise en place d'une procédure pour prendre une décision dans une situation aléatoire.

- a) identification de la chose certaine inconnue, sur laquelle on pose une question.
- b) modélisation de la situation par un modèle probabiliste connu mais de paramètres inconnus.
- c) traduction de la question posée en a) dans le modèle choisi.
- d) choix d'un écart entre ce qui est observé aléatoire et ce qui est attendu sous l'hypothèse.
- e) choix des valeurs extrêmes de cet écart.
- f) calcul de l'écart pour l'expérience faite et prise de décision.
- g) calcul de la loi de probabilité de l'écart, évaluation des risques d'erreur des décisions prises et retour sur le choix des valeurs extrêmes.

Il y a une relation entre niveau de risque accepté ou acceptable et grande valeur de l'écart : c'est le praticien (chef du personnel, chef de production, joueur, sociologue, décideur en général...) qui prend la décision et choisit en fin de compte le risque acceptable.

-II- MODALITES GLOBALES POUR LES FORMATIONS ET EXEMPLES D'ACTIVITES.

1) Des modalités globales

Nous avons proposé des hypothèses générales sur les formations d'enseignants, inspirées de nos recherches sur les pratiques³. Nous précisons ici ce que nous avons retenu et la manière particulière de décliner ces hypothèses compte tenu du contenu mathématique particulier.

a) Une condition nécessaire plus ou moins respectée : la durée

Nous essayons de construire des formations longues (un semestre). Cependant les formations continues sont de plus en plus réduites ce qui nous semble très dommageable.

b) Un travail spécifique sur les contenus

Nous avons deux objectifs : le premier est de faire expliciter les implicites mathématiques qui émaillent les contenus à enseigner – ce que nous appelons les savoirs cachés (même si cela ne change rien à ce qui sera enseigné aux élèves) ; cela donne souvent lieu à des cours, si possible préparés par des activités des participants qui amènent des questions sur ces savoirs cachés - comme des analyses d'exercices. Puis nous essayons de faire débusquer les glissements qui peuvent intervenir dans les tâches proposées aux élèves – là encore ce n'est pas pour nécessairement modifier ce qui est proposé mais pour faire prendre conscience des activités des élèves.

c) Pour les participants, des activités réelles, proches de l'expérience et des besoins

Les activités proposées aux participants doivent être des activités réelles, où ils peuvent s'investir et apprendre quelque chose de nouveau. Nous donnons dans le paragraphe 2 des exemples précis. Cependant on peut noter que c'est le formateur qui décide de ces activités qui restent relativement isolées les unes des autres, dans la mesure où ce sont les savoirs statistiques à enseigner au lycée qui servent de fil directeur au scénario.

³ Cf. Robert Aline, 2005

d) L'importance du collectif

En formation de formateurs, la plupart des activités sont collectives (en petits groupes ou en grand groupe). Au PAF, le manque de temps ne permet pas de mettre en place réellement des formes de travail collectives, ce que nous trouvons dommage.

e) Le déroulement de séances de statistique

Une séance est consacrée à quelques exemples analysés dans des classes (cf. Carranza, 2004). Cependant il manque encore à nos formations des études plus nombreuses de déroulements de séances effectives (par exemple sur vidéo). C'est une manière efficace pour mettre en évidence les contraintes, les marges de manœuvre et les alternatives envisageables dans ce type de séances, et cela manque à nos formations.

2) Des activités choisies

a) Activités des stagiaires à partir de l'étude de manuels et de documents sur les performances des élèves

i) Les analyses d'exercices ou activités des manuels : pour diagnostiquer

On propose aux stagiaires en séance d'analyser quelques énoncés choisis à partir de différents manuels. Ils ont utilisé par ailleurs une grille jointe en annexe⁴.

Plusieurs types de contenus peuvent être abordés ainsi que plusieurs paragraphes des manuels.

En statistique descriptive, il s'agit d'analyser des activités attendues et possibles des élèves dans les exercices et de mettre en évidence les glissements.

En hasard, il s'agit de découvrir les mathématiques cachées à partir de TP ou d'activités.

Ces activités sont complétées par quelques analyses plus systématiques (faites collectivement en dehors des séances) pour généraliser ou non les résultats observés sur tout un chapitre d'un manuel ou différents manuels.

Un bilan est effectué à la séance suivante.

ii) Du côté des élèves : pour mieux connaître les performances

Les résultats des analyses EVAPM ont été travaillés (en dehors des séances) et permettent de confirmer ou non les tendances déjà remarquées.

⁴ Voir annexe page 16.

iii) Travail de réécriture d'exercices ou d'activités

A partir d'un énoncé de manuel, il s'agit de proposer un autre énoncé, corrigeant les difficultés précédentes. Cette activité n'a pas été souvent proposée.

b) Travail sur les ressources

i) Un travail préalable de lecture en séance

A partir d'un petit article choisi sans difficulté de contenu particulière mais pouvant susciter des réactions diverses, on organise une lecture individuelle de 20 minutes suivi d'un temps de discussion. Une grille de lecture est distribuée⁵.

Un stagiaire résume et critique et une discussion s'organise ; le formateur demande que chaque intervenant précise le statut de ce qu'il dit.

Cela permet de s'approprier la grille et de prendre conscience de la diversité des points de vue qui apparaissent dans une discussion sur un article.

ii) Un travail de lecture hors séance avec une présentation résumée en séance

Les articles sont choisis pour leurs apports en termes de contenus.

Une lecture a lieu et un travail en petits groupes permet d'élaborer un compte rendu.

Les présentations permettent de mettre en fonctionnement les catégories critiques déjà utilisées précédemment.

iii) La recherche bibliographique sur un thème

Enfin, on demande aux stagiaires d'établir une bibliographie sur un contenu donné.

Toutefois le formateur partage les ressources à explorer entre les différents petits groupes (par exemple les années 2000-2001 des bulletins de l'APMEP et repères IREM).

Chaque article doit apparaître avec son titre, sa place et un résumé de 5 lignes.

Une présentation rapide est organisée en séance.

c) Des activités prévues mais non réalisées : observation et analyses de séances en classe

Plusieurs modalités sont envisageables : des analyses de vidéo de séances en classe peuvent être organisées. Ou bien on peut présenter des résultats d'analyses de telles séances.

⁵ Voir annexe page 15.

d) les apports mathématiques et complémentaires

Sur des points clefs dégagés soit en analyse d'exercices soit en biblio, on organise des exposés permettant une explication et une présentation détaillée des savoirs mathématiques concernés.

Les apports complémentaires peuvent être culturels, historiques, épistémologiques, didactiques, ou encore des statistiques « vécues »

e) Conception de documents de formation, construction de scénario, animations

Cette modalité nous semble importante dans une formation de formateurs même si elle est courte, et virtuelle.

On peut concevoir plusieurs types d'activités (ils seront illustrés dans la partie -III- Exemples de scénarios en formation de formateurs : quatre chroniques), par exemple :

- préparer un exercice bien choisi et le faire travailler pendant une séance en se répartissant les petits groupes de stagiaires (voir par exemple, la séance 1 page 37 et la séance 13 page 39 ou bien le déroulement du stage PAF page 44)

- concevoir un scénario sur un thème précis (voir page 36)

- animer une séance de formation (voir pages 39 et 43)

- élucider une question délicate et produire un document d'éclaircissement (voir page 36)

-III- EXEMPLES DE SCENARIOS EN FORMATION DE FORMATEURS : QUATRE CHRONIQUES

Nous décrivons ci-dessous chronologiquement 4 formations effectuées pendant les années 2003-2004 et 2004-2005.

Trois des formations ont été effectuées dans le cadre de l'option "Statistique et probabilité" du parcours professionnel "formation de formateur" du Master de l'université Paris 7 :

- Année 2003-2004 (DU2)
- Année 2004-2005, Groupe "Statistique descriptive"
- Année 2004-2005, Groupe "hasard"

La quatrième formation décrite a été proposée par l'IREM de Paris 7 dans le cadre des plans académiques de formation des académies de Paris, Créteil et Versailles en 2004-2005.

Notes :

- dans les descriptifs ci-dessous les séances dont le titre est en **gras** ("*séance 7*" par exemple) sont des séances en présence du formateur, les autres séances sont des séances pour lesquelles les étudiants se retrouvent pour travailler en groupe dans la salle habituelle ou en bibliothèque ("*séance 7*" par exemple).

- Afin de ne pas confondre les différentes chroniques qui se succèdent ci-dessous des "logos" différents ont été utilisés pour lister les séances.

1) Master Professionnel, Option statistique/Probabilité, 2003-2004

1ère partie : statistique, apports de contenu

✓ **Séance 1 et 2** (19/9/3 et 3/10/3)

Cours de statistique reprenant en les détaillant les contenus des programmes du secondaire (intervenant extérieur).

✓ **Séance 3** (15/9/3)

Exposé et débat sur l'histoire de la statistique et de son enseignement en France (intervenant extérieur).

2ème partie : Ouvertures, pistes de réflexion

✓ Séance 4 (14/11/3)

Intervention à propos du concours « À vos stats », réflexion sur les nouveaux programmes de lycée.

Travail sur trois affiches du concours « À vos stats » pour répondre à la question : quel travail statistique a été faites ? Comment aller plus loin compte tenu du niveau scolaire ?

✓ Séance 5 (21/11/3)

Compte rendu d'une enquête effectuée auprès d'enseignants de l'académie de Lyon en 2000 à propos de l'enseignement de la statistique en seconde.

✓ Séance 6 (5/12/3)

Bilan des quatre premières séances.

3ème partie : réflexion sur la formation, production de documents

(suite de la séance 6)

Présentation des travaux possibles, essentiellement :

- élaboration de documents de formation
- proposition de scénario de formation

Trois projets sont retenus :

- Pourquoi et comment faire des simulations en classe de seconde ? (scénario)
- Mode, moyenne, médiane au collège. (scénario)
- Le problème de l'adéquation à une loi équirépartie en terminale et son lien avec le test du Khi2. (document)

✓ Séances 7 à 11 (16/1/4, 6/2/4, 12/3/4, 2/4/4, 14/5/4)

Séances de travail en groupe. Le travail est encadré, surtout lors des premières séances.

✓ Séance 12 (28/5/4)

Présentation des différents dossiers (une édition papier est disponible, ainsi qu'une mise en ligne comme documents de travail).

Documents proposés en annexe⁶ :

Présentation des trois dossiers réalisés.

⁶ Page 18.

2) Master Professionnel, Option statistique/Probabilité, 2004-2005, Groupe "Statistique descriptive"

1^{ère} partie : Statistique descriptive dans les manuels et dans la classe

➤ Séance 1 (30/9/4)

Analyse d'exercices de manuels issus des chapitres de statistique descriptive de classe de seconde (ou de la partie des chapitres de statistique concernant la statistique descriptive le cas échéant). Une sélection d'exercices est proposée⁷. Ils ont été regroupés par thèmes. Les thèmes des pages ne sont pas annoncés, mais le travail est fait sur chaque page séparément :

- première page : utilisation de la géométrie en statistique (exemple typique : la "moyenne" de trois triangles rectangles)
- pages 2, 3, 4 : calculs de moyennes, d'écart types, constructions de graphes, utilisation de la linéarité de la moyenne, éventuellement avec un contexte transparent (glissement numérique, algébrique)
- page 5 : utilisation de l'algèbre en statistique (exemple typique : trouver une note manquante dans une série dont on connaît la moyenne)

Le travail se fait en groupe. Sur chaque page, chaque groupe choisit un ou deux exercices et en fait une analyse a priori. Un tour de table est ensuite fait, pour chaque page, chaque groupe complétant ce qui a été déjà dit.

"Ce n'est pas ça les statistiques" n'apparaît pas tout de suite et n'est pas d'emblée accepté par tous. La discussion s'engage, le terme de "glissements" est proposé par les enseignants.

Un travail sur la médiane, prévu, n'a pas été fait faute de temps. Il aurait permis d'insister sur deux difficultés liées aux mathématiques dans le travail de statistique descriptive : les notions de fonctions (et en l'occurrence de fonction réciproque) et la notion d'ordre.

Mise en place du travail de la semaine suivante : travail en groupe sur différents énoncés (les énoncés sont répartis équitablement).

- 8 exercices traitant de statistique issus de l'évaluation EVAPM⁸ (avec le codage des réponses et les résultats observés, item par item)

⁷ Cf. annexe page 23.

⁸ Cf. annexe page 23.

- 7 exercices (mot clef statistique) de la base EVAPMib⁹ (avec les résultats de différentes sessions et des commentaires trouvés sur le site web).

➤ *Séance 2* (7/10/4)

Travail en groupe sur les différents énoncés.

➤ *Séance 3* (14/10/4)

Chaque groupe présente rapidement par oral ses conclusions.

Intervention de Pablo Carranza : "Glissements dans l'enseignement de la statistique en classe de seconde"

➤ *Séance 4* (21/10/4)

Travail en groupe sur un chapitre de manuel (statistique descriptive, au choix des stagiaires : 3^{ème}, 2^{nde} ou 1^{ère}), analyse plus macroscopique des éventuels glissements.

➤ *Séance 5* (4/11/4)

Bilan du travail sur les chapitres.

2^{ème} partie : mathématiques cachées

(suite de la séance 5)

Intervention de Jacqueline Mac Aleese sur les différents indicateurs¹⁰.

3^{ème} partie : travail bibliographique

(suite de la séance 5)

Introduction au travail sur la biblio.

Lecture d'un petit article (*Les statistiques : de leur utilisation dans le domaine public en remontant vers leurs sources, exemple tiré du débat national sur l'école*, René Mulet-Marquis, repère IREM n°55, 2004). Réactions libres et retour sur ces réactions par les enseignants (classification des réactions, différents points de vue présents), la "grille" de lecture mise en place dans l'enseignement de l'année précédente est redistribuée¹¹.

Distribution d'articles à lire et répartition :

- *Moyenne, médiane, écart type. Quelques regards sur l'histoire pour éclairer l'enseignement des statistiques*, Anne Boyé, Marie-Claire Comairas, Repère IREM n°48, juillet 2002.

⁹ Cf. annexe page 23.

¹⁰ Voir page 13.

¹¹ Annexe page 15.

- *A bas la moyenne, à propos des paramètres de tendance centrale et de dispersion d'une série statistique*, Jean Claude Girard, repère IREM n°33, octobre 1999.

- *Mode, moyenne, médiane au collège : pourquoi ?*, Jean Claude Duperret, Repère IREM n°6, janvier 1992

➤ *Séance 6* (18/11/4)

Travail en groupe. Lecture d'un article, préparation du compte rendu.

➤ *Séance 7* (25/11/4)

Bilan du travail sur les articles. Chaque groupe résume puis commente l'article lu (en spécifiant les différents points de vue pris).

Présentation des séances de formation du mois de janvier. Proposition d'organisation des journées et de répartition du travail d'ici là.

Lancement du travail de recherche bibliographique (séance suivante). Répartition de revues et d'années par groupe.

➤ *Séance 8* (2/12/4)

Travail en groupe. Recherche bibliographique ciblée.

➤ *Séance 9* (9/12/4)

Bilan de la recherche bibliographique (½ séance)

4^{ème} partie : deux séances de formation

(suite de la séance 9)

Travail en groupe sur les scénarios de janvier.

➤ *Séances 10 et 11* (16/12/4 et 6/1/5)

Travail en trois groupe sur la séance du 27 janvier

➤ *Séance 12* (13/1/5)

Séance organisée par l'autre groupe du dispositif (notion de hasard).

➤ *Séance 13* (27/1/5)

Séance présentée par les étudiants à l'autre groupe du dispositif (qui a travaillé essentiellement sur la notion de hasard) :

- Présentation d'une bibliographie (organisée en trois rubriques : information générale, collège et lycée).

- analyse d'exercices (6 pages d'exercices distribuées)

Réflexion :

- un exemple de scénario sur site traitant de statistique descriptive.

Certains documents sont proposés en annexes¹². L'ensemble des documents est disponible sur le site du Master Professionnel.

3) Master Professionnel, Option statistique/Probabilité, 2004-2005, Groupe "hasard"

1^{ère} partie : Des exercices aux savoirs, mise en évidence de la transposition didactique

❖ Séance 1 (23/9/4)

Analyse d'exercices dans le but de :

- mettre en évidence les savoirs savants, les savoirs cachés
- tracer la frontière entre expériences et simulation
- distinguer les situations pour lesquelles une loi est connue des autres situations

Le travail se base sur des "activités" ou des "travaux pratiques" de manuels (peu d'exercices abordent ce thème, certains manuels ne proposent que des travaux pratiques, les énoncés naviguant entre instructions aux élèves et propositions de gestion de la classe pour l'enseignant).

Le travail se fait, dans un premier temps, en groupe. Il est demandé de faire une analyse a priori des énoncés et des activités des élèves. Ce type d'analyse est connu des étudiants : ils l'ont pratiqué lors de l'enseignement "DU1". Un des enjeux de la séance est donc aussi de faire le lien avec le travail déjà effectué. Les analyses sont cependant décalées du fait de la nature des notions abordées et du travail demandé.

Nous demandons donc aux étudiants une analyse permettant de dégager les savoirs mathématiques cachés, et d'essayer de distinguer, de ce point de vue, les différents type de situation (cf. objectifs ci-dessus).

TD n°10 - Déclic 2000 - Pile ou face, évolution des fréquences

Activité 2 - Bréal 2000 - Tirages dans une urne de composition connue

TD n°7 - Déclic 2000 - Fluctuation d'échantillonnage

Modalité :

Analyse en groupe d'une activité après l'autre.

Mise en commun

¹² Annexe page 59.

Conclusion par l'enseignant

La conclusion de la séance fait le lien entre les analyses développées par les étudiants et les objectifs de la séance.

❖ *Séance 2* (30/9/4)

Travail en groupe

Enoncé : faire le même travail sur un panel d'exercice plus large (les exercices d'un chapitre de manuel).

Production : préparation d'un petit compte rendu oral

❖ *Séance 3* (7/10/4)

Bilan des analyses d'exercices en groupe. De nombreuses questions sont dégagées¹³.
Durée 1h30.

Chaque groupe fait un compte rendu oral.

Un bilan est fait par les enseignants (reprise des objectifs du 23 septembre).

2^{ème} partie : Approches du cadre mathématique

Cours de mathématiques : Fluctuations, fourchettes de sondage¹⁴

1. Modélisation probabiliste de sondages dans le cas d'un lancer de pièce
2. Liens probabilité/statistique
3. travail sur des exercices de manuel (travail à faire)

Document distribué en fin de séance : Article "référendum" pour info

Travail à faire : faire et analyser trois exercices de Déclic 1S 2001 et Déclic 2nde 2000

❖ *Séance 4* (14/10/4)

Travail en groupe sur les trois exercices distribués

❖ *Séance 5* (21/10/4)

Intervention de Bernard Parzysz¹⁵ : Générateurs de nombres aléatoires (voir aussi les actes des journées APMEP d'Orléans, voir aussi le livre). Ce thème a été fortement évoqué dans bilan des retours sur les analyses d'exercices (le 7 octobre).

Discussion sur les trois exercices de manuels sur fourchettes (cf. séances 3 et 4).

¹³ Annexe page 66.

¹⁴ Voir page 15.

¹⁵ Voir Parzysz 2005

Complément de cours :

- symétrie de la relation entre :

$$f_n \in [p-a ; p+a] \text{ (} p \text{ connu et } f_n \text{ attendu)}$$

et

$$p \in [f_n-a ; f_n+a] \text{ (} p \text{ inconnu et } f_n \text{ observé)}$$

- influence (ou non) de la taille de la population

3^{ème} partie : travail bibliographique

Introduction à la lecture d'articles (rappel à propos de la fiche de lecture étudiée l'année précédente¹⁶).

Un petit article d'exemple était prévu mais il ne reste que trop peu de temps (*Les statistiques : de leur utilisation dans le domaine public en remontant vers leurs sources, exemple tiré du débat national sur l'école*, René Mulet-Marquis, repère IREM n°55, 2004)

Distributions d'articles (au choix d'article par groupe après une description succincte) pour préparer le travail des séances suivantes.

Enseigner les statistiques en seconde, Rémy Coste, APMEP 425, décembre 1999 (???)

La simulation en statistiques, Philippe Dutarte, Repère IREM 47, avril 2002

Modélisation et simulation en classe, quel statut didactique ? Jean Claude Girard, Michel Henry, à paraître

Sur quelques conceptions du hasard, Bernard Courtebras, ... (gros article)

La liaison statistique probabilités dans l'enseignement, Jean Claude Girard, Repère IREM 57, octobre 2004

Simulation et modélisation d'un problème de bouteilles, François Huguet, Repère IREM 36, juillet 1999

Ce dernier article n'a été choisit par aucun groupe

❖ *Séance 6 (4/11/4)*

Travail en groupe

Lecture d'un article. L'objectif est d'en faire un compte rendu oral reprenant les grandes lignes de la fiche de lecture : présentation de l'auteur et du contexte, résumé, commentaire (clarté, intérêt pour un enseignant, un formateur...).

¹⁶ Cf. annexe page 15.

❖ *Séance 7* (18/11/4)

Présentation des séances de janvier, proposition de scénario (et d'organisation de la séance), discussion.

Bilan de la séance biblio du 4 novembre (première partie)

Présentation du travail de recherche biblio.

Objectif : constituer une bibliographie succincte à présenter lors d'un stage (concrètement, la séance de janvier).

Le travail est réparti sur 5 groupes :

Repère IREM année 2000 et *Petit x* 1999-2000

Repère IREM année 2001 et *Petit x* 2000-2001

Repère IREM année 2002 et *Petit x* 2002

... etc.

❖ *Séance 8* (25/11/4)

Travail de recherche bibliographique.

❖ *Séance 9* (2/12/4)

Bilan de la séance biblio du 4 novembre (seconde partie : présentation de l'article de Courtebras)

Bilan oral de la recherche bibliographique.

Présentation d'une bibliographie sur « test » suite à une demande de complément :

Tests d'hypothèses, *Promenades mathématiques*, Frédéric Laroche, Ellipse

Tests d'adéquation à une loi de probabilité, pratique des tests du Khi-deux, Louis Maris

Bonneval, Michel Henry, à paraître

Introductions aux tests d'hypothèses, Michel Henry, à paraître

4^{ème} partie : deux séances de formation

Réflexion sur la séance du 13 janvier : répartition du travail, travail en groupe.

❖ *Séance 10 et 11* (9/12/4 et 6/1/5)

Travail en trois groupes sur la séance de janvier.

❖ *Séance 12* (13/1/5)

Séance présentée par les étudiants à l'autre groupe du dispositif (qui a travaillé essentiellement sur les objets de statistique descriptive depuis septembre) :

- bibliographie (liste d'articles commentée)
- commentaire des programmes
- expérience spaghetti, différentes modélisations

- point de vue plus théorique (cette partie est ajoutée à la maquette prévue par l'enseignant, à la demande des participants du groupe)

Certains documents sont proposés en annexes¹⁷. L'ensemble des documents est disponible sur le site du Master Professionnel.

❖ *Séance 13* (27/1/5)

Séance organisée par l'autre groupe du dispositif (statistique descriptive).

4) Déroulement des 3 jours du PAF 2004-2005 : "Statistique et probabilités de la seconde aux classes de terminales"

Jour 1 (18/11/4)

Matinée :

- Accueil et présentation des stagiaires, le stage et son contenu prévu
- Exposé : Stat au collège, stat au lycée (évolution des programmes dans le temps, évolution des contenus de la sixième à la terminale)

- Analyse d'activités sur hasard et modélisation en seconde

Après-midi :

- A propos d'échantillonnage, de fluctuations, de fourchettes (exposé théorique¹⁸)
- Travail en petits groupes sur deux activités de seconde à développer et mettre en forme pour utilisation en classe

- Mise en commun-bilan de ce travail

- Réflexion sur le contenu des deux jours suivants de stage

Jour 2 (27/1/5)

Matinée :

- Analyse d'exercices de statistique descriptive en seconde et introduction de la notion de "glissements"

- Travail de bibliographie sur des articles parus dans RDM ou APM

- Apports théoriques sur les tests¹⁹.

Après-midi :

- Apports théoriques sur les tests (suite)

- Analyse de deux séquences observées en classe de seconde (construction d'un histogramme, construction et comparaison de diagrammes en boîte) : mise en évidence de glissements. (Exposé de Pablo Carranza²⁰)

¹⁷ Voir annexe page 67.

¹⁸ Voir page 15.

¹⁹ Voir page 20.

Jour 3 (31/1/5)

Matinée :

Excel et calculatrices : des simulations (au CARFIM)

Après-midi :

- Apports théoriques sur les lois continues
- Exemple de travail effectif en classe de TS sur les lois exponentielles
- Compte-rendus bibliographiques des articles donnés le 27
- Bilan-évaluation du stage

L'équipe d'animateurs de ce stage :

Bernadette Denys, Bernard Parsysz, Christophe Hache, Jacqueline Mac Aleese, Brigitte Chaput, Michèle Boyer, Pablo Carranza

²⁰ Cf. note 2 page 3.

CONCLUSION

Tout d'abord il manque une évaluation ; les contraintes de temps n'ont pas permis de retour.

D'autre part les études didactiques sur le sujet sont encore rares, il y a par exemple peu de recherches sur les difficultés des élèves en statistique, que ce soit avant les cours ou en lien avec les cours ; on a vu que les notions statistiques sont rarement introduites comme outil, ce qui fait travailler les définitions, en revenant aux connaissances antérieures ; enfin les connaissances statistiques des enseignants sont encore hétérogènes.

De ce fait, les formations sont davantage des formations d'enseignants que de formateurs. La difficulté des contenus a même rebuté certains participants au DU ; même si le travail organisé autour du hasard a été plus facile pour les participants que celui de statistique descriptive, ce sont davantage les connaissances en statistique qui ont été travaillées dans tous les cas que des connaissances sur l'enseignement de la statistique. Les problèmes de gestion de la classe ont notamment été peu abordés, malgré l'importance des données dans ce domaine ainsi que celle des calculs qu'on peut faire sur ordinateur, toutes choses qui peuvent modifier le déroulement habituel.

BIBLIOGRAPHIE

- BONNEVAL Louis Marie, HENRY Michel ; Tests d'adéquation à une loi de probabilité, pratique des tests du Khi-deux, à paraître
- BOYE Anne, COMAIRAS Marie-Claire (2002) ; Moyenne, médiane, écart type. Quelques regards sur l'histoire pour éclairer l'enseignement des statistiques, *Repère IREM* n°48
- CARRANZA Pablo (2004) ; Glissements dans l'enseignement de la statistique en classe de seconde, *Mémoire de DEA* (non publié)
- COSTE Rémy (1999) ; Enseigner les statistiques en seconde, *APMEP* 425
- COURTEBRAS Bernard ; Sur quelques conceptions du hasard
- DUPERRET Jean Claude (1992) ; Mode, moyenne, médiane au collège : pourquoi ? *Repère IREM* n°6
- DUTARTE Philippe (2002) ; La simulation en statistiques, *Repère IREM* 47
- GIRARD Jean Claude (1999) ; A bas la moyenne, à propos des paramètres de tendance centrale et de dispersion d'une série statistique, *Repère IREM* n°33
- GIRARD Jean Claude (2004) ; La liaison statistique probabilités dans l'enseignement, *Repère IREM* n°57
- GIRARD Jean Claude, HENRY Michel (2005) ; Modélisation et simulation en classe, quel statut didactique, à paraître
- HENRY Michel ; Introductions aux tests d'hypothèses, à paraître
- HUGUET François (1999) ; Simulation et modélisation d'un problème de bouteilles, *Repère IREM* n°36
- MULET-MARQUIS René (2004) ; Les statistiques : de leur utilisation dans le domaine public en remontant vers leurs sources, exemple tiré du débat national sur l'école, *Repère IREM* n°55
- PARZYSZ Bernard (à paraître) ; actes des journées de l'APMEP de 2004, Orléans
- ROBERT Claudine ; *Comptes et décomptes de la statistique*, Vuibert
- ROBERT Claudine (2005) ; Des images dans les plaques de chocolat, *Repère IREM* n°58
- ROBERT Aline et POUYANNE Nicolas (2004) ; Formateurs d'enseignants de mathématiques du second degré : éléments pour une formation, *Document pour la formation des enseignants* n°5, IREM-Université paris 7.
- ROBERT Aline (2005) ; Recherches en didactique des mathématiques et formations professionnelles des enseignants du second degré en mathématiques - l'exemple d'une formation de formateurs, *Actes du séminaire de didactique National 2004-2005*.

Annexes

Annexe I	Pablo Carranza : Présentation du DEA	2
Annexe II	Grille de lecture d'article	15
Annexe III	Grille d'analyse d'exercices.....	16
Annexe IV	Dossiers réalisés lors de l'option statistique du Master (2003-2004) ...	18
1)	Adéquation	18
2)	Mode, moyenne, médiane	18
3)	Simulation	18
Annexe V	Documents de l'option statistique/Probabilité, 2004-2005, Groupe "Statistique descriptive"	23
1)	Exercices analysés (séance 1 et 2).....	23
2)	Documents relatifs à la séance 13	59
.a.	Plan de la séance	59
.b.	Bibliographie commentée.....	59
.c.	Réflexion à propos d'un stage "sur site" (plan du document)	61
Annexe VI	Documents de l'option statistique/Probabilité, 2004-2005, Groupe "hasard"	62
1)	Activités analysées lors de la séance 1	62
2)	Bilan de la séance 3	66
3)	Documents relatifs à la séance 12	67
.a.	Plan de la séance	67
.b.	Bibliographie commentée.....	67
4)	Expérience mise en oeuvre pendant la séance.....	69
.a.	Descriptif.....	69
.b.	Programmes des simulations	73

Annexe I

PABLO CARRANZA :
PRESENTATION DU DEA

*« Glissements dans l'enseignement de la statistique
en classe de seconde »*

Introduction

La statistique est caractérisée par un ensemble de particularités qui, *a priori*, semble s'opposer aux pratiques courantes en classe de mathématiques. Nous nous sommes intéressé au problème de la transposition didactique (Chevallard, 1985) de la statistique descriptive en classe de Seconde et spécialement aux transformations substantiellement importantes. Nous étudions cette distance entre " la statistique enseignée " et la statistique en tant que domaine du " savoir savant ", et ce, pour l'instant, en nous limitant à la statistique descriptive.

Nous le proposons via l'introduction d'un terme : le " glissement ". Nous employons ce vocable pour décrire l'observation d'un phénomène particulier caractérisé par :

- a) un déplacement de la problématique centrale (d'un exercice placé dans le chapitre statistique) vers autres domaines des mathématiques.
- b) une perte significative de sens statistique dans ce déplacement. Une supposition a été faite, les " glissements " ne se limiteraient pas à l'espace partagé par professeurs et élèves : ils se manifesteraient, en amont, dans les étapes successives de la transposition didactique.

Ceci nous a conduit à analyser le programme de la classe de Seconde ainsi que les manuels des éditions 2000 et 2004 en finissant par l'observation et l'analyse de séances ordinaires. Notre étude inclut donc cinq parties. La première consiste en une caractérisation épistémologique de la statistique. Ici nous précisons leurs principaux objets d'étude ainsi que

leurs critères de validation. Cette caractérisation nous permet d'identifier les " glissements " présents dans les successives étapes de la transposition didactique. Une deuxième partie est consacrée à l'analyse du programme de la classe de Seconde, en repérant plusieurs " glissements " dans ses propositions. La troisième partie consiste en analyser les activités du chapitre de statistique descriptive. Les respectifs analyses *a priori* (Robert, A. 1998) des exercices contribuent à identifier les " glissements " et à les classer par type. En la quatrième partie, nous présentons les " glissements " qui se sont produit dans des séances ordinaires (4 en total) observées durant le mois de février 2004 en deux lycées aux alentours de Paris. Finalement, nous présentons un résumé de l'ensemble de nos analyses en identifiant quelques aspects caractéristiques de la statistique descriptive traitée en classe de Seconde.

Une enquête épistémologique

Pour pouvoir déceler les possibles " glissements ", une étude épistémologique a été réalisée (Hacking, 1983 ; et autres). Nous proposons un bilan sous la forme d'une caractérisation de la " démarche statistique ".

Dans cette caractérisation nous distinguons un aspect partagé par la totalité d'auteurs consultés : La statistique intègre les statistiques théoriques aux statistiques contextualisées dans un dispositif cognitif et social. De cette manière, le modèle mathématique du traitement de l'information statistique n'est qu'un outil théorique de ce traitement, qui n'est lui-même qu'un moment d'une opération plus complexe qui lie connaissance et praxis.

C'est cette intégration théorie – contexte qui confère à la statistique un statut particulier. Armatte (Armatte, M. 1997) affirme :

« cette opération consiste à fonder une action ou une décision sur un raisonnement inductif contrôlé à partir d'observations et de mesures multiples ».

Les statistiques ont une composante nettement associée à la pratique, à sa pratique : l'emploi des outils statisticomathématiques pour la réduction de données.

Pour analyser l'opération socio-cognitive complexe qui recouvre les statistiques nous proposons trois phases suivantes :

Première phase

La première phase s'occupe de la construction des faits et de leur mesure. Elle traite de questions aussi fondamentales que celles des catégories, nomenclatures et taxinomies à travers lesquelles le réel est appréhendé.

Les statistiques traite aussi des méthodologies d'investigation, celles qui accouchent de relevés d'information qu'on appelle des "données". Cette même phase traite aussi des échelles de mesures sur lesquelles ces catégories sont projetées, et des méthodes numériques, tabulaires et graphiques sous lesquelles la même information peut être réduite et communiquée.

Seconde phase

La seconde phase consiste à produire sur la base des faits contingents un discours théorique

général. C'est un moment du travail statistique qui amène à manipuler ce que l'on appelle parfois inférentielle et qui a pour fonction de fournir les raisonnements, les enchaînements de la preuve qui prendront place dans un discours argumentaire fondé sur la répétition et les grands nombres. Le raisonnement peut être inductif s'il va des observations aux régularités et aux lois ou bien de type hypothétique-déductif si les observations servent à valider ou réfuter un jeu d'hypothèses a priori qui prend souvent la forme d'un modèle.

Troisième phase

La troisième phase est celle de la décision et de l'action. Dès l'origine, les statistiques se sont souciées de fournir des règles de décisions optimales dans des questions faisant intervenir des situations incertaines : assurances maritimes, rentes viagères, témoignages, décisions de justice sont l'occasion d'élaborer ces règles et de les fonder sur un calcul nouveau, celui de la valeur (ou l'utilité) espérée.

Cette caractérisation nous sert de guide pour examiner les propositions du programme, celles des manuels et le déroulement effectif en classe.

L'analyse du programme

Nous avons considéré le programme d'une façon large : le texte du programme, les documents d'accompagnement de ce programme et les documents de la commission responsable de la rédaction des programmes.

Bien que les documents d'accompagnement et ceux qu'inspirent les programmes font référence à la même démarche statistique que celle précisée en la première partie (« Une enquête épistémologique »), la présentation du programme présente à notre avis un " glissement ".

Dans la démarche statistique, une première étape est en effet nécessaire : celle de la construction des faits à étudier et de leur mesure. Cette phase traite les définitions des caractéristiques à observer et sa pertinence par rapport au sujet étudié sur la population cible.

Cette étape inclut aussi l'exploration dans l'ensemble de données de sous-populations, de données aberrantes, de données manquantes, afin de vérifier que l'on travaille sur une "matière première" appropriée à l'étude en question.

Aucune référence n'est faite sur ce sujet. En réalité cette contextualisation initiale de toute étude statistique est en général "externe" aux mathématiques, tant en spécifiant fondamentalement les choix mathématiques qu'il y a là une spécificité statistique de la dynamique "contextualisation-decontextualisation".

Certes, le professeur a la liberté de choix et c'est explicité dans le programme :

«...il proposera des sujets d'étude et des simulations en fonction de l'intérêt des élèves, de l'actualité et de ses goûts».

Néanmoins, aucune " capacité attendue " ne relève du champ statistique et probablement les contenus pour les évaluations seront limités à un travail des objets mathématiques.

Bref, l'actuelle présentation du chapitre statistique centre l'attention sur les objets mathématiques de la statistique, " objets " qui constituent un modèle mathématique du traitement de l'information statistique et qui n'est qu'un " outil " théorique de ce traitement. C'est ce qui nous amène à postuler la présence d'un " glissement " dans le programme de Seconde : ignorance d'une phase préliminaire fortement contextualisée entraînant des choix.

L'analyse de manuels

Nous avons analysé la section statistique descriptive de treize manuels édités en 2000 ou 2004. Des analyses *a priori* (Robert, 1998) nous ont permis d'identifier les tâches potentiellement concernées par les exercices. Nous avons alors identifié et caractérisé un certain nombre de glissements (suites et sommes, géométrie, pourcentage, proportionnalité, algèbre – arithmétique, calcul...), ainsi que quelques cas particulières (moyenne, médiane . En certains cas, nous présentons un exemple, à fin d'illustrer les critères qui ont guidé nos choix :

Glissement "suites et sommes"

Ces exercices sur le symbole Σ et son utilisation sont présentés dans le chapitre "statistiques" et ne font, en rien, travailler dans ce domaine. Ces exercices sont une réponse aux commentaires des programmes : "On peut commencer à utiliser le symbole Σ "

Considérons un exemple : (Belin 2000 : n° 7 p 183) :

Calculer la somme suivante $\sum_{i=1}^8 \frac{1}{i} =$

Le signifiant « Σ » est en effet utilisé en statistique pour sommer les termes d'une suite dont la progression n'est pas, par essence, "régulière" (on pourrait dire "complètement modélisable par une formule").

Glissement "géométrie"

Dans cette partie, les élèves sont amenés à travailler les statistiques dans un contexte géométrique. Considérons le exemple suivant: (*Déclic 2000, n° 21 page 192*) :

Soit trois triangles dont les côtés mesurent respectivement (6, 8, 10) ; (5, 12, 13) ; (8, 15, 17). Ces triangles sont-ils rectangles ? Calculer la moyenne des côtés les plus courts de ces trois triangles, puis la moyenne des côtés moyens, puis celle des côtés les plus longs. Le triangle ayant ces trois moyennes comme côtés est-il rectangle ?

La moyenne en statistique est une mesure de position centrale d'une population, elle offre un certain type d'information synthétique qui est interprétée dans le contexte des données. L'emploi donné ici à la moyenne est nettement différent de celui donné en statistique. Ici, la moyenne n'a pas le sens de mesure de position centrale. ici elle perd toutes ces utilités et se transforme en formule de calcul.

Glissement "pourcentage-proportionnalité"

A partir d'un travail sur un graphique statistique (à interpréter ou à construire), les tâches dévolues aux élèves vont essentiellement être centrées sur la proportionnalité entre la surface représentant un groupe d'"individus" et l'effectif (ou la fréquence) de ce groupe.

Glissement "Algèbre-arithmétique"

Ici, nous regroupons les exercices qui ont comme activité centrale le travail algébrique (ou arithmétique) et qui n'abordent pas des thèmes qui sont d'intérêt en statistique. En général, des exercices exigent une mise en équation de certaines valeurs et une (ou plusieurs) inconnue(s), pour tout de suite trouver la donnée manquante par résolution de cette équation. Ces exercices n'utilisent que la formule des indicateurs statistiques, oubliant de cette façon leur sens statistique.

Glissement "calcul"

Les exercices de ce groupe possèdent un dénominateur commun : quelques éléments sembleraient indiquer qu'il s'agit de véritables problèmes de statistique : les données sont fournies, un contexte est présenté et il faut calculer un indicateur.

Néanmoins, ils n'exigent que le calcul d'un ou plusieurs indicateurs statistiques.

Ces exercices ne laissent aucune place aux élèves pour le raisonnement inductif, pour l'inférence de lois qui sont l'un des objectifs d'une étude statistique. (voir étapes de la démarche statistique *Erreur ! Source du renvoi introuvable.*). Généralement, les données sont fournies déjà synthétisées et regroupées (dans deux tableaux). Le contexte indique la variable considérée, la population, etc. Les possibles interprétations (et/ou inductions) qu'on pourrait y tirer sont soit déjà fournies soit non sollicitées.

Nos conclusions seront développées après le traitement de certains cas ponctuels déjà annoncés.

Le cas particulier de la médiane

Les transformations que subit le processus de détermination de la médiane constituent un clair exemple de transposition didactique (Chevallard, 1991). La méthode proposée dans quelques manuels consiste à trouver les effectifs cumulés quelques notions de proportionnalité sont utilisées pour déterminer (méthode graphique) et calculer (méthode analytique) la médiane. C'est une opportunité propice pour l'intégration de contenus de proportionnalité : le théorème de Thalès gagne une place centrale dans la justification de la méthode, car après le graphique vient la mise en équation qui conduit à l'algèbre.

Propriétés de linéarité de la moyenne

Ces contenus sont dans le programme de la classe de seconde, listés dans la colonne « capacités attendues ». Cet ensemble de propriétés en statistique est courant dans l'étape de modélisation et même dans la construction d'une règle de décision mais toujours sous un contexte relativement important d'abstraction. Les propriétés de linéarité sont plus adaptées pour une étape de modélisation et abstraction, ce dernier ne se produit pas en classe de seconde. Nous interprétons les glissements qui interviennent concernant les propriétés de linéarité de la moyenne comme conséquence de cette inadéquation des contenus de statistique dans le programme de la classe seconde.

Notre travail continue avec l'étude de séances ordinaires. Nous présentons nos analyses sur

les glissements produits dans 4 séances enregistrées de deux professeurs en les alentours de Paris pendant l'hiver de l'année 2004.

L'analyse de séances ordinaires

Nos données sources consistent en des enregistrements sonores accompagnés d'une prise de notes effectuée sur place. Un regard transversal sur les séances nous a permis de mettre en évidence certaines régularités de la statistique traitée, tel le cas d'une typologie des données traitées, ainsi qu'une comparaison des emplois des concepts de " moyenne " et de " médiane ". Nous allons maintenant présenter les glissements trouvés dans les séances observées, pour en suite présenter les mentionnées régularités.

Glissement «Histogramme»

Nous avons observé une séance de Caroline (le 16 janvier 2004) dont la construction de cet histogramme a été le sujet central. Pour l'histogramme qui nous occupe, à cause des largeurs différentes des bases des colonnes, on est forcé d'utiliser la notion de proportionnalité et de géométrie dès la construction du graphique. Et lorsque la lecture des effectifs de chaque classe est demandée à partir du graphique, les élèves sont amenés à utiliser des intermédiaires (règle, grille, table). Dans cette séance, le sens du graphique (la lecture immédiate de l'information et de possibles relations) s'est dilué dans les techniques de calcul. Et ce qui aurait du être une façon rapide de visualiser les données s'est transformé en un problème d'aires et proportionnalité. Le graphique a perdu son sens primordial, celui de présenter des relations qui seraient difficiles à identifier ailleurs, et s'est transformé en une opportunité pour appliquer des notions de proportionnalité.

Nous allons maintenant présenter un autre cas, il s'agit aussi d'un objet graphique : la boîte à moustaches.

Glissement «Boîte à moustaches»

Dans cette séance, le professeur Bruno (le 17 janvier 2004) présente la boîte à moustaches comme un outil de comparaison de populations, après avoir traité les mesures de position centrale et de dispersion : moyenne, médiane et étendue. Le séance consiste à construire deux boîtes à moustaches représentant les notes de deux hypothétiques élèves.

Dans la séance observée, les boîtes ont été désignées une à côté de l'autre et sans une échelle générale de comparaison. Bref, dans la séance en question, nous n'avons pas eu la possibilité d'observer ni d'interprétations ni de propositions de possibles relations mises en évidence par

le graphique grâce à l'intégration d'information que ce genre de représentations fournit. La séance s'est centrée sur la construction de l'objet « boîte à moustaches ».

Regard transversal sur les séances observées

Nous nous intéressons ici à savoir quelle place occupent les différents objets de la statistique dans la statistique enseignée. Bien entendu, la liste n'est exhaustive et nous nous contentons en prenant quelques éléments typiques de statistique liés aux contenus de la classe de Seconde, nous tentons donc, une description générale de quelques concepts de statistiques.

Les données utilisées en classe

Sur la dimension

Dans les séances observées, les données ont été unidimensionnelles, présentés sous la forme d'une liste ou dans un tableau, elles ne sont jamais du type multidimensionnel ni présentent aucune chronologie spécifique.

Le professeur, qui prend à sa charge les interprétations et leur gestion (voir page **Erreur ! Signet non défini.**), qui les développe et les confronte.

Sur le nombre

Le nombre de données traitées en classe a été inférieur à 50. Dans les séances enregistrées professeurs et élèves utilisaient des calculatrices pour l'obtention de la moyenne. La médiane, les quartiles et l'étendue se retrouvaient directement sur la liste généralement ordonnée de données.

Sur la terminologie employée

Certains termes spécifiques à la statistique sont absents dans la classe (discours du professeur, exercices...). Les mots "population", "individu", "variable", "échantillon" etc. n'ont été pas utilisés pendant les séances enregistrées.

Comparaison Moyenne - médiane

Pendant les séances observées, professeurs et élèves effectuent les calculs de moyenne et médiane mais, en général, il n'y a pas d'interprétation sur les calculs réalisés, sur la différence des résultats. Le paramètre est calculé et les interprétations intra-séries n'ont pas lieu. Dans aucun des cas où ces deux indicateurs statistiques ont été calculés il ne s'est produit une interprétation des leurs différences (ou des similitudes) au-delà d'une comparaison numérique.

Résultats

Bien que les documents d'accompagnement et les idées des concepteurs des programmes soient proches de notre caractérisation de la statistique, le chapitre statistique des programmes focalise l'attention sur les parties mathématiques des concepts statistiques concernés.

En ce qui concerne les exercices des manuels, nous avons proposé une caractérisation des glissements observés. Nous constatons aussi que la construction et l'exploration de données sont généralement absentes des exercices des manuels, et que les problèmes qui placent les données dans un contexte sont exceptionnels. Les exercices transmettent majoritairement une image de transparence, produit de l'absence d'explicitation des choix et des interprétations. Ils semblent refléter un certain scénario de déroulement statistique : les données existaient avant même se poser les questions, la variable à observer est dénuée de subjectivité, les conclusions sont claires, précises et non contradictoires.

Dans les séances que nous avons observées, nous constatons une forte concentration sur les processus de calcul et sur l'application de contenus d'autres domaines des mathématiques, ainsi que l'absence de lecture immédiate de l'information, et que de possibles relations mises en évidence par les graphiques. La représentation symbolique des indicateurs est absente ; le vocabulaire partagé par professeurs et élèves concernant la statistique est autour des mots liés aux contenus visés (moyenne, médiane, étendue, quartiles, etc.), et n'incluent pas les termes population, variable, individu, tendance, distribution, etc. Les séances observées suggèrent une répartition des tâches en classe : les élèves ont à leur charge les calculs, ce sont eux qui trouvent la valeur de la moyenne, la médiane et les quartiles. Le professeur a tendance à conduire les conclusions et les interprétations de l'exercice. Les élèves ne choisissent pas les indicateurs utilisés. Pour l'étendue, la supposition d'une distribution uniforme des données est présente dans le langage. Pour la moyenne et la médiane, les critères de pertinence d'utilisation en tant que mesure de position centrale, n'ont pas été examinés dans les classes observées. La représentation graphique n'a pas joué le rôle d'instrument de mise en relation, ni

Conclusions

Nous avons observé l'épistémologie, les programmes, les manuels et la classe. Et nous avons caractérisé des glissements (à tous les niveaux et de diverses importances et type). Dans le cadre de notre thèse en didactique des mathématiques, nous poursuivons actuellement cette recherche en analysant la possibilité d'introduire la notion subjective (degré de croyance) de probabilité. Notion qui complémente l'approche fresquistes et qui peut être appliquée au calcul

de probabilité d'événement uniques.

Bibliographie

- [1] ARMATTE, M. (1995). Histoire du modèle linéaire. Formes et usages en statistique et économétrie. Thèse EHESS, Paris.
- [2] ARMATTE, M. (1997). Les développements de la raison statistique et leurs liens avec les mathématiques. *Colloque international d'histoire des mathématiques*. CIRM-Luminy.
- [3] BARTHELEMY, J.-P.(1993). *Bulletin de la Société Francophone de Classification*. N° 1 Juillet 1993.
- [4] BROUSSEAU, G. (1986). Fondements et méthodes de la didactique des mathématiques. *Recherches en Didactique des Mathématiques*, Vol 7 n° 2.
- [5] CHEVALLARD, Y. (1991). *La transposition didactique, du savoir savant au savoir enseigné*. Editions La Pensée Sauvage, Grenoble
- [6] CICOUREL, A. (1964). *Method and measurement in sociology*. New York, The Free Press of Glencoe. .
- [7] CROMBIE, A. (1980). Styles of thinking and historiography of science, *I Congreso de la Sociedad Española de Historia de las Ciencias*, S. Garma Editores, Madrid.
- [8] CROMBIE, A (1994), *Styles of Scientific Thinking in the European Tradition: The history of argument and explanation especially in the mathematical and biomedical sciences and arts*, London, Duckworth.
- [9] CROMBIE, A (1992), *Stili di pensiero scientifico agli inizi dell' Europa moderna*. Napoli.
- [10] DALLAS, E. (1998). *Métodos multivariados aplicados al análisis de datos*. Thompson Editeurs.
- [11] DOUADY, R. (1994). Ingénierie didactique et évolution du rapport au savoir. *Repères IREM* n° 15.
- [12] GASÇON, J. (1994). Un nouveau modèle d'algèbre élémentaire comme alternative à l'arithmétique généralisée. *Petit x* n° 37.
- [13] GRAS, R. (1996). *L'implication Statistique*. La Pensée sauvage éditeurs.
- [14] Commission de réflexion sur l'enseignement des mathématiques. Rapport d'étape. Statistique et Probabilité. Juin 2002 et Mars 2003

- [15] DESROSIERES, A. (2000). L'histoire de la statistique comme genre. *Revue Genèses* n° 30. Editions Belin.
- [16] DESROSIERES, A. (2001). Entre réalisme métrologique et conventions d'équivalence. *Revue Genèses* n° 43. Editions Belin.
- [17] DESROSIERES, A. (2002). *Revue Courrier de la statistique* n° 104
- [18] ESCOFFIER, B. (1997). *Initiation aux traitements statistiques*. PUR.
- [19] FISHER, R. (1925). *Statistical methods for Research workers*. Edinburgh, Oliver and Boyd.
- [20] HACKING, I. (1983). *The Accumulation of Styles of Scientific Reasoning*, Kant order Hegel, Dieter Heinrich ed., Stuttgart, Klett-Cotta.
- [21] HACKING, I. (1990). *The taming of chances*. Cambridge. Cambridge University Press.
- [22] HACKING, I. (1999). *The social construction of what ?* Cambridge, Mass., Harvard University Press.
- [23] HERAN, F. (1984). L'assise statistique de la sociologie. *Revue Economie et statistique*, n° 168.
- [24] LAHANIER-REUTER, D. (1999). *Conceptions du hasard et enseignement des probabilités et statistiques*. P.U.F.
- [25] MARTIN, O. (1997). *Les statistiques parlent d'elles mêmes. Regards sur la construction sociale de la statistique*. La Pensée confisquée, Paris, La Découverte
- [26] MORGAN, M. (1990). *The story of économétric ideas*. Cambridge, Cambridge University Press.
- [27] PARZYSZ, B. (1997). Les probabilités et la statistique dans l'enseignement secondaire, d'hier à aujourd'hui, in "*Enseigner les probabilités au lycée*" (collectif). Ed. IREM de Reims..
- [28] PARZYSZ, B. (1998). *Document de synthèse pour l'habilitation à diriger des recherches en Didactique des mathématiques*. Université Paris VII.
- [29] REGNIER, JC. (2002) A propos de la formation en statistique...Questions éducatives. *Revue du centre de recherche en éducation*, Saint-Etienne. Université Jean Monnet.
- [30] ROBERT, C. (2002). *Conférence sur les nouveaux programmes des lycées*. 16 janvier 2002 à Clermont-Ferrand).

- [31] ROBERT, A. (1992). Problèmes méthodologiques en didactiques des mathématiques. *Recherches en didactiques des mathématiques*, vol 12 n° 1.
- [32] ROBERT, A. (1998). Outils d'analyse des contenus mathématiques à enseigner au lycée et à l'université. *Recherches en didactique des mathématiques*, vol 18 n° 2.
- [33] STIGLER, S.M., (1986). *The story of statistiques. The mesasurement of uncertainly before 1900*. Belknap Press of Harvard University Press, Cambridge et London

L'objectif général d'un travail de lecture est de renseigner, à l'oral ou à l'écrit, sur le contenu présenté par les auteurs, en permettant la critique et en dégageant éventuellement une synthèse " à retenir ".

Lorsqu'on expose un article (ou un livre), le fait d'annoncer à l'avance le plan et la tonalité générale de l'exposé, et le fait de proposer systématiquement les citations sur transparent peuvent contribuer au suivi de la présentation !

A) Partie descriptive " en amont " : contexte de l'article, auteurs, objectifs de l'article (pour qui, pour quoi), statut, contenus

1) Sur quoi ça porte (contenus) ?

Plusieurs sujets peuvent être abordés de manière plus ou moins imbriquée

- les mathématiques (d'un niveau donné, transversal, ...), points de vue épistémologique, historique etc.
- Les programmes, instructions, commentaires, documents d'accompagnements (idem)
- Les élèves :
diagnostics de connaissances ou de performances,
diagnostics de connaissances ou de performances en relation avec un enseignement ou un programme,
compte tenu d'une théorie de l'apprentissage, diagnostics de connaissances ou de performances, avec ou non élaboration de stratégies
- Stratégies d'enseignement (motivées ou non ; si oui, motivées par rapport aux programmes, aux mathématiques, à une théorie ; expérimentées ou non, limitées aux contenus ou accompagnées d'éléments de gestion), à une échelle globale (plus d'un mois) ou locale (on parlera alors de séquence)
- Les formations

Les échelles concernées (une séance, une année, une décennie..., un élève, une classe, un établissement, une classe d'âge) complètent utilement ces informations.

2) Qui a écrit (auteurs) ?

Dans nos revues, plusieurs publics s'expriment ou sont visés : les enseignants, la noosphère et l'institution, les formateurs, les chercheurs, des équipes mixtes.

Ainsi pourra-t-on renseigner les questions suivantes :

Qui sont les auteurs ? A quelle occasion le livre a-t-il été écrit (commande, manifestation, rien) ? Cela permet de se demander quelles ambitions a le livre, et d'en tenir compte pour les interprétations éventuelles.

La date d'écriture (ou de publication) complète utilement ces renseignements.

Exemples : dans le bulletin de l'APM, écrivent des enseignants et des représentants de la noosphère (certains) souvent sur les mathématiques et les programmes, ou bien des enseignants pour donner des exemples bruts de cours ou d'exercices.

Dans petit x, plutôt des chercheurs, et des enseignants, avec des articles sur les mathématiques et les élèves (stratégies, ou séquences motivées, quelquefois par une théorie, voire expérimentées).

Dans Repères-IREM, plutôt des enseignants, et des chercheurs, avec des articles à destination des enseignants pour leurs élèves (séquences, motivées plutôt par les programmes).

Certains écrits émanent d'enseignants et n'affichent aucune prétention théorique, même si une véritable réflexion est proposée. D'autres écrits peuvent se situer dans plusieurs " paradigmes " qui traversent les recherches actuelles en didactique des mathématiques : dialectique outil/objet et jeux de cadres, Théorie des Situations, anthropologie, théorie de l'activité...

On peut ainsi chercher à préciser quel est le statut revendiqué du texte : recherche (diagnostic, expérimentation, ingénierie), informations, réflexion, récit, innovation, synthèse... Cela contribue à indiquer le poids relatif qu'on pourra accorder à cet écrit.

Dans le même ordre d'idées, on peut signaler s'il existe une bibliographie ou des annexes précises.

4) Cas d'un texte de recherche

Pour mieux apprécier un tel texte, il est nécessaire de savoir si les présupposés théoriques sont bien précisés, développés ou s'ils manquent (implicites).

On peut aussi dégager s'il y a des preuves apportées au propos, des références, si les auteurs explicitent des hypothèses admises, ou bien testées.

Enfin, il est important de signaler si les auteurs apportent eux-mêmes des limites, des nuances, des critiques, ou suggèrent des perspectives.

B) Le résumé de tout ou partie du texte.

C'est un exercice difficile, quelle que soit la taille du résumé. Attention : bien distinguer ce qui est résumé du livre ou l'avis du lecteur n'est pas toujours évident !

C) Critiques, synthèse et perspectives

Plusieurs questions permettent d'organiser cette partie essentielle.

1) Lisibilité

Est-ce facile à lire ?

La lecture est-elle autonome (ou bien faut-il lire ou connaître autre chose pour comprendre) ?

Que dire de la biblio ?

Y a-t-il des références théoriques à connaître ou que l'article permet de comprendre ?

2) Fiabilité

Est-ce que les ambitions des auteurs sont remplies ?

Les apports sont-ils convaincants, que ce soit des connaissances, des démarches, des questionnements ?

3) Utilisations ultérieures

Quelles critiques de votre part, compte tenu de tout ce qui précède ?

Quoi retenir " au final " pour un formateur, un enseignant ?

Quels " transferts " ou suites peut-on imaginer ?

Soulignons que ces analyses ne permettent pas à elles seules de comprendre ce que les élèves ont eu finalement à faire en statistique : le déroulement dans les classe peut modifier considérablement les activités déclenchées par un énoncé.

Ce peut être par des ajouts de commentaires qui restituent le cadre statistique d'exercices techniques, ou par des découpages des tâches simples qui donnent lieu à plusieurs activités successives. Par exemple un calcul de médiane peut nécessiter d'ordonner les valeurs avant de compter les effectifs : l'enseignant peut rappeler rapidement cette étape ou non, cela modifie l'activité des élèves. Cela permet à certains d'aborder la deuxième étape (compter), mais cela peut aussi faire oublier à d'autres la première étape qu'ils n'ont pas eu à retrouver seuls.

Ces analyses ont pour ambition de contribuer ainsi à renseigner sur les activités proposées en classe aux élèves, à l'échelle de la séance, pour comparer avec ce qui se passe en classe et permettre de saisir soit des adaptations imprévues soit des modifications dues au déroulement. A terme cela renseigne, en conséquence, sur leurs apprentissages à condition de replacer ces analyses dans des scénarios plus globaux.

Ces analyses sont **relatives à un niveau scolaire donné**, un programme donné, et même une classe donnée : cela explique l'importance de la prise en compte des programmes.

1) Les énoncés proposés aux élèves portent sur des connaissances qui peuvent être **anciennes, en cours d'acquisition** : c'est une première distinction à prendre en compte vis-à-vis des activités et des apprentissages qu'ils peuvent induire. Nous sommes plus précis sur les notions nouvelles que sur les anciennes.

Souvent en statistique la tâche proposée est relativement simple et isolée, par exemple mettre en œuvre une définition, mais ce qui nécessite de nombreuses adaptations est la mise en œuvre de connaissances **non statistiques** pour résoudre.

2) De plus ces connaissances à mettre en fonctionnement peuvent être indiquées (directement ou indirectement par leur place) ou non : on parle de fonctionnement de type **mobilisable ou disponible**. Et le travail des élèves n'est pas analogue selon qu'ils doivent rechercher les connaissances à utiliser, c'est-à-dire mettre en relation une propriété à démontrer et différents outils (travail du pourquoi ou du quoi) ou mettre en fonctionnement en l'adaptant une propriété (travail du comment).

3) **Dans tous les cas**, que ce soit sur des connaissances anciennes ou en cours d'acquisition, et qu'elles doivent être mobilisables ou disponibles, les mises en fonctionnement peuvent varier, avec des conséquences sur les activités et les apprentissages. Un travail sur ces conséquences potentielles nous amène à distinguer des **grands types d'adaptations, caractérisés par les niveaux de mises en fonctionnement définis ci-dessous**.

4) Nous distinguons ainsi pour chaque question d'un énoncé plusieurs caractéristiques. Elles ont été retenues dans la mesure où nous supposons qu'elles interviennent de manière discriminante dans les activités provoquées chez les élèves par les énoncés en relation avec les apprentissages ultérieurs.

Nous prenons ainsi en compte :

La manière dont la question est posée (fermée ou non) – montrer que ? est-ce que ?

Cela renseigne notamment sur le fait que les élèves auront ou non à faire des conjectures ou à modéliser eux-mêmes. Entre « calculer... » et « que peut-on dire de ... ? » il peut y avoir beaucoup de différence de travail. Une question fermée sans aucune indication peut cependant ne pas être abordée immédiatement, s'il manque une indication de méthode, ce qui nous intéresse en termes d'activité.

Les indications (en utilisant...)

Cela renseigne sur le fait que les élèves auront à choisir ou non les méthodes à mettre en œuvre.

Le découpage (donné, à établir, à rétablir...)

Il y a plusieurs cas : ou bien il s'agit de restituer une suite de sous-tâches classiques qui interviennent toujours dans le type de tâche considéré, ou bien il faut concevoir un découpage adéquat. Cela renseigne notamment sur le degré d'organisation des connaissances à mettre en œuvre.

Le niveau de mise en fonctionnement de différentes connaissances à mettre en jeu (compte tenu de ce que les élèves ont à leur disposition)

Cela renseigne sur les adaptations de leurs connaissances que les élèves peuvent être conduits à faire.

5) Un zoom sur les adaptations

Si le travail des élèves consiste à appliquer une propriété sans calcul supplémentaire ni reconnaissance (remplacer les données générales par des données particulières) on parle d'application simple et isolée ou immédiate. Il faut encore souligner qu'une telle application peut enclencher une activité qui n'est pas simple : calculer une médiane par exemple peut amener à utiliser des connaissances anciennes de manière nouvelle (avec des adaptations nombreuses) !

Six types se dégagent, qui peuvent intervenir simultanément, qui ont chacun un spectre assez large (encore une fois tout ceci est relatif à des élèves donnés) :

a. Les reconnaissances (partielles) des modalités d'application des notions, théorèmes, méthodes, formules... Cela peut aller de reconnaissances de variables, de notations à des reconnaissances de formules ou de conditions d'applications de théorèmes. Le contexte, la place dans le manuel ou le cours notamment, sont importants pour déterminer la mesure de ces reconnaissances.

b. L'introduction d'intermédiaires – notations, points, expressions. En statistique ce peut être d'ordonner ou de regrouper des données en classe par exemple.

c. Les mélanges de plusieurs cadres ou notions..., les changements de points de vue, les changements ou jeux de cadres, les mises en relation ou interprétations : typiquement utiliser du calcul algébrique pour obtenir le résultat (par exemple reconnaître un changement affine de données et l'appliquer à la moyenne). En statistique on trouve des jeux de cadres entre numérique et graphique.

d. L'introduction d'étapes, l'organisation des calculs ou des raisonnements (cela va d'utilisation répétée (in) dépendante d'un même théorème à un raisonnement par l'absurde faisant intervenir le théorème). Les étapes peuvent être classiques ou à imaginer.

e. L'utilisation de questions précédentes dans un problème.

f. L'existence de choix – forcés (un seul convient finalement) ou non.

DOSSIERS REALISES LORS DE L'OPTION
STATISTIQUE DU MASTER (2003-2004)

1) Adéquation

Il s'agit d'un document de formation creusant la notion de test d'adéquation à une loi équirépartie en terminale.

Voir page 19 le préambule et le sommaire.

2) Mode, moyenne, médiane

Il s'agit d'un scénario de formation destiné à des enseignants de collège à propos de valeurs centrales.

Voir page 20 le préambule et le sommaire.

3) Simulation

Il s'agit d'un scénario de formation à propos du programme de seconde en statistique et plus particulièrement de la simulation.

Voir pages 21 et 22 les quatre premières pages du dossier.

Le problème de l'adéquation à une loi équirépartie en terminale et son lien avec le test du χ^2

Ce document a été élaboré par cinq professeurs de mathématiques, qui ont suivi une formation de formateurs en 2003-2004, sous l'égide d'Aline Robert et Nicolas Poyanne, à l'intérieur d'une double structure, Paris VII et Université de Versailles Saint-Quentin. Ils ont poursuivi leur travail à l'intérieur d'un module centré sur l'enseignement de la statistique et des probabilités, et encadré par Christophe Hache, à l'IREM de Paris VII. Cette formation a été complétée par un stage du PAF, intitulé « Statistique et probabilités : continuités et ruptures de la seconde à la terminale », toujours à l'IREM de Paris VII.

Il porte sur le problème de l'adéquation de données expérimentales à une loi équirépartie, évoqué très brièvement dans les programmes de Terminale S et ES et développé à travers deux exemples dans l'annexe aux programmes traitant des probabilités et de la statistique. Par sa nouveauté dans l'enseignement des mathématiques au lycée et par son appropriation rapide dans les sujets du baccalauréat, cette partie des programmes soulève un certain nombre d'interrogations pour de nombreux professeurs.

Ce document se veut donc un complément aux ressources existantes sur le problème de l'adéquation à une loi équirépartie. Il s'adresse à tous les enseignants, qui, comme ceux qui l'ont conçu, éprouvent la nécessité d'en savoir un peu plus sur ce thème afin d'éclaircir, d'approfondir, voire de dépasser les consignes des programmes de terminale. Il résulte d'apports théoriques dispensés en particulier par Brigitte Chaput, et d'une réflexion commune sur l'appropriation de ces apports et de leurs transpositions pour les élèves.

PLAN

Introduction :

Qu'y a-t-il dans les programmes de lycée ? Adéquation à une loi équirépartie,
+ Hors programme : test du χ^2 .

I - Mise en place du modèle théorique :

Poser la problématique. Pourquoi le carré des écarts ? Passage du χ^2 au n^2 .
Mise en place du problème par une simulation.

II - Activités élève :

- 1) Dé équilibré :
Séries de lancers pour obtenir le neuvième défile.
Quelles questions se poser pour l'adéquation à une loi équirépartie ?
- 2) Activité sur les naissances.

III - Analyses d'exercices

- Exercice 1 ES Pondichéry 2003
- Exercice ES juin 2003
- Exercice commun S et ES
- Exercice d'un manuel

IV - Activité complémentaire :

Adéquation à une loi normale

Conclusion

ANNEXES

Sujets de baccalauréat.
Table du χ^2 .

Analyse du sujet de terminale de l'apmp.
Des documents du céderom d'accompagnement des programmes.

Stage 'DU2' : Formation de formateurs en mathématiques

Françoise GONJIN
Catherine LEON YAKER
Véronique NICOLLE
Jean François CHESNE
Philippe MERY

1 PREAMBULE

Le dossier présenté ici s'inscrit dans le cadre de la deuxième année de formation au diplôme de formateur d'enseignants de mathématiques assurée à l'IREM de Paris VII par Christophe Haache que nous remercions tout particulièrement de remercier.
 Nous remercions également Brigitte Chaput pour ses interventions toujours très intéressantes ainsi que tous les autres formateurs qui sont intervenus durant cette année de formation.
 Nous remercions aussi Nicolas Pouganne et Aline Robert dont les connaissances transmises lors de la première année de cette formation nous ont été de précieux apports.

Les changements des programmes de mathématiques du collège en 1986 sont marqués par l'introduction des statistiques que l'on trouve sous la rubrique : « Gestion de données ». Successivement mises en place dans les différentes classes de collège, les statistiques font leur entrée en classe de seconde au lycée à partir de 1990. Cette orientation sera confortée par les nouveaux programmes de collèges en 1996 et ceux de la classe de seconde appliquables en septembre 2000 prenant une place importante dans les objectifs d'enseignement et introduisant notamment les simulations.

L'enseignement des statistiques s'est très vite confronté à deux problèmes, celui de la formation des enseignants et celui de la place à lui accorder dans la formation globale dispensée aux élèves. Les raisons qui militent en faveur de l'enseignement des statistiques sont nombreuses. Citons entre autres : la formation du citoyen, l'incitation au travail interdisciplinaire, la possibilité d'utiliser les calculatrices et l'outil informatique.

Malheureusement, au collège, les statistiques restent encore trop souvent le « parent pauvre » de l'enseignement. En effet, leur apprentissage consiste dans le meilleur des cas à remplir des tableaux, à faire des graphiques, à appliquer des formules et dans le pire des cas il est tout simplement oublié et l'impression que les statistiques ne sont pas de vraies mathématiques y est encore très forte.

C'est pourquoi, l'un des objectifs de ce stage est que les professeurs de collège prennent conscience de l'importance de la place des statistiques dans le déroulement général des programmes des classes du second degré de la sixième à la terminale et qu'ainsi ils soient convaincus de l'intérêt de faire des statistiques dès le collège assurant par ce fait la continuité nécessaire de cet enseignement.

TABLE DES MATIERES

1	PREAMBULE	3
2	OBJECTIFS DU STAGE	4
3	PLANNING DU STAGE	5
4	DESCRIPTION DETAILLEE DE LA DEMI-JOURNEE 1	7
4.1	Introduction	7
4.2	Etude n°1 : Les paramètres de tendance centrale : conception et difficultés prévisibles.	9
4.3	Compléments	21
5	DESCRIPTION DETAILLEE DE LA DEMI-JOURNEE 2	26
5.1	Etude n°2: Médiane et moyenne : aspect théorique	26
5.2	Une utilisation de la médiane et des quartiles : les boîtes à moustaches	31
6	DESCRIPTION DETAILLEE DE LA DEMI-JOURNEE 3	40
6.1	Exemples traités par l'outil informatique	40
7	BILAN DU STAGE	61
8	ANNEXES	62
8.1	Un peu d'histoire	62
8.2	Quelques repères historiques	63
8.3	Usage de la calculatrice graphique	64
9	BIBLIOGRAPHIE	69
9.1	Ouvrages généraux	69
9.2	Ouvrages scolaires	69
10	SITES INTERNET	69
10.1	Sites généralistes	69
10.2	Sites spécialisés	69

DOCUMENTS

document 1	Moyenne et médiane	13
document 2	Quelles moyennes ?	14
document 3	Emplois des différentes moyennes	15
document 4	Savoir choisir le caractère de position qui est le mieux adapté	16
document 5	Recherche de résumés statistiques	29
document 6	Boîtes à moustaches	39
document 7	Représentations graphiques de données statistiques en utilisant le tableur Excel	42
document 8	Prise en main du tableur pour le traitement statistique	43
document 9	Observation de l'erreur commise sur la moyenne (groupements par classes)	48
document 10	Moyenne et médiane	52
document 11	Calcul de moyennes avec le tableur	55
document 12	Détermination de la médiane d'une série statistique à variable continue	57
document 13	Comment l'amplitude des classes influe sur la moyenne	60
document 14	Représentations historiques à propos de quelques mois	63
document 15	Utiliser la calculatrice graphique	64

ANNEXES

annexe 1	Corrigé des documents 1, 2 et 3	17
annexe 2	Découvrir d'autres moyennes que la moyenne arithmétique	18
annexe 3	Triangle rectangle et moyenne	19
annexe 4	Que de moyennes	20
annexe 5	Corrigé du texte « Recherche de résumés statistiques »	30

QUESTIONNAIRES

questionnaire 1	Préparation de la troisième demi-journée TICE	8
questionnaire 2	Bilan du stage	61

Annexe page 20

1. Objectif général de la formation :

A. Nos objectifs en tant que formateurs :

Le public auquel nous nous adressons est constitué de professeurs en activité dans le secondaire et qui ont ou auront en charge l'enseignement des statistiques/probabilités en seconde générale. Beaucoup d'entre eux ont reçu une formation sur ce sujet. Celle-ci n'est pas nécessairement dans l'esprit des programmes actuels. Les enseignants/stagiaires se posent des questions sur le pourquoi et le comment de l'enseignement de cette partie du programme.

Il nous semble essentiel que les enseignants durant ce stage :

- prennent conscience de l'importance de la place des statistiques et perçoivent l'intérêt de faire des statistiques en seconde,
- s'approprient les processus de modélisation et l'intérêt des simulations,
- inscrivent leurs objectifs de séances dans le cadre d'un apprentissage plus large, celui défini par les instructions officielles,
- puissent transposer ces savoirs dans leur classe.

C'est pourquoi notre objectif sera de donner la possibilité aux stagiaires/professeurs de créer des situations d'apprentissage pour que les élèves de seconde puissent, en accord avec les objectifs généraux des instructions officielles, comprendre les trois phases expérimentation, modélisation et simulation.

B. Ce qui est annoncé aux stagiaires :

Objectif du stage : permettre l'élaboration par les stagiaires/professeurs de situations d'apprentissage à propos des statistiques pour les élèves de seconde.

Modalités : Après avoir testé des activités proposées par les formateurs, les stagiaires/professeurs en élaboreront eux-mêmes. Le travail de groupe sera privilégié.

C. Concrètement, il est important de :

- Annoncer le plan au début des séances pour exposer fond et forme.
 - Débuter la séance par un temps d'échange pour aborder tout aspect relatif à la formation (question, échange d'expériences, d'informations et de références).
 - Prévoir une bibliographie incluant des liens internet en vue de prolonger la réflexion.
 - Répartir précisément le rôle des formateurs : par exemple, l'un est chargé de la forme, il anime la séance tandis que l'autre se concentre sur le fond, apporte des arguments, enrichit le débat.
- En ce qui concerne notre posture de formateur, il nous semble important de :
- Ne pas être prescriptif : il n'y a pas de (bonnes) recettes pour enseigner.
 - Ne pas considérer qu'on peut répondre simplement et rapidement aux attentes exprimées des stagiaires.
 - Ne pas considérer les enseignants qui vont constituer le groupe de stagiaires comme des étudiants classiques, mais comme des adultes, des pairs qui sont dans un processus de formation qui devra se poursuivre tout au long de leur carrière.

Ces objectifs et modalités seront communiqués aux stagiaires avant le début du stage (voir annexe : fac-similé du dispositif de formation).

2. Caractéristiques

Public: une vingtaine d'enseignants de mathématiques en lycée

Durée: six demi journées (un mardi en février puis le jeudi de la semaine suivante et pour finir un mercredi vers la fin mai)

Matériel: rétroprojecteur, transparent, affiches vierges, un ordinateur disposant d'un tableur par personne, une calculatrice programmable par personne, un dé par personne, 20 manuels scolaire divers de mathématiques de seconde, 20 jeux des 10 fiches activités (voir documents annexes).

Lieu: à déterminer

Nombre de formateurs: deux, avec éventuellement des intervenants extérieurs (universitaires, inspecteurs, chercheurs, enseignants...)

Documents fournis et ressources : fiches activités élèves (venant des formateurs, voir documents annexes) et celles construites par les stagiaires, une bibliographie (articles d'universitaires), brochure IREM Paris Nord, sites internet.

Annexe page 21

3. Trame du scénario de formation

Les formateurs se présentent. Ils reformulent avec les stagiaires les objectifs de la formation. Ils proposent ensuite que la formation se déroule en plusieurs temps dont le contenu peut évoluer suivant les attentes des stagiaires.

Ainsi, pour le début du stage, une séance d'initiation à un tableur peut être mise en place.

Dès le départ une liste des différents méis est programmée pour optimiser la communication. Cette liste sera donc communiquée aux stagiaires afin qu'ils puissent échanger sur les séances qu'ils auront à construire et/ou à tester, et cela par le biais des formateurs.

• Premier jour

1^{er} temps : Travail en groupes sur différentes activités élèves proposées par les formateurs sur la simulation en seconde (voir documents annexes).

2^{ème} temps : Discussion, synthèse autour du travail effectué dans le 1^{er} temps.
Distribution d'article(s) à lire pour la 2^{ème} journée.

• Deuxième jour

3^{ème} temps : Discussion autour des articles. Analyse d'exercices de manuels scolaires

4^{ème} temps : Construction d'activités sur la simulation en seconde (devoir à la maison; contrôle en classe; séance de cours, de module, d'aide)

• Troisième jour

5^{ème} temps : Bilans et prolongements.

• Articulation des différents jours

Chaque jour est construit autour d'un pôle cohérent en lui-même et chacun de ces pôles s'articule avec les autres pour former un ensemble cohérent qui doit amener les stagiaires à pouvoir atteindre les objectifs fixés.

Le premier jour, mardi, a pour but de :

- créer une dynamique de groupe;
- installer les participants dans une posture de recherche;
- faire une mise au point sur les connaissances et rassurer chacun quant à celles qui sont requises;
- comprendre l'esprit des programmes, en percevoir les objectifs, les replacer dans le cursus scolaire des élèves.

Le deuxième jour, jeudi de la semaine suivante, a pour but de :

- provoquer un regard critique sur ce qui est proposé dans les manuels;
- construire une séance (voire une séquence);
- installer ainsi chacun dans une position d'échange et de recherche constructive.

Le troisième jour (deux mois plus tard) a pour but de :

- partager les expériences (puisque les séquences élaborées le 2^{ème} jour ont dû être testées)
- proposer des prolongements;
- faire un bilan de la formation, communiquer avec d'autres acteurs du système éducatif (chercheurs, inspecteurs, etc...).

L'intervalle de 8 jours entre le 1^{er} et le 2^{ème} jour de la formation est :

- suffisamment court pour que chacun reste dans la dynamique de formation, sans être trop happé par le travail quotidien,
- suffisamment long pour que chacun ait pu lire au moins certains des articles proposés.

L'intervalle de 2 mois entre le 2^{ème} jour et le bilan doit permettre à chacun d'expérimenter ce qui a été construit au 2^{ème} jour.

Une journée entière de bilan nous semble possible car après une longue interruption de la formation, les stagiaires ont besoin d'un certain temps pour se replonger dans le sujet.

Annexe page 22

1) Exercices analysés (séance 1 et 2)

- Exercices de manuels pages 24 à 28
- Exercices EVAPM avec codages et résultats..... pages 29 à 34
- Exercices extraits de la base EVAPMib (ctug48.univ-fcomte.fr/evapmib/). Les exercices analysés pendant le Master sont les exercices : 575 (page 41), 651 (page 42), 697 (page 43), 713 (page 43), 732 (page 45), 904 (page 50) et 922 (page 50) pages 35 à 58

Belin 2000 p 182

20 ★ La composition linguistique de la Belgique est indiquée ci-après :

Néerlandais : 58,5 % ; Français : 33,6 % ;
 Allemand : 0,6 % ; Autres : 7,3 %.

Si on représente les belges francophones par un carré de côté 10 cm, déterminer les côtés des carrés représentant les trois autres communautés linguistiques sachant que les aires des carrés sont proportionnelles aux effectifs.

Délic 2000 p 192

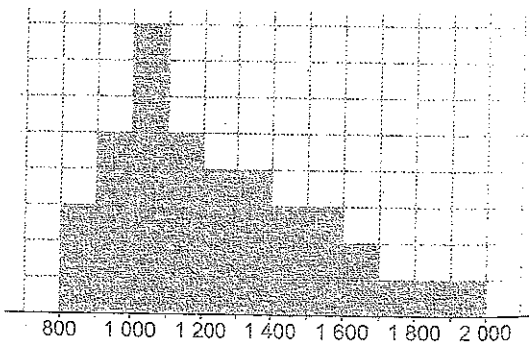
21 ★ Soit trois triangles dont les côtés mesurent respectivement : (6, 8, 10) ; (5, 12, 13) ; (8, 15, 17).

- 1° Ces triangles sont-ils des triangles rectangles ?
- 2° a) Calculer la moyenne des côtés les plus courts de ces trois triangles, puis la moyenne des côtés moyens, puis celle des côtés les plus longs.
- b) Le triangle ayant ces trois moyennes comme côtés est-il rectangle ?



Délic 2004 p 206

On considère la répartition des salariés d'une entreprise selon le salaire net mensuel en €, donnée par l'histogramme ci-dessous, où un carreau représente 10 salariés.



- 1° Donner le tableau des fréquences et des fréquences cumulées.
 - 2° Faire un regroupement en classes d'amplitude 200 à partir de 800 €.
- Représenter la nouvelle série par un histogramme où un carreau représente 10 salariés.
- Expliquer comment réaliser cet histogramme à partir du précédent par une simple manipulation géométrique.



Des cubes

Modulo 2004 p 194

On considère un cube d'arête a (en cm) ; on note S l'aire (en cm^2) d'une face et V le volume (en cm^3).

1. Recopier et compléter le tableau suivant :

a	2	3	5	6	7	9
S						
V						

2. Calculer la moyenne des arêtes, la moyenne des aires et la moyenne des volumes.
3. La moyenne des aires est-elle le carré de la moyenne des arêtes ?
4. La moyenne des volumes est-elle le cube de la moyenne des arêtes ?

Tailles

Préal 2000 p 160

On a mesuré les tailles en mètre des 26 élèves d'une classe :

Albert	Ben	Carole	Dan	Elsa	Franck	Gilles
1,81	1,85	1,64	1,75	1,66	1,88	1,86
Hervé	Inès	Jean	Karine	Leila	Max	
1,83	1,73	1,76	1,72	1,65	1,74	
Noé	Olga	Pierre	Quentin	Raoul	Sylvie	Tarek
1,83	1,75	1,77	1,78	1,82	1,62	1,89
Ulla	Valérie	William	Xavier	Yseult	Zoé	
1,71	1,69	1,85	1,88	1,70	1,72	

On pourra utiliser un tableur ou une calculatrice.

- Déterminer la taille moyenne et la taille médiane.
- Si l'on exprime toutes les tailles en cm, que deviennent la moyenne et la médiane ? Peut-on généraliser cette propriété ?
- Déterminer la taille moyenne des filles et la taille moyenne des garçons. Peut-on, à partir de ces deux valeurs, retrouver la taille moyenne de la classe ? A-t-on la même propriété pour la médiane ?
- Si l'on supprime la plus grande et la plus petite tailles, que deviennent la moyenne et la médiane ?
- Toutes les valeurs sont supérieures à 1,60 m. Si l'on considère les surplus par rapport à 1,60, comptés en cm : 21, 25, 4, 15, etc., peut-on retrouver la moyenne ? la médiane ? Peut-on généraliser cette propriété ?

Saut en hauteur

Préal 2000 p 160

Le professeur d'EPS compare les résultats de ses élèves en saut en hauteur.

- Dans une classe A, au premier trimestre, la moyenne des 20 garçons est 1,40 m, la moyenne des 10 filles est 1,25 m. Quelle est la moyenne de la classe ?
- Au deuxième trimestre, la moyenne des 20 garçons est passée à 1,45 m, et celle des 10 filles est passée à 1,30 m. Quelle est la nouvelle moyenne de la classe ?
- Dans une autre classe B, la moyenne des 5 garçons est 1,45 m, et celle des 25 filles est 1,30 m. Quelle est la moyenne de cette classe ? Comparer avec les deux résultats précédents.

Fractal 2000 p 22

21. On a relevé la taille de dix joueurs de basket : 1,80 m ; 1,95 m ; 1,97 m ; 1,88 m ; 1,98 m ; 2,0 m ; 2,05 m ; 1,87 m ; 1,94 m ; 2,0 m.
- Calculer la taille moyenne en utilisant la taille 1,80 m pour référence.
 - Calculer la taille moyenne en utilisant la taille 2 m pour référence.
 - Calculer la taille moyenne en utilisant la taille 1,90 m pour référence.

Médianes *Dima Hine 2000 p 25-29*

22. Dans une classe de 30 élèves, il y a 15 filles et 15 garçons. On a relevé la taille en cm des filles et des garçons, et obtenu les listes suivantes :

Filles : 155, 164, 162, 168, 158, 164, 152, 160, 158, 162, 155, 160, 158, 160, 150

Garçons : 161, 159, 168, 170, 170, 159, 162, 167, 168, 165, 174, 170, 168, 171, 172

- Quel traitement faut-il faire subir à chacune de ces deux listes pour trouver leur médiane ?
- Calculer la médiane des tailles des filles, puis celle des garçons.
- Calculer la médiane des tailles des élèves de la classe.
- Peut-on calculer la médiane de la taille des élèves de la classe en faisant la moyenne de la médiane de la taille des filles et de celle de la taille des garçons ? (Cet exercice peut se traiter en utilisant un tableur).

Fractal 2000 p 22

24. 1) Calculer la moyenne des nombres : 13 ; 15 ; 8 ; 12 ; 9 ; 9 ; 14 ; 8.
- 2) Calculer la moyenne obtenue en enlevant 2 à chacun des nombres de la liste.

27

Un candidat a obtenu à un examen les notes suivantes :

Écrit : 07 ; 08 ; 08 ; 10 ; 12 ; 14.

Oral : 10 ; 12 ; 13 ; 15.

- Calculer sa moyenne à l'écrit et sa moyenne à l'oral.
- Calculer sa moyenne générale sachant que l'écrit a pour coefficient 5 et l'oral 2.

Podeto 2004 p 189

Delagrave 2000 p 130

25 Contrôle

Voici la distribution des notes obtenues à un contrôle de statistiques.

Note	6	8	9	10	12	14	18	20
Effectif	1	8	3	2	5	3	2	1

- Calculer la moyenne et l'étendue de cette série. Déterminer sa médiane.
- Regrouper cette série en classes d'amplitude 2 : $[0 ; 2]$, $[3 ; 5]$, ..., $[18 ; 20]$ et la représenter par un histogramme.
- En prenant le centre des classes comme valeur représentative, calculer la moyenne de la série. Calculer l'écart du résultat avec la moyenne de la série.
- Comment modifier les notes pour que la moyenne soit de 10 ?
- Calculer la moyenne et l'étendue de la série élaguée des notes extrêmes. Indiquer son étendue.
- En modifiant le moins possible de notes dans la série initiale, proposer une série ayant la même étendue, la même moyenne et la même médiane.

Delagrave 2000 p 134

26 Une ou plusieurs solutions ?

Un professeur de gymnastique mesure la taille de ses 10 gymnastes avec qui elle veut réaliser une figure acrobatique. La plus petite mesure 1 m 58; la plus grande 1 m 78. trois gymnastes mesurent 1 m 65. Elle sait de plus que la taille modale est 1 m 65, que la taille moyenne est 1 m 70, que cinq gymnastes mesurent moins de 1 m 69 et que les cinq autres mesurent plus de 1 m 69.

- Donner une taille possible de toutes les gymnastes.
- Y a-t-il d'autres solutions ?

Belin 2000 p 188-189

513 Une entreprise de services à domicile en plomberie et électricité a établi le relevé suivant de ses interventions journalières pour une période de 52 jours ouvrables:

Nombre d'interventions	14	15	16	17	18	19	20	21	22	23	24	25
Nombre de jours	1	2	4	4	5	7	8	7	6	5	2	1

- Effectuer la représentation en bâtons de cette série.
 - Déterminer la médiane et les quartiles Q_1 et Q_3 .
2. On effectue le regroupement des effectifs par classes de largeur 3 : $[14 ; 17[$; $[17 ; 20[$; $[20 ; 23[$; $[23 ; 25]$.
- Dresser le tableau des effectifs croissants et tracer le polygone correspondant.
 - Déterminer graphiquement puis par le calcul la médiane et les quartiles Q_1 et Q_3 .

Dimathém 2000 p 29

514 Comparaison de la dispersion des populations dans deux régions françaises
Cet exercice utilise les tableaux de l'exercice 25

1. Déterminer l'étendue des populations des 8 départements de la région Midi-Pyrénées.

Procéder de même pour les 8 départements de la région Rhône-Alpes.

2. Dans les deux régions, il y a un ou plusieurs départements beaucoup plus peuplés que les autres. Retirer de la liste ce ou ces départements et refaire le même travail qu'à la question 1.

515 Les deux tableaux ci-dessous donnent la population des 8 départements de la région Midi-Pyrénées et des 8 départements de la région Rhône-Alpes (source : Recensement INSEE 1999).

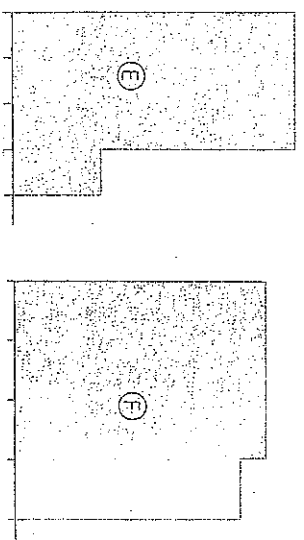
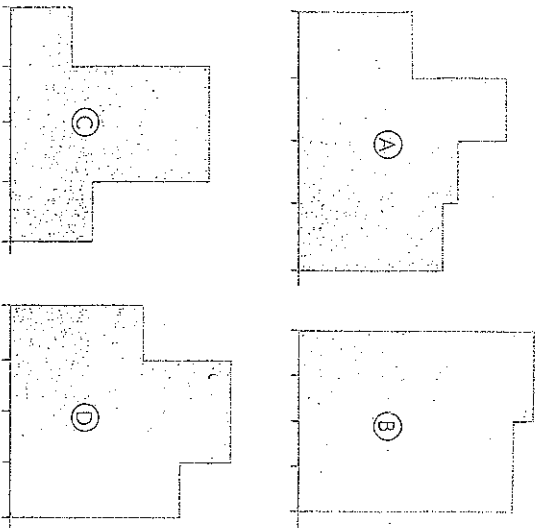
Région Midi-Pyrénées			
Département	Population	Département	Population
Ariège	137 200	Lot	159 700
Aveyron	263 300	Hautes-Pyrénées	222 000
Haute-Garonne	1 046 200	Tarn	342 400
Gers	174 900	Tarn-et-Garonne	205 800

Région Rhône-Alpes			
Département	Population	Département	Population
Ain	514 700	Loire	728 100
Ardeche	286 000	Rhône	1 575 100
Drôme	436 500	Savoie	372 400
Isère	1 090 600	Haute-Savoie	631 000

Diamètres de noix *brezel 2000 P 158*
 On a prélevé 100 noix sur un arrivage pour mesurer leurs diamètres. Le tableau ci-dessous donne les résultats (en mm) :

12	27	15	32	23	15	11	9	39	12
17	34	32	18	11	8	24	27	16	14
35	39	17	22	39	8	24	29	9	37
8	29	27	15	34	18	20	11	9	11
33	10	9	9	12	18	32	30	20	12
17	14	23	38	17	12	16	24	8	9
38	34	20	8	30	11	12	9	8	15
28	34	29	11	20	8	13	21	9	16
22	33	8	11	24	29	36	9	9	21
39	12	30	37	18	24	16	39	24	21

Cette statistique est représentée ci-après par divers histogrammes. Certains sont exacts, d'autres faux. Lesquels et pourquoi ?



La consommation d'énergie en mtep (millions de tonnes d'équivalent pétrole) de six pays d'Europe est indiquée, pour 1997, dans le tableau suivant. (Source OCDE) :

Allemagne	Belgique	Espagne	France	Grèce	Italie
244,34	40,41	74,93	61,16	17,96	125,45

De 1987 à 1997, la consommation a baissé en Allemagne de 5,4 %, mais a augmenté en Belgique de 25,4 %, en Espagne de 47,8 %, en France de 16,7 %, en Grèce de 33,8 % et en Italie de 14,6 %.

a. Calculer la consommation de chacun de ces pays en 1987.
 b. Calculer la consommation moyenne par habitant en 1997, dans chacun de ces pays, puis dans l'ensemble de ces six pays.

Podolski 2004 P 189

12. On a interrogé les 31 élèves d'une classe de seconde sur leur taille (en cm) et on a obtenu les réponses suivantes :
 165 ; 182 ; 174 ; 166 ; 175 ; 168 ; 160 ; 172 ; 184 ; 169 ; 170 ;
 168 ; 177 ; 162 ; 167 ; 171 ; 176 ; 169 ; 165 ; 167 ; 175 ; 171 ;
 168 ; 169 ; 180 ; 175 ; 162 ; 178 ; 162 ; 173 ; 172

Déterminer les modes, la médiane, la moyenne arrondie à 10⁻¹ et l'étendue de cette série.

Frachale 2000 P 21

13. En 1985, en France métropolitaine et dans les Dom, on comptait 257 826 bacheliers, toutes séries confondues.
 En 1998, le nombre de bacheliers est passé à 501 516. Calculer le pourcentage d'augmentation du nombre de bacheliers.

Pythagore 2000 P 235
 a) Les étudiants d'une école de commerce ont relevé les prix de vente pour une même raquette de tennis, dans divers magasins.

Prix p	Nombre de magasins
$300 \leq p < 320$	2
$320 \leq p < 340$	3
$340 \leq p < 360$	4
$360 \leq p < 380$	5
$380 \leq p < 400$	21
$400 \leq p < 420$	9
$420 \leq p < 440$	11
$440 \leq p < 460$	5
$460 \leq p < 480$	4
$480 \leq p < 500$	6

Quel est le prix affiché moyen (à 1 F près) de cette raquette ?
 b) Les clients étant généralement attirés par les magasins les moins chers, les chiffres de ventes sont finalement les suivants.

Prix p	Nombre de raquettes vendues
$300 \leq p < 320$	2570
$320 \leq p < 340$	1780
$340 \leq p < 360$	2300
$360 \leq p < 380$	1250
$380 \leq p < 400$	780
$400 \leq p < 420$	510
$420 \leq p < 440$	90
$440 \leq p < 460$	110
$460 \leq p < 480$	60
$480 \leq p < 500$	250

Quel est le prix de vente moyen d'une raquette ?

Point 2000 P 204

19. La moyenne de l'ensemble des élèves des cinq classes de seconde à un devoir commun de français est de 10,74.

Quatre des classes connaissent leurs résultats :
 seconde 1 (35 présents) – moyenne $\approx 11,45$;
 seconde 2 (25 présents) – moyenne $\approx 9,34$;
 seconde 3 (32 présents) – moyenne $\approx 10,75$;
 seconde 5 (20 présents) – moyenne $\approx 9,85$.

Quelle est la moyenne des 34 élèves de la classe de seconde 4 ?
 Fractions 2000 P 22
 La moyenne de cinq notes d'un élève est 12. Les quatre premières notes sont 13, 10, 8 et 15.
 Quelle est la cinquième ?

Examen d'entrée Hypothèse 2000 P 25
 Un candidat à un examen a passé les quatre premières épreuves ci-dessous.

Matière	Mathématiques	Français	Hist.-Géo.	Langue
Coefficient	3	3	2	1

Sa moyenne est 9,7. Il lui reste à passer l'épreuve d'éducation physique qui est affectée du coefficient 2.
 Quelle note doit-il obtenir pour que sa moyenne finale soit supérieure à 10 ?

1.0 Optimisme

« Pour l'instant, j'ai 13 de moyenne en mathématiques. Avec un 20 au prochain contrôle, je monte à 14. »
 Combien y a-t-il eu de contrôles pour l'instant ?

Martial le malheureux ! Repère 2004 P 2

Martial fait les courses avec ses parents dans une grande surface. Pour réviser son contrôle de statistique, il décide de calculer avec le ticket de caisse le prix moyen d'un article.

Il trouve 4,37 € de moyenne pour 73 articles achetés. Malheureusement, pendant qu'il faisait ses calculs, sa mère revient avec 2 CD audios qui ont coûté 16 € et 23 €. Or il ne les a pas comptés dans sa moyenne.

Comment Martial peut-il faire pour trouver le prix moyen des 75 articles achetés sans refaire tous les calculs ?

Fractions 2000 P 23

28. Un élève a obtenu les notes suivantes au premier trimestre : 8 ; 9 ; 8 ; 14 ; 7 ; 15.
 1) Calculer sa moyenne.
 2) Au deuxième trimestre, il améliore chaque note de 20 %. Calculer sa nouvelle moyenne. Trouver une relation entre les deux moyennes.

Douce inconnue Repère 2004 P 2

Trouver un réel x positif tel que la moyenne des nombres $-3 ; 4 ; 2 ; 7 ; x$ et $\sqrt{2}$ soit x .

Ce tableau donne la répartition des prix de micro-ondes dans différents points de vente.

Prix x (en euros)	Effectif
€300 ; 400€	10
€400 ; 500€	32
€500 ; 600€	25
€600 ; 700€	13

Calculer le prix moyen des fours à micro-ondes :
 a) en posant $y = \frac{x}{100}$
 b) en posant $z = \frac{x - 500}{100}$

Hypothèse 2000 P 26

La moyenne de 3 nombres a, b et c est 29 ; celle des nombres a et b est 22 et celle des nombres b et c est :
 Quels sont ces 3 nombres ?

Question STA001

Voici un tableau donnant, en millions de tonnes, la production mondiale de quatre céréales pour cinq années.

Utiliser ces données pour répondre aux questions ci-dessous :

année	blé	riz	orge	maïs
1962	250	227	88	204
1972	343	307	133	309
1982	476	422	161	449
1992	565	528	165	533
2000	583	598	133	594

Source : Données F.A.O.

a) Parmi ces 4 céréales, celle dont la production est la plus élevée est-elle la même pour chaque année étudiée ? Préciser.

21

b) Quelle est la céréale dont la production a le plus augmenté, en millions de tonnes, entre 1962 et 2000 ?

22

c) La production de chacune de ces 4 céréales augmente-t-elle d'une année étudiée à la suivante ?

23

d) Calculer en pourcentage l'accroissement de la production mondiale de blé entre 1962 et 2000.

24
25

STA001

Item Identification	Conditions d'attributions du code 1
1 a) R.E.	Non. Justifié : Plus forte production : le blé de 1962 à 1992 puis le riz en 2000.
2 b) R.E.	Le maïs.
3 c) R.E.	Non. Justifié : La production d'orge a baissé entre 1992 et 2000.
4 d) R.E.	133,2% ou 133%. On accepte toute réponse exacte à 1 près.
5 d) Démarche	Calcul posé : $(583-250)/250$.

	21	22	23	24	25
STA001	0	0	0	0	0
A	01	02	03	04	05
...	75%	88%	73%	18%	23%

Question STA002ter

Le tableau ci-contre résume une petite enquête effectuée dans les 11 classes de seconde d'un lycée, concernant le nombre de livres que chaque élève a lus pendant les vacances de l'été 2002.

nombre de livres lus	nombre d'élèves
0	45
1	54
2	120
3	78
4	23
5	13
6	11
7	12
8	9

a) Combien d'élèves ont lu au moins 4 livres pendant vacances de l'été 2002 ?

18

b) Combien y a-t-il d'élèves en seconde dans ce lycée sachant qu'ils ont tous répondu à l'enquête ?

19

c) Quelle est l'étendue de cette série ?

20

d) Quelle est la moyenne de cette série ?

21

e) Quelle est la médiane de cette série ? Justifier.

22
23

f) Quel est le mode de cette série ? Justifier.

24
25

STA002ter

Item	Identification	Conditions d'attributions du code 1
1	a) R.E.	68
2	b) R.E.	365
3	c) R.E.	8
4	d) R.E.	2,485. Accepter toute réponse exacte à 0,1 près.
5	e) R.E.	2
6	e) R.E.	Justification correcte
7	f) R.E.	Réponse : 2
8	f) R.E.	Justification correcte

STA002ter	18	19	20	21	22	23	24	25
A	01	02	03	04	05	06	07	08
100%	68%	85%	30%	34%	24%	20%	29%	26%

Question STA003

- a) Fabriquer une série de 13 nombres pris dans l'intervalle $[0; 10]$ telle que sa médiane soit inférieure à sa moyenne.
 b) Fabriquer une série de 13 nombres pris dans l'intervalle $[0; 10]$ telle que sa médiane soit 5 et telle que sa moyenne soit la plus élevée possible.
 Donner la valeur de cette moyenne maximale.
 c) Une série de 13 nombres, tous dans l'intervalle $[0; 10]$, peut-elle avoir pour médiane 5 et pour moyenne 2,5 ?

13
14
15
16
17

STA003

Item	Identification	Conditions d'attributions du code 1
1	a) R.E.	La série proposée vérifie la condition imposée.
2	b) R.E.	La série proposée est la série de moyenne maximale : 7 valeurs 5 suivies de 6 valeurs 10.
3	b) R.E.	La moyenne maximale donnée en b) est exacte, à 0,1 près (valeur exacte : 7,308...)
4	c) R.E.	Non.
5	c) R.E.	Justification correcte. On peut s'attendre au calcul de la moyenne minimum pour une telle série : elle est de 2,69 environ. Accepter toute justification complète.

	10	11	12	13	14
STA003	0	0	0	0	0
A	01	02	03	04	05
	86%	23%	32%	44%	18%

On veut simuler, à partir de données obtenues avec la touche random de la calculatrice, une promenade aléatoire sur les sommets d'un carré ABCD.

Un « déplacement élémentaire » se fera le long d'un côté du carré, d'un sommet à un sommet voisin, et on choisit au hasard à partir de chaque sommet un des deux « déplacements élémentaires » possibles. Par exemple, à partir du sommet B, les deux déplacements élémentaires possibles sont B-C et B-A

Le point de départ d'une promenade aléatoire est toujours D.

Exemple de promenade aléatoire de longueur 4 partant de D et arrivant à B : D-C-B-C-B

I) La touche random de la calculatrice a fourni le nombre : 0,9435974025

Trouver deux méthodes différentes pour générer à l'aide de ce nombre une promenade aléatoire de longueur 5 partant de D.

Décrire chaque méthode et construire la promenade correspondante.

Votre méthode peut ne pas utiliser tous les chiffres.

II) Un jeu consiste à faire une promenade aléatoire de longueur 5 à partir de D.

Une partie est gagnante si la promenade arrive en A.

a) Est-ce possible ?

b) Voici 10 nombres fournis par la touche random d'une calculatrice :

0,9083188611 0,3393625254 0,1468878292 0,7338123112
0,9439919875 0,2003402018 0,9954663411 0,7980701009
0,4058096418 0,5147019505

Choisissez une des modélisations imaginées en I) et fabriquez ainsi 10 parties.

Combien sont gagnantes ?

c) Cette manipulation vous conduit, peut-être à une conjecture intéressante. Laquelle ?

Une conjecture est une propriété que l'on pense vraie sans l'avoir démontrée.

18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	

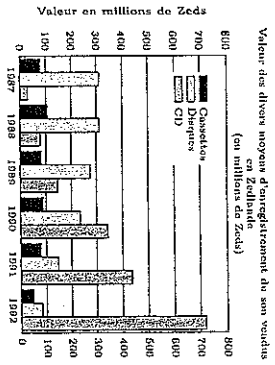
STA004

Item	Identification	Conditions d'attributions du code 1
1	I) R.E.	Les deux méthodes proposées sont correctes et différentes (elles ne génèrent pas toujours le même chemin.
2	I) R.P.	Une seule méthode correcte.
3	I) Démarche	La partie a été utilisée.
4	I) Démarche	La comparaison à 5 a été utilisée.
5	II) a) R.E.	Réponse exacte.
6	II) a) R.E.	La réponse est justifiée par un exemple.
7	II) b) R.E.	Application correcte de l'algorithme, sans erreur.
8	II) b) R.P.	Application correcte de l'algorithme, avec 1,2, ou 3 erreurs.
9	II) c) R.E.	Conjecture : ou arrive forcément en A ou en C.
10	Démarche	Conjecture exacte sur les probabilités d'arriver en A ou en C.
11	Démarche	Découverte d'un algorithme permettant de prévoir rapidement le point d'arrivée.

	15	16	17	18	19	20
STA004	0	0	0	0	0	0
A	01	02	03	04	05	06
	39%	13%	10%	12%	24%	17%

Question STA011

Les graphiques donnent des informations sur les ventes de CD et d'autres moyens d'enregistrement du son en Zedlande. Le Zed est l'unité monétaire de la Zedlande.



Ventes observées selon les âges en 1992

À l'aide des deux graphiques, calculez combien les jeunes entre 12 et 19 ans ont dépensé d'argent en CD en 1992. Montrez votre travail.

STA011

Item	Identification	Conditions d'attributions du code 1
1	R.E.	86,4 millions de zeds. Explication ou méthode indiquée.
2	R.E.	Réponse donnée dans l'intervalle de 84 à 87,6 millions de zeds. AVEC explication ou présentation de méthode.
3	R.P.	Réponse donnée dans l'intervalle de 84 à 87,6 millions de zeds, SANS explication NI présentation de la méthode.
4	Erreur	Réponse donnée dans l'intervalle de 84 à 87,6 zeds, c'est-à-dire oubli du facteur million, AVEC explication ou présentation de méthode.
5	Erreur	Réponse en dehors de l'intervalle précédent, due à une erreur de virgule, AVEC explication ou présentation de méthode.

	22	23	24	25	26
STA011	0	0	0	0	0
A	01	02	03	04	05
	25%	23%	1%	6%	5%

Question STA013

Dans une entreprise, la répartition des salaires est la suivante :

- salaires compris entre 2 000 F et 6 000 F : 51 personnes
- salaires compris entre 6 000 F et 10 000 F : 5 personnes
- salaires compris entre 10 000 F et 14 000 F : 5 personnes
- salaires compris entre 14 000 F et 18 000 F : 40 personnes

En utilisant les centres des classes, le patron de l'entreprise a calculé une estimation du salaire moyen. Combien a-t-il trouvé ?

Calculs

Réponse :

Les syndicats disent : « Le salaire médian n'est que de 6 000 F. »

Est-ce possible ? (justifier) (salaire médian signifie : médiane de la série statistique des salaires)

STA013

Item	Identification	Conditions d'attributions du code 1
1	R.P.	Moyenne : Erreur de calcul en cours d'exercice, mais démarche correcte.
2	R.E.	Moyenne : 9346,... (accepter toute valeur arrondie, si calcul exact). Ne pas tenir compte d'un oubli éventuel de l'unité.
3	R.E.	Salaire médian : accepter toute réponse montrant que l'élève a compris la notion de médiane. Par exemple : si le salaire de la 51 ^e personne est 6 000 F Remarquons qu'il n'est que possible que le salaire médian soit égal à 6 000 F, ce n'est pas certain. Ne pas tenir compte d'un oubli éventuel de l'unité.

	38	39	40
STA013	0	0	0
A	01	02	03
	75%	16%	26%

Question STA014

Un journaliste de télévision a montré le graphique suivant en disant : « Il y a eu une énorme augmentation du nombre de vols cette année »

Nombre de vols par an

520
515
510
505

Cette année

l'année dernière

515
510

Considérez-vous que l'affirmation du journaliste est une interprétation correcte de ce graphique ? Expliquez votre réponse.

Item	Identification	Conditions d'attributions du code 1
1	R.E.	NON, justifié par le fait que seul une partie du graphique est présentée.
2	R.E.	NON, justifié par des arguments corrects en termes de rapport ou de pourcentage d'augmentation.
3	R.E.	NON, autres arguments corrects.

STA014	27	28	29
A	01	02	03
	89%	19%	31%

Épreuve AC8 : en QCM

Q06 ♦ Dans une classe de 30 élèves, la moyenne trimestrielle des notes de mathématiques est 9,4. Le professeur décide d'ajouter 1 point aux trois notes les plus basses.

a	La moyenne ne change pas	Oui	Non	Jusp
b	La moyenne augmente de 0,1 point	Oui	Non	Jusp
c	La moyenne augmente mais on ne peut pas savoir de combien	Oui	Non	Jusp
d	La moyenne augmente de 1 point	Oui	Non	Jusp

Q07 ♦ Dans une classe de 30 élèves, la médiane de la série des notes de mathématiques est 10. Le professeur décide d'ajouter 1 point aux trois notes les plus basses, lesquelles restent les plus basses.

a	La médiane ne change pas	Oui	Non	Jusp
b	La médiane augmente de 0,1 point	Oui	Non	Jusp
c	La médiane augmente mais on ne peut pas savoir de combien	Oui	Non	Jusp
d	La médiane augmente de 1 point	Oui	Non	Jusp

Item	Ident.	Conditions d'attribution du code 1
6	Q06	a : NON ; b : OUI ; c : NON ; d : NON
7	Q07	a : OUI ; b : NON ; c : NON ; d : NON

Q06	21	22	23	24
a				
b				
c				
d				
	63%	35%	44%	55%

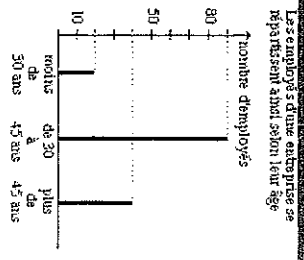
Q07	25	26	27	28
a				
b				
c				
d				
	55%	52%	50%	52%

- Recherche avec les mots
- Fréquence
- Diagramme
- Histogramme
- Ecart-type
- Médiane
- Moyenne
- Statistiques

Clé3

Origine : APMEP
 Sens : Communication
 Type d'activité : Traductif
 Forme : Fermée
 Durée estimée : De 2 à 5 min
 Facilité : 70
 Complexité : A4
 Mots-clés : Graphique; Statistique;
 Capacités : S008;
 Ages : 14; 16;

Quel est le nombre total des employés ?
 Quel est, par rapport à ce nombre total, le pourcentage des employés âgés de 36 à 45 ans ?



EVAPM3/90 - A19 :
 Nombre d'employés R.E. 150 : R = 86 % NR = 03 %
 R = 84 % (orientés en LP)
 Pourcentage R.E. 60%
 on accepte 0,6 et 60 : R = 51 % NR = 17 %
 R = 34 % (orientés en LP)
 EVAPM1/95 - E06 :
 Nombre d'employés R.E. 150 : R = 81 % NR = 02 %
 Pourcentage R.E. 60%
 on accepte 0,6 et 60 : R = 70 % NR = 02 %

page 80

Dans la question A19-20, si l'on rapporte le pourcentage de bonnes réponses du 2° à celles du 1°, on trouve qu'il y a environ 60% des élèves qui savent passer de l'effectif d'une classe à la fréquence de cette classe exprimée en pourcentage.
 Or il est important de savoir calculer un pourcentage. Et les nouveaux programmes, dès la sixième, insistent beaucoup sur cette capacité. Un score de l'ordre de 60 % dans une situation aussi simple que possible, n'est donc pas de nature à nous satisfaire.
 On trouve d'autres calculs de ce type dans les questionnaires composites et dans les questionnaires thèmes, et on observe des résultats inférieurs à ceux étudiés ici, en particulier lorsque les élèves doivent penser eux-mêmes à calculer l'effectif total.
 Les mauvais résultats obtenus tiennent-ils à une difficulté de raisonnement des qu'on ne peut pas faire fonctionner un mécanisme? Il semble d'après des entretiens oraux que l'élève ait besoin de "voir" l'indice 100, "bien placé".

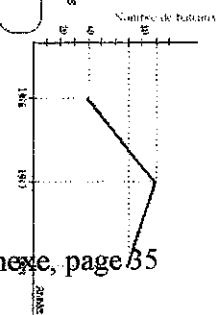
Clé : 10

Origine : APMEP
 Sens : Communication
 Type d'activité : Traductif
 Forme : Fermée
 Durée estimée : De 2 à 5 min
 Facilité : 70
 Complexité : A3
 Mots-clés : Graphique;
 Pourcentage;
 Capacités : S004;
 Ages : 13;
 Voisines : 814

Une usine a fabriqué des bateaux, tous parés, pendant trois ans. Le graphique ci-contre se concerne pour chacune de ces trois années.

1) Combien de bateaux cette usine a-t-elle fabriqués en 1987 ?

2) Parmi les bateaux fabriqués par cette usine au cours de ces trois années, quel est le pourcentage de ceux qui ont été fabriqués en 1987 ?



Annexe page 85

EVAPM4/89 - B36 :
 R.E. 90 : R = 71 % NR = 18 %
 R.E. 43% ou 0,45 : R = 14 % NR = 51 %
 EVAPM4/91 - A31 :
 R.E. 90 : R = 88 % NR = 06 %
 R.E. 45% ou 0,45 : R = 33 % NR = 24 %
 EVAPM4/91 - B36 :
 R.E. 90 : R = 80 % NR = 15 %
 R.E. 45% ou 0,45 : R = 27 % NR = 38 %

voir jumelle Q.C.M. 814

Les résultats de cette question illustrent bien la comparaison entre les passations de 1989 et de 1991 puisque le questionnaire B d'EVAPM4/91 a été repris intégralement d'EVAPM4/89. Les écarts observés pour les taux de non-réponse et les scores de réussite sont significatifs. Il semble que le domaine statistique ait été mieux pris en compte, pour le niveau Quatrième, en 91 qu'en 89. Les autres reprises de 1989 dans ce thème le confirment. Par ailleurs, la différence des scores entre les questions EVAPM4/91 - B36 et EVAPM4/91 - A31 peut étonner. La place de EVAPM4/91 - B36 comme dernière question du questionnaire a certainement joué, et les écarts de réussite sont liés, pour l'essentiel, aux écarts des taux de non-réponse. Si on considère en effet les réussites relatives, le tableau suivant (2) fait apparaître des résultats équivalents.

Tableau 2

B36 B37 A31 A32 C37 C38	80 27 88 33 89 42
5 35 6 43 8 47	15 38 6 24 3 11
94 43 94 43 92 47	

Enfin les résultats à la question EVAPM4/91 - C37 pourraient faire croire que le style QCM réussit nettement mieux aux élèves que le questionnement classique. Mais le tableau 2 ci-dessus relativise les scores. En effet, si les élèves sont plus nombreux à réussir, ils sont aussi plus nombreux à échouer. On peut tout simplement constater que les élèves prennent davantage de "risques" dans un QCM.
 Pour la deuxième partie de cette question EVAPM4/91 - B36, les élèves n'ayant pas répondu, ou mal, à la première, étaient pratiquement hors-jeu. Mais l'échec s'est, ici, lourdement aggravé. L'erreur la plus fréquente est une reprise du résultat du 1°) (90 bateaux) en 90 % ou 0,9. Les réponses fantaisistes ou mal analysables sont nombreuses.
 Il est vrai qu'il fallait traduire un effectif en pourcentage en ayant soi-même à faire intervenir et à trouver l'effectif total. Il y avait donc tout un calcul à organiser.

C1647

Origine : APMEP
 Sens : Communication
 Type d'activité : Traductif
 Forme : Feuille
 Durée estimée : De 2 à 5 min
 Facilité : 70
 Complexité : A3
 Mots-clés : Graphique;
 Pourcentages;
 Capacités : S004;
 Ages : 13;

Occupations des dépenses de la KULTUR en 1988.
 Pour le cinéma, les livres et les disques.
 Elles se sont réparties selon le diagramme circulaire ci-contre.
 Parmi les pourcentages suivants:
 (par rapport au total des dépenses)

5% : 20% : 25% : 50% : 60% : 90%
 choisit celui qui le semble le mieux
 correspondre :

- au cinéma :
 - aux livres :
 - aux disques :

EVAPM4/89 - A22 :
 Les deux réponses exactes : R = 65 % NR = 10 %

EVAPM4/89 - A22 : page 77
 SAVOIR LIRE des données statistiques présentées sous la forme de tableaux ou de diagrammes d'effectifs ou de fréquences. (S 321).
 Dans l'item A22, certains élèves confondent un pourcentage (%) et la mesure de l'angle au centre représentant graphiquement ce pourcentage.
 Ainsi trouve-t-on un pourcentage de "90 %" pour le cinéma...
 Telle est l'erreur la plus fréquente.
 Les élèves qui répondent "90 %" pour le cinéma, auraient, dans leur logique, dû répondre "216 %" pour les livres, ce qui aurait dû choquer leur sens critique. Mais "216 %" ne figurait pas dans la liste proposée. Ces élèves ont donc répondu par n'importe quoi de la liste, par exemple 50 %, ce qui est tout aussi choquant, sinon plus, puisque la dépense en pourcentage serait plus forte pour le cinéma que pour le livre!
 Ce qui montre une absence totale d'esprit critique ou une totale méconnaissance de la signification des pourcentages ou des comparaisons sur graphiques.
 Cela dit, l'exercice est dans l'ensemble bien réussi.
 La comparaison avec l'item P 21-22 (voir plus bas) conduit cependant à s'interroger sur l'appui qu'il a pu apporter la liste proposée. Nous en reparlons.

C1648

Origine : APMEP
 Sens : Communication
 Type d'activité : Classificatoire
 Forme : Feuille
 Durée estimée : De 2 à 5 min
 Facilité : 30
 Complexité : A3
 Mots-clés : Statistique;
 Capacités : S005;
 Ages : 13; 14;

On propose sur 6 cartes 10 données.
 Il a été de 3 notes sur 10 :
 5 ; 8 ; 8 ; 14 ; 5 ; 17 ; 8 ; 14 ; 14 ; 8.
 Prélever systématiquement 5 dans un tableau dans lequel les données sont les suivantes :

Annexe, page 36

EVAPM4/89 - A21 :
 tout tableau regroupant les données
 d'une façon ou d'une autre : R = 24 % NR = 48 %
 EVAPM4/91 - A33 :
 tout tableau regroupant les données
 d'une façon ou d'une autre : R = 36 % NR = 29 %
 EVAPM3/90 - E04 :
 tout tableau regroupant les données
 d'une façon ou d'une autre : R = 54 % NR = 20 %
 EVAPM3/92 - E04 :
 tout tableau regroupant les données
 d'une façon ou d'une autre : R = 52 % NR = 16 %

EVAPM3/90 - E04 : page 82-83
 Lorsqu'il s'agit d'organiser la démarche visant à présenter des résultats dans un tableau qu'il faut construire (question E4), les résultats restent médiocres, mais il faut insister sur le saut considérable existant dans ce domaine, qui a quelque chose à voir avec l'autonomie acquise par l'élève, entre le niveau de compétence observé en fin de quatrième (24%) et celui observé en fin de troisième (34%).
 EVAPM4/89 - A21 : page 78 voir Q8
 A partir de données statistiques, PRESENTTER les effectifs ou les fréquences dans des tableaux. (S 322).
 Environ les 3/4 des élèves ne comprennent pas la question A21 :
 - il y a près de la moitié de non-réponses
 - parmi les élèves qui répondent, une bonne moitié font un tableau qui reproduit telle quelle la suite des notes indiquée, et avec des effectifs fantaisistes : quelques élèves font fâcheusement jouer une mémoire visuelle sur effectifs et effectifs cumulés croissants : ainsi, sous les notes de la suite trouve-t-on successivement 5 ; 13 ; 21 ; 35 ; ... tout cela avec 10 dévots !
 Les bonnes réponses (1/4 des élèves) proposent soit un tableau avec les notes 5 ; 8 ; 14 ; 17 soit des tableaux avec des intervalles (0 < note = 5 ; ...) d'amplitude 5 ou 4 (il y a alors 5 classes, dont une vide).
 Curieusement, il y a excès de zèle pour la moitié environ des bonnes réponses: elles donnent non seulement les effectifs mais aussi les effectifs cumulés croissants, ce qui semble signaler un conditionnement abusif...

Origine : ANRMEP
 Sens : Objet
 Type d'activité : Critique
 Forme : Semi-ouverte
 Durée estimée : De 10 à 20 min
 Facilité : 40
 Complexité : E2
 Mots-clés : Fréquence;
 Probabilité; Statistique;
 Capacités : S020; S043; S045;
 Ages : 16;
 Visines : 10067

Dans une urne, il y a des boules rouges et des boules blanches

On extrait 100 fois de suite une boule de cette urne

Chaque fois, la boule est remplacée dans l'urne et l'ensemble est à nouveau mélangé.

Après chaque tirage, on note la couleur de la boule tirée

Dans ces conditions, on a tiré 60 fois une boule rouge et 40 fois une boule blanche.

a) Cette information est-elle utile pour donner une idée de la composition de l'urne (proportion de boules rouges et de boules blanches) ?

Arguez et justifiez.

b) Voici trois énoncés concernant la situation décrite ci-dessus.

Pour chacun d'eux, indiquez si vous pensez qu'il est vrai ou faux.

Arguez et justifiez.

Énoncé A : La probabilité de tirer une boule rouge change à chaque tirage

Énoncé B : Comme il n'y a que des boules rouges et des boules blanches dans l'urne, si l'on tire une 101^{ème} boule, il y a une chance sur deux pour que cette boule soit rouge

Énoncé C : Toutes les boules de l'urne ayant la même probabilité d'être extraites, la probabilité de tirer une boule rouge est égale à la probabilité de tirer une boule blanche

c) Voici deux énoncés concernant la situation décrite ci-dessus (extraction de 100 boules).

Énoncé D : À chaque tirage, il y a 6 chances sur 10 de tirer une boule rouge

Énoncé E : À chaque tirage, la probabilité de tirer une boule rouge est supérieure à celle de tirer une boule blanche

Quel est, de ces deux énoncés, celui qui est le plus vraisemblable ?

Arguez et justifiez.

EVAPM11/93 - CC22 :
 Question a) réponse OUI
 expliquée ou non
 toutes séries : R = 43 %
 série S : R = 49 %
 série E : R = 46 %
 série B : R = 38 %
 sérieA1 : R = 36 %
 série F : R = 39 %
 réponse NON expliquée ou non
 toutes séries : R = 40 %
 série S : R = 35 %

série E : R = 44 %
 toutes séries : R = 54 %
 série S : R = 46 %
 sérieA1 : R = 41 %
 série F : R = 44 %
 Explication raisonnable pour OUI ou NON
 OUI parce que l'observation faite permet de penser qu'il y a plus de boules rouges que de boules blanches
 toutes séries : R = 12 %
 série S : R = 14 %
 série E : R = 12 %
 série B : R = 15 %

sérieA1 : R = 17 %
 série F : R = 19 %

Question b) énoncé A réponse

toutes séries : R = 73 %

série S : R = 76 %

série E : R = 81 %

série B : R = 77 %

sérieA1 : R = 66 %

série F : R = 64 %

Question b) énoncé B

toutes séries : R = 64 %

série S : R = 71 %

série E : R = 71 %

série B : R = 61 %

sérieA1 : R = 67 %

série F : R = 40 %

Question b) énoncé C réponse

OUI expliquée ou non

toutes séries : R = 28 %

série S : R = 22 %

série E : R = 19 %

série B : R = 31 %

sérieA1 : R = 39 %

série F : R = 30 %

Question b) énoncé B réponse

NON expliquée ou non

toutes séries : R = 58 %

série S : R = 65 %

série E : R = 69 %

série B : R = 57 %

sérieA1 : R = 48 %

série F : R = 50 %

Question b) énoncé B

explication correcte pour NON

toutes séries : R = 45 %

série S : R = 54 %

série E : R = 47 %

série B : R = 40 %

sérieA1 : R = 40 %

série F : R = 27 %

Question b) énoncé C réponse

OUI expliquée ou non

toutes séries : R = 32 %

série S : R = 26 %

série E : R = 33 %

série B : R = 33 %

sérieA1 : R = 37 %

série F : R = 41 %

Question b) énoncé C réponse

NON expliquée ou non

toutes séries : R = 52 %

série S : R = 59 %

série E : R = 55 %

série B : R = 52 %

sérieA1 : R = 44 %

série F : R = 50 %

Question b) énoncé C

explication correcte pour NON

toutes séries : R = 41 %

série S : R = 48 %

série E : R = 37 %

série B : R = 39 %

sérieA1 : R = 39 %

série F : R = 24 %

Question c) D est plus

vraisemblable que E

toutes séries : R = 21 %

série S : R = 19 %

série E : R = 22 %

série B : R = 23 %

sérieA1 : R = 18 %

série F : R = 25 %

Question c) E est plus

vraisemblable que D

toutes séries : R = 69 %

série S : R = 65 %

série E : R = 65 %

série B : R = 55 %

sérieA1 : R = 60 %

série F : R = 49 %

Question c) explication correcte

" l'énoncé E est plus

vraisemblable que D "

toutes séries : R = 36 %

série S : R = 42 %

série E : R = 33 %

série B : R = 33 %

sérieA1 : R = 34 %

série F : R = 26 %

Russite conjointe pour les

explications correctes

toutes séries : R = 10 %

série S : R = 15 %

série E : R = 12 %

série B : R = 07 %

sérieA1 : R = 06 %

série F : R = 04 %

Les programmes mettent l'accent sur le processus de modélisation de situations aléatoires et sur l'introduction de la notion de probabilité, en s'appuyant sur une approche fréquentiste. La question CC22-36 teste plutôt les conceptions des élèves que leurs connaissances, qui sont sur ce sujet en Première assez limitées. La question a) évite l'idée de probabilité et le libellé "donner une idée de la composition de l'urne" permet aux élèves de répondre de manière qualitative. L'opinion des élèves est très partagée : 43% OUI, 40% NON. On semble sentir une résistance à l'approche fréquentiste, qui pourtant est fortement soutenue par les programmes. Tous les élèves sont loin d'accepter facilement la stabilisation d'une fréquence. D'ailleurs on peut s'interroger sur le sens que les élèves donnent à la notion de stabilisation: au bout de combien d'expériences la fréquence semble-t-elle stabilisée? Ici, cent tirages sont-ils suffisants? Ce qui est important ici, ce n'est pas la réponse, mais c'est que l'élève se pose la question, car cela signifie qu'il relativise l'utilité de l'information pour pouvoir construire la composition de l'urne. Ainsi 54% des élèves fournissent une explication cohérente que leur réponse ait été OUI ou NON.

Dans les énoncés B et D, nous avons utilisé de manière délibérée le mot "Chance". Les élèves font-ils une différence entre chance et probabilité? L'énoncé B décrit la même conception (erronée) que l'énoncé C. La différence est le passage du terme chance à celui de probabilité. Le taux de réponses OUI (c'est à dire une réponse fautive) est sensiblement le même pour les deux énoncés, ce qui peut laisser supposer que l'hypothèse de départ n'est pas confirmée: il y a peu de différence pour les élèves entre le mot chance et probabilité en ce qui concerne les énoncés A et B. L'étude implicite montre que les élèves qui répondent correctement à l'énoncé B donnent une réponse correcte à l'énoncé C. Cette assimilation "chance" = "probabilité" révèle le caractère subjectif attaché à cette notion.

Cependant l'étude implicite de cet exercice montre que les élèves qui semblent refuser l'existence de loi de probabilité pensent majoritairement qu'il y a une chance sur deux que la boule soit rouge. Ainsi ils réitèrent

Clé124

l'idée d'un modèle probabiliste et en même temps se comportent comme si en présence de deux éventualités, il y avait une chance sur deux....

De même, ceux qui pensent que la probabilité de tirer une rouge est égale à la probabilité de tirer une bleue affirment la plupart du temps que l'information initiale n'est pas suffisante pour avoir une idée de la composition de l'urne. Ces élèves semblent avoir mis une barrière solide entre la formalisation mathématique et la description du réel.

L'énoncé A semble montrer que pour 15% des élèves, en fin de Première, il n'y a pas de loi de probabilité et que le hasard ne peut pas être quantifié, modélisé. Il faut noter que cette résistance, prévisible, est sensiblement la même dans toutes les séries et elle renforce l'idée que le temps que l'on doit passer sur l'introduction des probabilités n'est pas compressible et ne doit pas être proportionnel aux nombres de lignes écrites dans le programme ou qu'une sensibilisation beaucoup plus précoce serait utile.

L'étude des croisements montre que pratiquement aucun élève répondant correctement à la question portant sur l'énoncé A trouve que l'énoncé D est plus vraisemblable, ainsi les élèves qui semblent accepter des lois probabilistes montrent une grande prudence sur l'utilisation "concrète" de celles-ci.

L'énoncé D a été rédigé avec le mot "chance" pour faire appel au bon sens alors que le mot "probabilité" peut avoir pour conséquence de placer l'élève dans le cadre mathématique. Les élèves privilégient l'énoncé en terme de probabilité (énoncé D) montrant ainsi une confiance bienvenue, alors que les 21% qui donnent leur préférence à l'énoncé A accordent, sans doute une confiance trop grande dans l'estimation fréquentiste de la probabilité ainsi qu'une trop grande précision sur la valeur à retenir pour celle-ci.

L'étude implicite de cet exercice montre que les élèves qui semblent refuser l'existence de loi de probabilité pensent majoritairement qu'il y a une chance sur deux que la boule soit rouge et que la probabilité de tirer une rouge est égale à la probabilité de tirer une bleue".

Origine : APMEP	Sur un parking, on s		A vitres			
Sens : Communication	relevé la marque et	le numéro	marque	numéro		
Type d'activité : Calculatoire	d'immatriculation	(de partement), des	et			
Forme : Feuille	voitures en	stat. On mesurait	les résultats sont	resemblés dans		
Durée estimée : De 2 à 5 min	Facilité : 90	Compléxité : A4	Mois-clés : Statistique;	Capacités : S001;		
Agés : 12;	Combien de voitures sont stationnées dans ce parking ?					
	Deux séries (79)	28	12	42	15	97
	Yenne (86)	12	15	21	10	58
	Yenne (86)	13	8	10	5	36
	Autres départements	17	16	12	13	58
	TOTAL	70	51	85	43	124

Quelle est la marque la plus représentée ?

Parmi les voitures stationnées, combien y-a-t-il de peugeot immatriculées dans le Yenne ?

Dans quel département ont été immatriculées le plus grand nombre de voitures ?

Annexe page 38

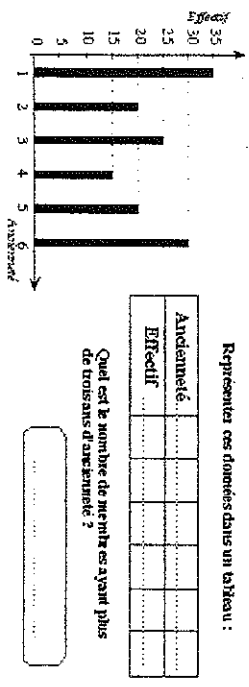
EVAPM5/88 - C08 :
voitures : R = 83 % NR = 05 %
manques : R = 94 % NR = 03 %
Yenne : R = 85 % NR = 03 %
département : R = 92 % NR = 03 %
EVAPM5/90 - C08 :
voitures : R = 87 % NR = 02 %
manques : R = 94 % NR = 02 %
Yenne : R = 89 % NR = 02 %
département : R = 94 % NR = 02 %

La lecture du tableau C ne semble pas avoir posé de difficultés aux élèves. Les pourcentages de réussite sont très élevés et l'exercice est dans l'ensemble très bien réussi. On peut s'étonner du pourcentage plus faible de la question sur les voitures. On trouve fréquemment le nombre 498 qui correspond à 2 fois le total indiqué dans le tableau.

Ref : Brochure 5/1988 p67

Origine : APMEP
 Sens : Communication
 Type d'activité : Traductif
 Forme : Fermée
 Durée estimée : De 2 à 5 min
 Facilité : 70
 Complexité : A3
 Mots-clés : Diagramme; Graphique; Statistique;
 Capacités : S002;
 Ages : 12;

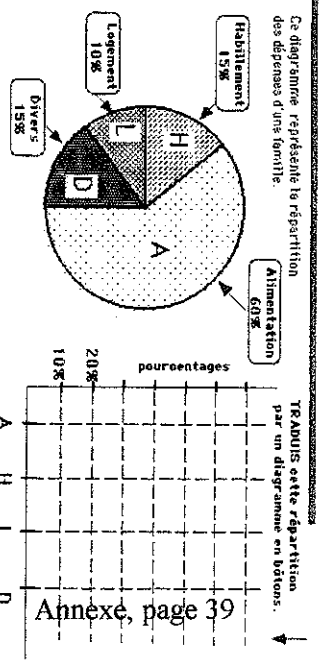
Le diagramme en bâtons représente les résultats d'une enquête.
 La population étudiée est composée de membres d'un club de tennis.
 Le caractère étudié est l'ancienneté au club (en années).



EVAPM5/90 - D23 :
 tableau correctement complété : R = 86 % N.R. = 02 %
 R.E. 65 : R = 38 % N.R. = 07 %

page 90
 La question D23-24 ne manquera pas d'attirer l'attention. En effet, les élèves mis en présence d'un graphique en bâtons sont quasiment tous capable de le traduire par un tableau statistique. Mais quel sens donnent-ils ensuite à ce tableau ? C'est sans doute là que les choses se gâtent !
 A partir d'un tableau correctement rempli, on trouve comme réponses à la question "Quel est le nombre de membres ayant plus de trois ans d'ancienneté ?" :
 0 (ils ont tous plus de trois ans d'ancienneté),
 3 (sans doute parce qu'il y a 3 "bâtons" concernés),
 21 (somme des nombres de la première ligne),
 25 (effectif de la classe "3 ans d'ancienneté"),
 90 (correspond à "au moins 3 ans...")
 "tous" (ils ont tous plus de trois ans d'ancienneté).
 Sans compter les erreurs d'addition qui donnent 60 ou 70 (au lieu de 65).
 Finalement, il n'y a que 38% des élèves qui donnent la réponse exacte. On voit ainsi à quel point la lecture d'un tableau, qui paraissait d'autant plus clair qu'il avait été rempli par les élèves eux-mêmes, a pu mettre les élèves en difficulté.

Origine : APMEP
 Sens : Communication
 Type d'activité : Technique
 Forme : Fermée
 Durée estimée : De 2 à 5 min
 Facilité : 60
 Complexité : A1
 Mots-clés : Graphique;
 Pourcentage; Statistique;
 Capacités : S003;
 Ages : 12;
 Voisines : 482



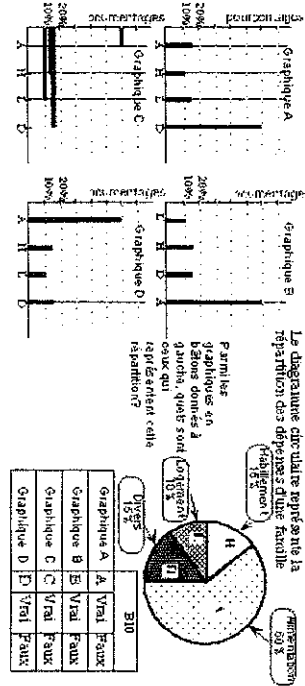
EVAPM5/88 - B23 :
 une erreur au plus de n'importe quelle nature : R = 85 % NR = 05 %
 pas d'erreur du tout : R = 42 % NR = 05 %
 EVAPM5/90 - A18 :
 une erreur au plus de n'importe quelle nature : R = 89 % NR = 02 %
 pas d'erreur du tout : R = 52 % NR = 02 %

voir jumelle Q.C.M. 482
 EVAPM5/88 - B23 : page 68

Pour le questionnaire B, on doit noter le faible taux de réussite. Beaucoup d'élèves n'ont pas traité l'exercice et pour ceux qui ont répondu à la question, il y a confusion entre les différentes représentations graphiques notamment diagramme en barres et polygone statistique.
 On peut s'étonner que la seule compétence exigible retenue porte sur le diagramme en bâtons. En effet, les exemples concrets empruntés à la vie courante trépassent très rarement. De plus, les graphiques présentés dans les journaux et ceux des manuels de géographie présentent des confusions dans l'emploi du vocabulaire. On peut donc avoir tendance à penser que les élèves n'ont pas bien fait la différence entre histogramme et diagramme en bâtons. Il faudrait sans doute que les utilisateurs de relevés statistiques se mettent d'accord sur un vocabulaire de base et nos élèves auraient peut-être une chance de bien interpréter la question posée.
 Ce domaine mérite de notre part une attention plus soutenue d'autant que l'on retrouve le même type de questions dans le nouveau programme de 4ème.
 EVAPM5/90 - A18 : page 89
 La question C8-11 reprise de 88 obtient des résultats très voisins de ceux obtenus à cette époque. Par contre la question A18-19 voit son taux de réussite baisser de façon significative. Cette constatation concernant une question particulière n'est sans doute pas suffisante pour pouvoir affirmer que les compétences développées chez les élèves de Cinquième, dans le domaine des statistiques, sont maintenant inférieures à ce qu'elles étaient il y a deux ans. Elle fait toutefois écho à une information issue du questionnaire destiné aux professeurs. En effet, d'après les réponses à ce questionnaire, le temps consacré à l'enseignement des statistiques se serait en effet réduit de façon significative au cours des deux dernières années.
 En fait, dans A18-19, la plupart des élèves lisent correctement la représentation en secteurs circulaires et les erreurs proviennent essentiellement d'une utilisation abusive des histogrammes. Les histogrammes, on le sait, ne sont justifiés que dans le cas des variables numériques continues. Certes des représentations du type histogrammes (bâtons larges et accolés) peuvent être utilisés dans les autres cas, mais cela risque d'amener les élèves au type d'erreur présentée ci-dessous.
 images de graphiques page 90
 Une petite partie des erreurs est aussi due à l'utilisation d'un polygone d'effectifs. Une habitude contestable s'est répandue dans les médias, mais aussi dans les établissements scolaires (bulletins de notes par exemple) : cette habitude consiste à joindre les points de n'importe quel graphique par une ligne polygonale, même lorsque ce figure n'a aucun sens. Cela est en particulier le cas lorsque la variable est nominale. Cette habitude demande à être interrogée par notre enseignement, mais certainement pas à être encouragée.

CI6482

Origine : APMEP
Sens : Communication
Type d'activité : Classificatoire
Forme : Q.C.M.
Durée estimée : De 2 à 5 min
Facilité : 40
Complexité : A3
Mots-clés : Graphique; Pourcentage;
Capacités : S003;
Âges : 12;
Voisines : 431



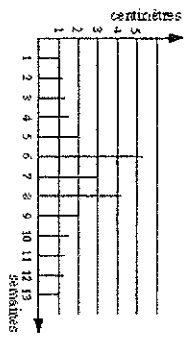
EVAPM5/90 - B10 :
-B- -D : R = 45 % NR = 04 %

voir janelle non Q.C.M. 431

page 90
Il est intéressant de constater que la Q.C.M. (B10) est moins bien réussie que la question classique correspondante. Ici, l'identification est plus difficile que la production. Vérification faite, cet accroissement n'est pas dû au fait qu'il y a deux réponses exactes. En effet, beaucoup d'élèves ont donné A et D comme bonnes réponses, ce qui se comprend mieux si l'on observe que les deux représentations sont symétriques (à une translation près, et à la place des lettres...). Nous n'avons pas poussé la créativité jusqu'à proposer un graphique du type C (battons horizontaux) qui soit exact. Pourtant, il y a tout intérêt à varier les formes de représentations proposées et à habituer les élèves à passer d'une forme à une autre.

CI6556

Origine : I.E.A.
Sens : Objet
Type d'activité : Calculatoire
Forme : Feuille
Durée estimée : De 2 à 5 min
Facilité : 70
Complexité : A4
Mots-clés : Graphique; Moyenne;
Capacités : S004;
Âges : 13;
Voisines : 500186



Sur ce graphique, la hauteur de plus en plus d'unités est relevée pendant 13 semaines.

La hauteur moyenne de pluie tombée par semaine durant cette période est environ :

Encadre la bonne réponse

1 centimètre
2 centimètres
3 centimètres
4 centimètres
5 centimètres

I.E.A. 4/82 E-88:
R.E. "2 centimètres est entouré les autres réponses sont barrées : R = 68 % N.R. = 05 %
France : R = 50 %
Japon : R = 67 %
U.S.A. : R = 46 %
EVAPM4/91 - P31 :
R.E. "2 centimètres est entouré les autres réponses sont barrées : R = 68 % N.R. = 05 %

voir aussi 500186 (ce numéro n'existe pas)

EVAPM4/91 - P31 :
La situation de la question 4P31 nous avait paru très intéressante en ce qui concerne la lecture et l'exploitation d'un tel document. Plus de deux élèves sur trois ont su estimer la hauteur moyenne de pluie à partir d'une lecture de valeurs qui ne pouvaient qu'être approchées. Bien sûr les réponses proposées les ont aidés, mais on peut tout de même saluer la performance, d'autant plus que ce type d'exercice n'est pas habituel pour eux.
Il semble donc que la lecture d'un graphique est toujours aussi bien maîtrisée, sinon mieux, par les élèves de Quatrième, mais que le calcul d'un pourcentage reste une tâche difficile à surmonter malgré les progrès observés.

Clé571

Origine : APMEP
Sens : Objet
Type d'activité : Calculatoire
Forme : Q.C.M.
Durée estimée : De 2 à 5 min
Facilité : 20
Complexité : A4
Mots-clés : Ecart-type;
Capacités : S021;
Âges : 16;

Le tableau ci-dessous donne les effectifs, par âge, d'un club de lycéés.

Âges des élèves	4 ans	15 ans	16 ans	17 ans
Nombre d'élèves par âge	1	3	15	5

Soit σ l'écart type de la série statistique donnée. On peut alors affirmer que :

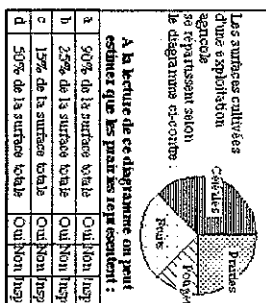
a	$\sigma < 0$	Oui	Non	Jasp
b	$0 \leq \sigma < 1$	Oui	Non	Jasp
c	$1 \leq \sigma < 2$	Oui	Non	Jasp
d	$2 \leq \sigma$	Oui	Non	Jasp

EVAPM1/93 - CA05 :

-b- -
toutes séries : R = 15 %
série S : R = 15 %
série E : R = 08 %
série B : R = 26 %
sérieA1 : R = 07 %
série F : R = 17 %
série G : R = 09 %

Clé575

Origine : APMEP
Sens : Communication
Type d'activité : Calculatoire
Forme : Q.C.M.
Durée estimée : De 2 à 5 min
Facilité : 80
Complexité : B3
Mots-clés : Diagramme; Graphique; Pourcentage; Proportionnalité;
Capacités : S008;
Âges : 14; 15; 16;
Voisines : 1023



EVAPM3/92 - H09 :

-b- - : R = 77 % N.R. = 01 %

EVAPM1/93 - CA07 :

-b- - :
toutes séries : R = 91 %
série S : R = 95 %
série E : R = 93 %
série B : R = 88 %
sérieA1 : R = 89 %
série F : R = 97 %
série G : R = 84 %
a en particulier répondu oui à a
toutes séries : R = 08 %
série S : R = 05 %
série E : R = 03 %
série B : R = 07 %
sérieA1 : R = 07 %
série F : R = 20 %
série G : R = 08 %

voir aussi la question 1023

EVAPM3/92 - H09 :

En troisième/1992, la seule question nouvelle concernant cette partie est la lecture du diagramme circulaire sous forme de Q.C.M. Peu de commentaires de notre part sinon pour constater qu'un élève sur cinq s'est laissé "piéger" par les réponses erronées. Les "pièges" auraient-ils été les mêmes s'il avait été question des fruits ou des céréales ? A suivre ... peut-être !
Ref : Brochure 3/1992 p78

Origine : APMP
 Sens : Objet
 Type diachronique : Traductif
 Forme : Feuille
 Durée estimée : De 5 à 10 min
 Facilité : 70
 Complexité : A3
 Mots-clés : Graphique;
 Statistique;
 Capacités : S024; S025;
 S026;
 Ages : 16;

La production annuelle nationale de station (en milliers de tonnes) est donnée par le tableau suivant :

Année	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
Production	20,7	21,5	20,0	19,6	19,0	16,9	16,7	16,0	14,0	13,9

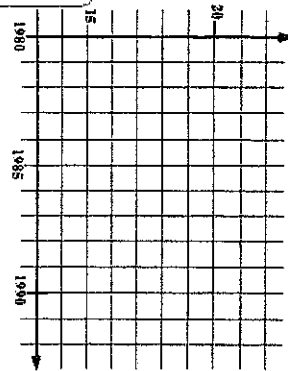
a) Sur le graphique de droite, représentez les données par un nuage de points.

Soit G_1 le point moyen du nuage constitué des 5 points correspondant aux années 1980 à 1984.

Soit G_2 le point moyen du nuage constitué des 5 points correspondant aux années 1985 à 1989.

b) Calculer les coordonnées de G_1 et de G_2 .

c) Placer les points G_1 et G_2 sur le graphique et tracer la droite (G_1, G_2) .
 d) Quelles remarques concernant l'évolution de la production cette droite vous suggère-t-elle ?



EVAPM1/93 - SF12 :
 Nuage : au plus deux erreurs grossières.
 Ne pas attacher trop d'importance à la précision
 série B : R = 72 % N.R. = 12 □ %
 Nuage : aucune erreur
 série B : R = 84 % N.R. = 06 %
 Coordonnées de G_1 : R.E. (1982 ; 20,16)
 série B : R = 58 % N.R. = 12 %
 Coordonnées de G_2 : R.E. (1987 ; 15,5)
 série B : R = 59 % N.R. = 12 %
 Droite $(G_1; G_2)$ correctement tracée
 série B : R = 63 % N.R. = 16 %

Interprétation : Toute remarque indiquant une bonne compréhension de la situation : décroissance ayant tendance à s'accélérer ou chute "décrochage" en 85 non compensée ensuite, ou... Cet item a été ajouté sans expérimentation, mais il était sans intérêt de faire calculer des points moyens et de faire tracer une droite de façon absolument gratuite.
 série B : R = 60 % N.R. = 19 %

La question SF12-17 est la seule qui fasse partie du programme spécifique de la section B, cependant la formulation de celle-ci ne fait appel à aucun concept nouveau. Les pourcentages de réussite montrent que les élèves qui parviennent pratiquement tous à tracer la droite (G_1, G_2) donnent une interprétation pertinente de celle-ci.

Origine : I.E.A.
 Sens : Objet
 Type diachronique : Calculatoire
 Forme : Q.C.M.
 Durée estimée : De 2 à 5 min
 Facilité : 30
 Complexité : A4
 Mots-clés : Moyenne;
 Capacités : S018;
 Ages : 15; 16;
 Voisines : 650

On se sera servi d'un décompte dans deux classes. Le premier classe, composée de 50 élèves, a obtenu une moyenne de 12,50. Le deuxième classe, composée de 30 élèves, a obtenu une moyenne de 14,00. Quelle est la moyenne de chaque classe par les 80 élèves combinés ?

12,50	Oui	Non	Je ne sais pas
13,30	Oui	Non	Je ne sais pas
13,25	Oui	Non	Je ne sais pas
13,80	Oui	Non	Je ne sais pas

I.E.A. Ter. S. 84 item 60
 JAPON : R = 85 %
 USA : R = 32 %
 EVAPM2/91 - R25 :
 -- -d : R = 33 % N.R. = 07 %
 EVAPM1/93 - CA03 :
 -- -d
 toutes séries : R = 40 %
 série S : R = 57 %
 série E : R = 72 %
 série B : R = 35 %

voir I.E.A. 650

La notion de moyenne n'est pas perçue comme difficile par les élèves, comme semble le montrer le faible taux de non-réponses à la question CA03-04, mais elle leur pose cependant des problèmes, ainsi 57% des élèves font $(12,30+14,80)/2$ sans tenir compte de la composition des classes, montant ainsi une conception erronée de la moyenne. La très grande familiarité de cette notion devient peut-être un obstacle à la compréhension. Le taux de non-réponses est élevé dans la question CA05 au sujet de l'écart-type. Cette notion de Seconde n'est pas a priori revue en Première et de plus la formulation est sans doute peu habituelle.

La question CB09-12 et la question CA03-04 qui proviennent d'EVAPM3/92 permettent de croiser plusieurs domaines essentiels aussi bien pour la formation de base du citoyen que pour toute initiation scientifique : la géométrie dans l'espace et la proportionnalité.

Les réussites sont très contrastées (la réussite conjointe passe de 12% à 54% et les élèves de G ont un taux fort de non-réponses - dont nous avons déjà parlé).

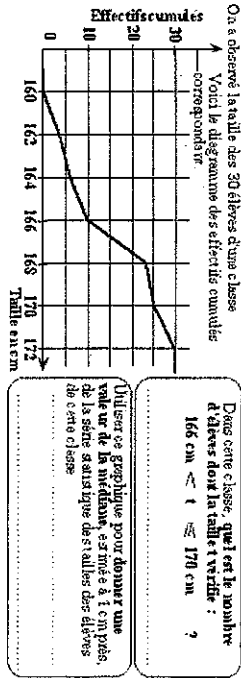
Les résultats ont finalement peu évolué depuis la classe de Troisième, sauf en ce qui concerne la perception de l'angle.

Une très grande partie des élèves semble percevoir correctement l'invariance de l'angle, ce qui montre une meilleure appréhension de l'espace qu'en Troisième. Par contre l'effet de l'agrandissement sur les arcs et les volumes pose toujours des difficultés (sauf en E et S peut-être grâce au travail sur les homothéties), montrant ainsi que les concepts d'aire et de volume ne sont pas encore bien acquis en fin de Première.

À ce sujet, nous renvoyons le lecteur à l'article "Longueur - Aire - Volume" du bulletin n° 402 de l'APMP (Février 1996), article qui s'appuie largement sur les études EVAPM de la Sixième à la Seconde.

Cl6697

Origine : APMEP
Sens : Communication
Type d'activité : Classificatoire
Forme : Feuille
Durée estimée : De 2 à 5 min
Facilité : 40
Complexité : A3
Mots-clés : Graphique; Médiane; Statistique;
Capacités : S008;
Âges : 15; 16;

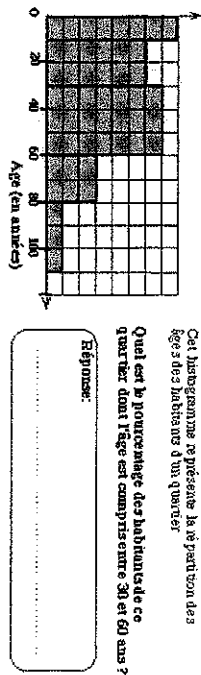


EVAPM291 - F39 :
nombre R.E. : 15 ; R = 45 % N.R. = 26 %
médiane R.E. : toute valeur appartenant à [166;168] : R = 20 % N.R. = 72 %
EVAPMLP95 - L12 :
nombre R.E. : 15 ; R = 30 % N.R. = 14 %
médiane R.E. : toute valeur appartenant à [166;168] : R = 33 % N.R. = 14 %

Le premier item de la question F39-40 montre que les élèves ont des difficultés à lire un polygone de fréquences cumulées. Ce type de diagramme est pourtant d'utilisation courante dans le monde social. Parmi les erreurs, on trouve souvent 25 : confusion avec le nombre d'élèves dont la taille est inférieure à 170cm. On trouve aussi 48 : somme des "images" de 167 et de 168 lues sur le graphique.
Très peu d'élèves répondent à la question concernant la médiane (28%) ; par contre la plupart de ceux qui répondent le font correctement. Rappelons que la notion de médiane figure au programme de Troisième, quelle peut faire l'objet d'activités en Seconde, mais quelle ne constitue une connaissance exigible ni dans une classe, ni dans l'autre.

Cl6713

Origine : APMEP
Sens : Communication
Type d'activité : Traductif
Forme : Semi-ouverte
Durée estimée : De 5 à 10 min
Facilité : 30
Complexité : B3
Mots-clés : Histogramme; Pourcentage; Statistique;
Capacités : S008;
Âges : 15; 16;



EVAPM291 - E22 :
R.E. 42 % : R = 31 % N.R. = 27 %
EVAPMLP95 - D11 :
R.E. 42 % : R = 23 % N.R. = 14 %

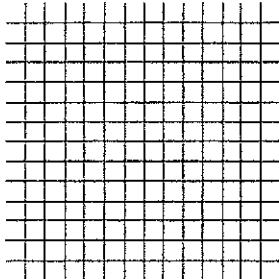
Non exigible en seconde (1991)
vérifier et compléter les codes APM et EVAPMIB
Les élèves ont tendance à utiliser des histogrammes ou des diagrammes de type histogramme lorsque cela n'est pas nécessaire, mais ils éprouvent de la difficulté à lire un histogramme simple. D'après le programme, on ne peut pas considérer la lecture d'un tel diagramme comme étant exigible en Seconde. On peut cependant penser qu'il n'y a la aucune difficulté particulière et que la compétence correspondante présente une certaine utilité sociale.
Parmi les 31% de bonnes réponses, on observe essentiellement deux démarches :
- passage par le pourcentage correspondant à un carreau,
- rapport entre le nombre de carreaux du rectangle correspondant aux âges considérés et le nombre total de carreaux.
Parmi les démarches erronées, citons :
- prise en compte de la seule hauteur de la classe (7 carreaux, donc 70%),
- prise en compte de la seule largeur de la classe (3 carreaux, donc 30%),
- comparaison de la hauteur de la classe avec la plus grande hauteur (7/8 * 100),
- comparaison de la largeur de la classe avec la plus grande largeur (3/11 * 100),
- calcul effectué comme si toutes les classes avaient la même amplitude.

Origine : APMEP
 Sens : Communication
 Type d'activité : Traductif
 Forme : Fermée
 Durée estimée : De 2 à 5 min
 Facilité : 70
 Complexité : A4
 Mots-clés : Diagramme, Statistique;
 Capacités : S006;
 Ages : 13;

Voici la répartition des élèves de sixièmes d'un collège, se lon la langue vivante étudiée.

Langue vivante	Anglais	Allemand	Russe
Nombre d'élèves	50	45	15

TRACER le diagramme en bâtons
 représenter ces données.
 au-dessous du tableau pour 10 lignes.



EVAPM4/91 - D01 :
 (on acceptera des bâtons dessinés en rectangles)
 L'axe des effectifs est gradué correctement
 et exactement 2 bâtons corrects : R = 44 %
 Les trois bâtons sont corrects : R = 74 % N.R = 05 %

Origine : APMEP
 Sens : Communication
 Type d'activité : Traductif
 Forme : Fermée
 Durée estimée : De 5 à 10 min

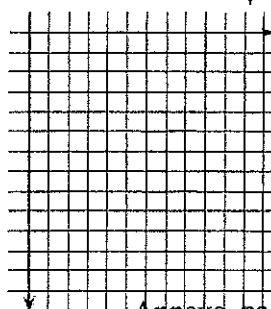
Facilité : 40
 Complexité : B3
 Mots-clés : Echelle;
 Statistique;
 Capacités : S007; S011;
 S014;
 Ages : 15;
 Voisines : 758

La tableau ci-dessous donne les effectifs, par âge, d'un club de bridge.

Âges des élèves	14 ans	15 ans	16 ans	17 ans
Nombre d'élèves par âge	1	3	15	5

COPIER le tableau de façon à obtenir le tableau des effectifs cumulés croissants.

TRACER le polygone des effectifs cumulés croissants ou une autre représentation de ces effectifs cumulés croissants.



CALCULER l'écart type de la série statistique donnée
 Résultat arrondi :

EVAPM2/91 - D01 :
 tableau complété correctement [14;19;24
 même si la case de gauche n'est pas ou mal remplie : R = 40 % N.R. = 43 %
 démarrage : une échelle apparaît sur les axes, ainsi qu'un diagramme ou un polygone manifestant que l'élève a bien compris la possibilité de traduire par un graphique le tableau : R = 32 % N.R. = 51 %
 polygone complet et correct : R = 16 % N.R. = 51 %
 autre représentation correcte (en escalier ou autre) : R = 13 % N.R. = 51 %
 écart-type R.E. 0,7 an (accepter tout arrondi ainsi que l'absence d'unité) : R = 05 % N.R. = 74 %

voir aussi question 758
 Les effectifs cumulés croissants de la question D1-5 ont visiblement fait fuir les élèves. Cette question reprise d'EVAPM3/90, pour la première partie, était mieux réussie en fin de Troisième (35% de réussite pour l'ensemble de la question, au lieu de 29% cette année).
 En ce qui concerne les effectifs cumulés, une erreur fréquente consiste à donner les compléments à l'effectif total (24) de l'effectif correspondant à chaque âge. On obtient alors la réponse suivante :
 tableau page 97
 Contrairement à la notion de médiane, celle d'écart type figure parmi les capacités exigibles de Seconde. Il est certain que cette notion avait été très rarement vue au moment de l'évaluation et il est peu probable qu'elle ait pu être vue de façon efficace ensuite.
 Concernant l'écart type, nous avons accumulé ici plusieurs difficultés. Nous aurions donc dû, à côté de cette question, poser une question plus simple présentant une seule série, non pondérée. Toutefois, l'examen des réponses proposées par les élèves montre qu'en général les erreurs ne viennent pas de là, mais bien d'une méconnaissance de la notion elle-même.

EVAPM3/90 - P10 : page 79

Le calcul d'effectifs cumulés, demandés dans la question P10, ne fait pas partie des capacités exigibles officielles. Au moins doit-il y avoir, en troisième, reprise de ces notions déjà vues en quatrième à propos d'activités. Or il y a ici 29 % de non-réponses, et beaucoup d'erreurs parmi les réponses.

- Ces erreurs sont essentiellement de trois types :
 - confusion avec les "fréquences",
 - confusion avec le rangement des effectifs par ordre croissant,
 - cumul non pas des effectifs, mais des âges : 11 ; 23 ; 36 ; 50.
- Cela ne relevait-il pas d'une "bonne" mécanique née de tableaux où la ligne qui précède les effectifs cumulés est celle des effectifs ?
- A noter cependant des progrès par rapport à la quatrième : il n'y avait alors que 17 % de hommes réponses. L'analyse relative au graphique sera faite plus loin.

page 83

A propos de la question P12-13. Notons qu'en quatrième, les bonnes réponses pour la seule compréhension de la démarche étaient de 11 % (contre 39 % en 3ème) et il n'y avait que 7 % de réponses correctes pour le graphique (contre 28 %).

EVAPM4/89 - Q15 : page 80-81 voir P21

La question Q15-16-17 n'a pas eu beaucoup de succès. Pourtant, la question est simple dès lors que l'on sait de quoi il s'agit. Or il y a 60 % de non-réponses et la moitié des réponses données sont d'interprétation difficile. Le taux de non-réponses à cette question est particulièrement important (60 %). C'est le plus important enregistré pour l'évaluation EVAPM4, (avec de façon très voisine le taux de non-réponses de Q8...11. Or la question est plutôt placée en début de questionnaire, et les questions suivantes sont largement traitées. On ne peut que voir dans ce taux le signe d'un rejet massif de la question. Pourquoi ce rejet ? Serait-ce que les collègues auraient fait l'impasse sur les statistiques, surtout pour la partie non expressément citée en "capacités exigibles", ou qu'ils les réservent pour les derniers jours de juin ?

Pour revenir à la question Q15...17, signalons que le codage était on ne peut plus large: on acceptait que la case de gauche soit mal remplie et on exigeait seulement la présence de la suite: 3; 11; 23; 25.

Parmi les réponses curieuses, citons un "cumuli" par colonne : 3 + 11 ; 8 + 12 ; 12 + 13 ; 2 + 14. L'élève avait retenu "qu'on additionnait" ... mais quoi ? et avec quelle signification ?

La case de gauche est parfois remplie par "effectifs cumulés croissants" ce qui n'est évidemment pas faux mais qui n'explique rien, si bien que seul un réflexe visuel peu jouer pour fournir des résultats corrects alors même qu'on ne saurait pas ce qu'ils signifient (on a vu l'impact de ce réflexe, à tort et à travers, pour l'item A 21). Les autres copies jonglent avec "au moins" et "au plus" d'une façon qui frise l'aléatoire... Il y a donc des gros progrès à faire pour faire comprendre "en français" le sens de "effectifs cumulés croissants" (pour les "décroissants" aussi sans doute !)? Sinon, que signifient-ils ?

Si l'on devait retenir uniquement les réussites comportant aussi une case de gauche explicitement et correctement remplie il ne resterait que bien peu d'élèves...

Il était quasiment impossible de répondre à la dixième partie de Q15...17 sans avoir bien répondu à la première. Seuls 1/3 environ des élèves qui ont réussi la première partie réussissent aussi la seconde, les deux autres tiers (non-réponse ou erreur) sont sans doute désarçonnés par l'expression "polygone des effectifs..." employée ici.

Un élève a, par exemple, tracé un quadrilatère aux sommets duquel il a marqué les 4 effectifs cumulés croissants.

En acceptant une bonne mise en place du polygone, sans que celui-ci soit pour autant complet et correct les taux de réussite ne sont guère plus importants que pour un polygone complet et correct : 11 % au lieu de 7 %).

Clé760

Origine : APMEP

Sens : Communication

Type d'activité : Traductif

Forme : Semi-ouverte

Durée estimée : De 5 à 10 min

Facilité : 20

Complexité : B3

Mots-clés : Pourcentages, Statistique,

Capacités : S005;

Âges : 13;

On rejette les jets de 400 tirés à tableaux.

Parmi eux : 30 % contiennent entre 500 F et 2000 F (catégorie I)

50 % contiennent entre 2000 F et 3500 F (catégorie II)

Les 2 autres contiennent entre 3500 F et 5000 F (catégorie III)

1) Présente ces résultats en un tableau des fréquences par catégories.

2) Présente ces résultats en un tableau des effectifs par catégories.

EVAPM4/89 - Q08 :

exactement une erreur dans premier le tableau : R = 04 % NR = 54 %

premier tableau correct : R = 16 % NR = 54 %

exactement une erreur dans le deuxième tableau : R = 01 % NR = 68 %

deuxième tableau correct : R = 13 % NR = 68 %

EVAPM4/89 - Q08 : page 79 voir A21

La question Q8...11 oblige à manipuler simultanément des fréquences et des pourcentages... et des effectifs. Il est clair que cette juxtaposition de difficultés a fait fuir les élèves, et bien que cette question soit placée en début de questionnaire, plus de la moitié d'entre eux ne l'aborderont pas. Dans le 1^{er}, même si l'on accepte une erreur, le taux de réussite n'est que de 20%. Ce taux est de 14% pour la deuxième partie.

La difficulté relative étant reconnue, les résultats restent inquiétants et nous questionnent gravement : En sixième, les mêmes questions, rédigées autrement (1^{er}) : indique le pourcentage pour la catégorie III ; 2^o : combien y a-t-il de réfringents dans chaque catégorie ? auraient probablement été perçus comme faciles et généralement réussies.

Ici interviennent un vocabulaire spécifique (fréquences, effectifs) pas courant du tout... pour des objets ! et les résultats sont catastrophiques.

Force est donc de constater que ce vocabulaire "fréquence", "effectifs", dès lors qu'il est ignoré ou mal assimilé, perturbe totalement la compréhension des faits et détériore massivement celle des pourcentages.

Non seulement il n'apporte rien, mais il détruit !

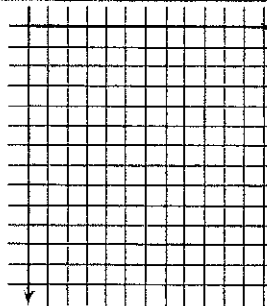
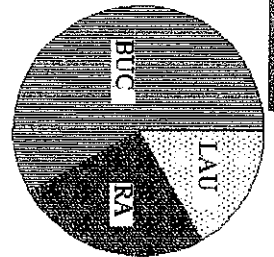
Les réponses déficientes témoignent d'ailleurs d'une perte totale du sens : l'élève ne sait plus ce qu'il fait. Ne pouvons-nous donner à cette remarque une portée générale sur les excès ou les parachutages de vocabulaire mathématique ?

Origine : APMEP
 Sens : Communication
 Type d'activité : Technique
 Forme : Femme
 Durée estimée : De 5 à 10 min
 Facilité : 20
 Complexité : B1
 Mots-clés : Diagramme;
 Graphique; Pourcentage;
 Proportionnalité; Statistique;
 Capacités : S004;
 Ages : 13;

Le professeur doit donner à l'élève LAU, RA et BUC le matériel pour tracer ces trois angles, conformément à son degré de compétence et contre

REPRÉSENTE, sur la partie quadrillée, cette répartition par un diagramme et bâtons portant sur les pourcentages.

Explique ce que tu fais pour passer d'un pourcentage à un diagramme



EVAPM4/89 - P21 :
 explication correcte pour le calcul d'au moins un pourcentage même si le résultat est faux : R = 23 % NR = 47 %
 calcul correct des trois pourcentages : R = 21 % NR = 47 %
 mise en place d'échelles et tracé d'exactement 1 ou 2 bâtons corrects : R = 08 % NR = 47 %
 correct des 3 bâtons : R = 18 % NR = 47 %

EVAPM4/89 - P21 : page 80 voir Q15

Questions mettant en jeu les deux compétences S321 et 322.

Dans la question P21..24, aux 21% des élèves qui calculent correctement les trois pourcentages et qui manipulent ainsi une certaine maîtrise, il faut opposer les 2% des élèves qui expliquent correctement au moins un des calculs de ces pourcentages, et qui ne parviennent pas à calculer les trois pourcentages demandés. Ce taux de 2% montre assez que les difficultés ne sont pas d'ordre calculatoire, mais bien d'ordre conceptuel. Un élève qui sait expliquer le calcul d'un pourcentage obtient généralement très bien les trois.

Mais cet item n'est pas du tout réussi (près de la moitié de non-réponses, et moins d'un quart des élèves pour savoir de quoi il retourne). Une fois de plus, les élèves qui se trompent confondent pourcentages du total et mesures en degrés des angles au centre. De là, des pourcentages voisins de 34%, 90 %, 216%. Certes la proportionnalité est respectée mais le concept de pourcentage s'effondre!

Comparons avec l'item A22 (voir plus haut), qui propose, à partir du même dessin, la recherche des pourcentages ; les résultats sont incompréhensiblement meilleurs pour le A22.

Cela ne peut qu'être attribué à deux facteurs :

- le caractère plus fin, plus indirect, de l'énoncé des P21-22, par une liste de réponses possibles (méthode "Q.C.M.");
 - le caractère plus flou, plus indirect, de l'énoncé des P21-22, le second facteur paraît décisif pour l'augmentation des mauvaises réponses, du A22 au P21-22 (de 29% à 60% des réponses).

Si le premier facteur aggrave la proportion des non-réponses au P21-22, le second facteur paraît décisif pour l'augmentation des mauvaises réponses, du A22 au P21-22 (de 29% à 60% des réponses).

En ce qui concerne la construction du diagramme en bâtons, l'énoncé induisait un passage par un calcul de pourcentages, donc seuls les élèves ayant réalisé, correctement ou non, ces calculs se sont lancés dans le dessin des bâtons.

La référence faite, pour le codage, à une graduation correcte, avec de bon pourcentages, de l'axe des ordonnées, a fait compter faux les dessins établis à partir d'une mauvaise réponse au 1°. De là, la faiblesse du score. (les pourcentages de réussite étant portés à 26% si l'on acceptait des erreurs sur 1 ou 2 bâtons, mais avec toujours de bonnes échelles).

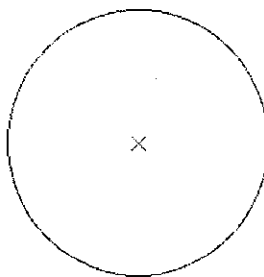
Remarquons que si l'on n'exigeait pas une échelle indiquant des pourcentages, il serait facile de fournir des bâtons proportionnels à des pourcentages non calculés, en donnant aux bâtons des longueurs proportionnelles aux angles au centre des secteurs LAU, RA et BUC.

Origine : APMEP
 Sens : Communication
 Type d'activité : Calculatoire
 Forme : Femme
 Durée estimée : De 2 à 5 min
 Facilité : 30
 Complexité : A3
 Mots-clés : Diagramme; Proportionnalité; Statistique;
 Capacités : S006;
 Ages : 13;

Le professeur doit donner à l'élève LAU, RA et BUC le matériel pour tracer ces trois angles, conformément à son degré de compétence et contre

note	0	10	20
fréquence	0,1	0,6	0,3

Explique cela par un diagramme circulaire.



EVAPM4/89 - M11 :
 conforme au cahier de tolérances : R = 30 % NR = 36 %
 EVAPM4/91 - M16 :
 conforme au cahier de tolérances : R = 30 % NR = 42 %

Il s'agit ici de tracer un diagramme circulaire à partir de données statistiques exprimées sous forme de fréquences.

Cet item est réussi par seulement 30% des élèves de quatrième (EVAPM 4/89). Le pourcentage de réussite passe à 39% en ce qui concerne le tracé correct d'un seul secteur (aux tolérances près).

Plusieurs éléments peuvent expliquer cette difficulté :
 - Construire une représentation graphique est une tâche plus difficile que de lire et interpréter un graphique déjà présent.

- Certains élèves ignorent ce qu'est une fréquence et, soit ne dessinent rien, soit les reportent sans prendre en compte la relation partie-tout. Cette procédure leur permet de réussir les représentations de type graphique bâton où il suffit de reporter directement les fréquences. Ici par exemple faute de pouvoir tracer des angles de 0,1 ou 1 degré, ces élèves ont tracé des angles de 10, 60 et 30 degrés qui respectent la proportionnalité, mais qui ne sont plus des diagrammes circulaires.

- Le diagramme de type représentation circulaire peut amener certaines confusions par exemple : un transit par des pourcentages qui augmentent les risques d'erreurs voire un mélange degré pourcentage, une utilisation inadaptée du rapporteur (dépassement de la tolérance de un degré, lecture de 120° au lieu de 60°) etc.

EVAPM4/89 - M11 : page 79 voir C15

Pour la question M11-12, en acceptant jusqu'à un seul secteur correct (aux tolérances près), le pourcentage de réussite passe à 38%. C'est vraiment peu.

Ici aussi, on pouvait partager le disque sans connaître le sens du mot "fréquence" et cela a souvent dû se produire, à en juger par les analyses précédentes.

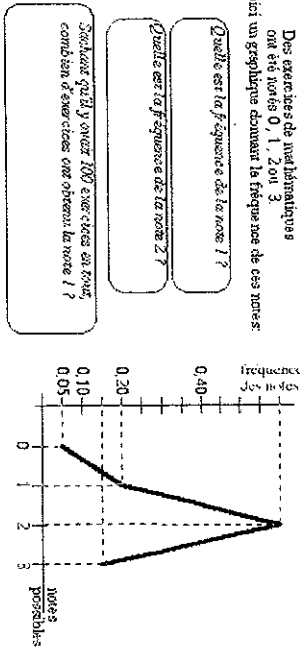
Mais ce partage n'est parfois pas un : fautive, sans doute, de pouvoir tracer des angles de 0,1° ou de 1°, des élèves ont tracé des angles au centre de 10°, 60° et 30°. La proportionnalité est alors respectée, ce qui est bien, mais il ne s'agit plus d'un "diagramme circulaire". (Le diagramme de ces élèves n'en serait pas moins acceptable comme représentation correcte de la situation, et on utilise effectivement de tels diagrammes, en séparant alors souvent les trois secteurs circulaires).

Il reste les 36% de non-réponses, dues sans doute, une fois de plus (cf. analyses précédentes) à l'inhibition provoquée par le mot "fréquences" ignoré ou mal assimilé.

Clé795

Origine : APMEP
 Sens : Communication
 Type d'activité : Calculatoire
 Forme : Fermée
 Durée estimée : De 2 à 5 min
 Facilité : 40
 Complexité : A4
 Mots-clés : Statistique;
 Capacités : S004;
 Ages : 13;

Des exercices de mathématiques ont été notés 0, 1, 2 ou 3. Voici un graphique donnant la fréquence de ces notes.



EVAPM4/89 - D17 :
 R.E. 0,20 ou 20% : R = 40 % NR = 07 %

EVAPM4/89 - D17 : page 78 voir B36
 Le troisième item de D19 demande en fait de traduire une fréquence en effectif. C'est plus facile qu'à l'item B37 analysé plus haut puisqu'il l'effectif total est donné et avec un nombre plus qu'accommodant ! Il y a donc des résultats quelque peu meilleurs, mais ils sont encore très faibles. Cela témoigne d'une profonde méconnaissance de ce qu'est une fréquence ou d'une ignorance de ce mot. Sinon la réponse est immédiate. L'erreur la plus fréquente assimile effectifs et fréquences et fait curieusement répondre "0,20", mais un tiers des élèves n'a pas répondu...
 Du premier item au troisième, le nombre de non réponses passe de 10% à 34%. Il semble donc bien qu'environ 1/4 des élèves ont répondu à la question 1 sans savoir ce qu'est une fréquence et uniquement parce qu'ils maîtrisent ce type de graphique.
 La capacité testée sous le code S 321 se retrouve ensuite dans un exercice plus complexe, proposé dans le questionnaire (complémentaire) P, qui sera analysé plus loin.

Clé813

Origine : APMEP
 Sens : Communication
 Type d'activité : Classificatoire
 Forme : Q.C.M.
 Durée estimée : Moins de 2 min
 Facilité : 50
 Complexité : B3
 Mots-clés : Statistique;
 Capacités : S005;
 Ages : 13;

Un professeur a corrigé 10 devoirs.

Il a mis les notes suivantes :
 5 ; 8 ; 8 ; 14 ; 5 ; 17 ; 8 ; 14 ; 14 ; 14 ;
 Les tableaux ci-dessous présentent les notes et les effectifs correspondants :

Note	5	8	14	5	17	8	14	14	14	14
Effectif	1	2	1	1	1	1	1	1	1	1

Note	5	8	14	5	17	8	14	14	14
Effectif	1	2	1	1	1	1	1	1	1

Note	5	8	14	5	17	8	14	14	14
Effectif	1	2	1	1	1	1	1	1	1

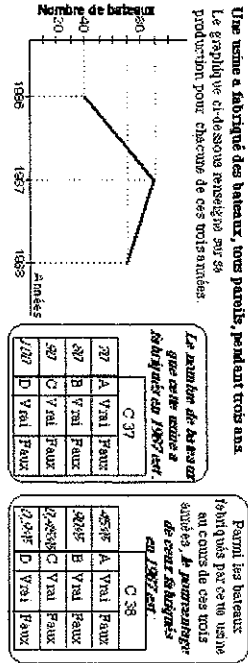
Note	5	8	14	5	17	8	14	14	14
Effectif	1	2	1	1	1	1	1	1	1

	A	Vrai	Faux
Tableau 1	A	Vrai	Faux
Tableau 2	B	Vrai	Faux
Tableau 3	C	Vrai	Faux
Tableau 4	D	Vrai	Faux

EVAPM4/91 - C39 :
 B et D : R = 47 % NR = 12 %

C16814

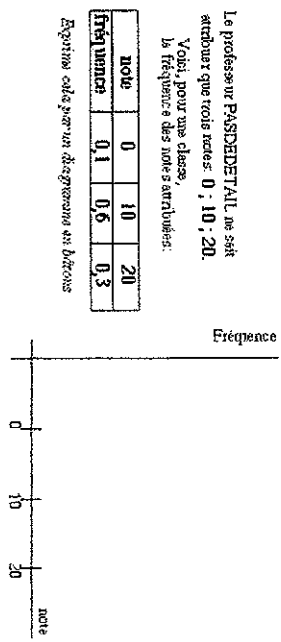
Origine : APMEP
Sens : Communication
Type d'activité : Traductif
Forme : Q.C.M.
Durée estimée : De 2 à 5 min
Facilité : 60
Complexité : A3
Mots-clés : Graphique; Pourcentage;
Capacités : S004;
Âges :13;
Voisines : 10



EVAPM4/91 - C37 :
nombre de bateaux C : R = 89 % NR = 03 %
pourcentage de bateaux A : R = 42 % NR = 11 %
voir jumelle non Q.C.M. 10

C16833

Origine : APMEP
Sens : Objet
Type d'activité : Technique
Forme : Semi-ouverte
Durée estimée : De 2 à 5 min
Facilité : 50
Complexité : B3
Mots-clés : Graphique; Statistique;
Capacités : S006;
Âges :13;



EVAPM4/89 - C15 :
une échelle figure sur l'axe des fréquences et exactement deux bâtons sont corrects : R = 07 % NR = 16 %
une échelle figure sur l'axe des fréquences et les trois bâtons sont corrects : R = 48 % NR = 16 %

EVAPM4/89 - C15 : page 79 voir M11
A partir de données statistiques... TRACER les diagrammes correspondant (S323).
Pour la question C15-16, si l'on accepte une erreur (pour un bâton) le pourcentage de réussite passent à 55%.
Cela reste faible, mais corrobore les analyses précédentes :
- Certains élèves ignorent sans doute ce qu'est une fréquence mais maîtrisent les graphiques bâtons et s'en tirent, puisqu'il suffit de reporter les fréquences sans avoir besoin de savoir ce que c'est.
- D'autres répondent en sachant ce qu'est une fréquence (d'après les analyses précédentes, les deux groupes d'élèves auraient des effectifs sensiblement voisins).
- Quelques élèves sont inhibés par le mot fréquence et ne dessinent rien, ou n'importe quoi (plaçant par exemple, sur l'axe des fréquences, 0,1 x 0 ; 0,16 x 10 ; 0,3 x 20 c'est-à-dire 0 ; 6 ; 6).
- Enfin, quelques autres ne maîtrisent pas du tout les graphiques bâtons, ne savent pas décider d'une échelle et dessinent apparemment n'importe quoi (certains élèves ont placé 0,1 et 0,6 et 0,3 au même endroit sur l'axe des fréquences).

Clé904

Origine : APMEP
Sens : Objet
Type d'activité : Calculatoire
Forme : Fermée
Durée estimée : De 2 à 5 min
Facilité : 70
Complexité : A4
Mots-clés : Moyenne;
Capacités : S020;
Âges : 15; 16;

Dans une classe de 40 élèves la moyenne d'un devoir était 10,6.
En fait le professeur avait noté 1 point à 2 élèves.

Quelle est, en réalité, la moyenne de la classe? (2,2 de manière de valeur exacte.)

Calculs :

Moyenne de la classe :

EV/PM2/91 - B18 :
calcul de la moyenne en passant par $40 \times 10,6 + 2$: R = 58 % NR = 25 %
ajout direct de 2/40 de point : R = 23 % NR = 25 %
R.E. 10,65 quelle que soit la démarche : R = 66 % NR = 25 %
EV/PM1/95 - A03 :
calcul de la moyenne en passant par $40 \times 10,6 + 2$: R = 60 % NR = 15 %
ajout direct de 2/40 de point : R = 33 % NR = 15 %
R.F. 10,65 quelle que soit la démarche : R = 69 % NR = 15 %

Lorsque nous avons posé la question B18-20, elle nous semblait porter sur la signification même de la moyenne. En effet, l'application sans plus de réflexion de la définition de la moyenne conduisant à retrouver la somme totale des points, à ajouter les deux points oubliés et à diviser par le nombre d'élèves. Une réflexion élémentaire sur la notion de moyenne devait conduire à faire effectuer à cette moyenne, directement, une translation de 2/40 de points.
En fait, nous nous sommes aperçus que même des enseignants n'envisageaient que la première méthode, non pas qu'ils ne puissent avoir accès à la seconde, mais parce que, comme pour les élèves, le principe d'économie qui guide nos démarches nous empêche de nous éloigner d'une méthode standard, peu élégante mais dont on est sûr qu'elle marche, si nous n'y sommes pas expressément invités. Finalement, il est tout à fait compréhensible que, dans le cas de la question B18-20, la plupart des élèves reviennent à la définition de la moyenne.
Le codage prévoyait de coder séparément les deux démarches, mais, non préparés à faire la distinction, 30% des enseignants ont implicitement considéré que la seconde impliquait la première. Correction faite de ces erreurs de codage, il ne reste que 07% des élèves qui aient trouvé un résultat exact par ajout direct de 2/40 points.
Le taux important de non-réponses (25%) montre bien le peu d'intérêt des élèves pour ce genre de question. La difficulté n'est pas en cause car 90% des élèves qui répondent le font correctement.
Une autre question concernant la notion de moyenne et sa signification a été posée dans le cadre de l'épreuve recherche de problèmes (question 1/2). On en trouvera une analyse dans le chapitre correspondant.

Clé922

Origine : APMEP
Sens : Objet
Type d'activité : Calculatoire
Forme : Semi-ouverte
Durée estimée : De 2 à 5 min
Facilité : 70
Complexité : B3
Mots-clés : Statistique;
Capacités : S007; S011; S014;
Âges : 15;

Le tableau ci-contre donne la répartition des 2 500 élèves d'un lycée selon la direction de leur famille (nombre de personnes vivra sous le même toit)

Direction	Nombre de personnes par famille
Est	2
Sud	3
Ouest	4
Nord	5
Autre	6
Indéterminé	7
Effectifs	10% 23% 42% 15% 5% 2%

a) Compléter la dernière ligne du tableau

b) Représenter les effectifs par un diagramme en bâtons

Indiquer le quadrillage donné

EV/PM2/91 - A04 :
au plus une erreur dans le tableau : R = 65 % NR = 21 %
aucune erreur (250,575,1125,375,125,50) R = 63 % NR = 21 %
diagramme ébauchant correctement le tableau obtenu, avec éventuellement oublié des indications concernant les axes et/ou une erreur de longueur sur un bâton R = 36 % NR = 32 %
diagramme correct, y compris en ce qui concerne les indications nécessaires à sa lecture R = 22 % NR = 32 %

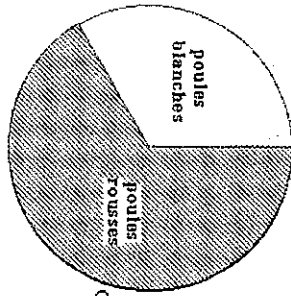
Dans la question A4-7, on obtient 63% de réussite pour le passage de fréquences, exprimées en pourcentages, à des effectifs. Les élèves ayant la possibilité d'utiliser leur calculatrice, le plus surprenant est le taux de non-réponses (21%). Voyant arriver des pourcentages et des traces de statistiques, un élève sur cinq préfère passer directement à la question suivante.
Pour les réponses fausses, on trouve des effectifs qui ne respectent pas les ordres de grandeurs relatifs des fréquences données, et dont la somme est différente de l'effectif total de la population. Pour quelques élèves, il y a encore confusion entre fréquences et effectifs : une fréquence de 10% se traduit alors simplement par un effectif de 10.

Pour les diagrammes, les erreurs proviennent essentiellement d'une utilisation abusive des histogrammes. Les histogrammes, on le sait, ne sont justifiés que dans le cas des variables numériques continues, ou considérées comme telles. Certes des représentations du type histogrammes (bâtons larges et accolés) peuvent être utilisées dans les autres cas, mais cela risque d'amener les élèves au type d'erreur présentée ci-dessous.
On observera que le diagramme présenté au milieu de la figure ci-dessus peut difficilement être banni compte tenu des habitudes culturelles (presse, publicité...) et scolaires (Sciences Humaines, Sciences Economiques et Sociales...), et que l'essentiel réside dans la lecture que les élèves font de telles représentations.

Clé948

Origine : APMEP
Sens : Communication
Type d'activité : Technique
Forme : Fermée
Durée estimée : De 2 à 5 min
Facilité : 30
Complexité : A4
Mots-clés : Diagramme; Graphique; Proportionnalité; Statistique;
Capacités : S002; S003;
Âges : 12;

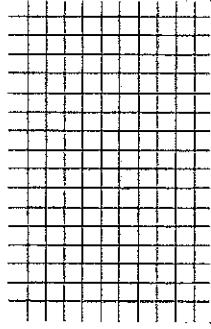
Le diagramme ci-dessous représente les résultats d'une enquête sur les 240 poules d'un élevage.
Le cercle de droite est le cobain.



A quelle fraction du disque correspond le secteur hachuré ?
Écrivez votre réponse.

Quel est le nombre de poules noires ?

Utilisez le quadrillage de droite pour construire un diagramme en bâtons.
Écrivez votre réponse.
L'axe des ordonnées est en mètres.



EVAPM3/90 - M16 :
fraction R.E. 2/3 ou fraction voisine
(erreur de 2° tolérée) : R = 33 % N.R. = 12 %
R.E. 1/6 ou nombre de [158;162] : R = 33 % N.R. = 12 %
compatible avec le nombre trouvé avant : R = 33 % N.R. = 12 %
(accepter deux bâtons larges si la légende permet la lecture)

page 90

La question M16-17 demande que l'élève organise lui-même la tâche. Les copies que nous avons consultées montrent que les élèves ont compris la question. La première source d'erreur provient de leur difficulté à associer une fraction à un secteur de disque. À ce sujet, le lecteur pourra se reporter à notre analyse du thème fractions dans le chapitre Domaine Numérique. Ici aussi, la plupart des représentations produites sont du type histogramme. Confronté aux consignes de codage, ces graphiques sont considérés comme exactes dans la mesure où la légende est suffisante pour une bonne interprétation.

Clé997

Origine : APMEP
Sens : Objet
Type d'activité : Calculatoire
Forme : Fermée
Durée estimée : De 2 à 5 min
Facilité : 60
Complexité : A1
Mots-clés : Médiane; Moyenne; Statistique;
Capacités : S010;
Âges : 14;
Voisines : 1046

Voici les masses respectives de 11 poules d'un lot.

1,2 kg; 1,4 kg; 1,6 kg; 1,8 kg; 3,5 kg; 3,2 kg; 1,4 kg; 1,5 kg; 3,2 kg; 3,5 kg; 3 kg.

Quelle est la moyenne de ces masses ?

La médiane de cette liste est 1,6 kg.
Pourquoi ?

EVAPM3/90 - M12 :
moyenne R.E. tout nombre compris entre 2,3 et 2,4 : R = 84 % N.R. = 07 %
médiane explication du type 5 valeurs en dessous,
5 valeurs au dessus : R = 38 % N.R. = 42 %
EVAPM3/92 - R27 :
moyenne R.E. tout nombre compris entre 2,3 et 2,4 : R = 84 % N.R. = 06 %
médiane explication du type 5 valeurs en dessous,
5 valeurs au dessus : R = 26 % N.R. = 54 %

C161028

Origine : APMEP
Sens : Objet
Type d'activité : Traductif
Forme : Fermée
Durée estimée : De 5 à 10 min
Facilité : 50
Complexité : A1
Mots-clés : Diagramme; Fréquence; Graphique; Statistique;
Capacités : S011; S012; S014;
Âges : 14;

La correction des 28 copies d'un devoir surveillé en composition française a fourni la liste des notes suivantes :

12 : 9 ; 11 ; 12 ; 10 ; 8 ; 11 ; 10 ; 9 ; 11 ; 10 ; 9 ; 10 ; 10 ;
11 ; 11 ; 6 ; 10 ; 9 ; 12 ; 10 ; 11 ; 10 ; 10 ; 10 ; 11 ; 10 ; 9 .

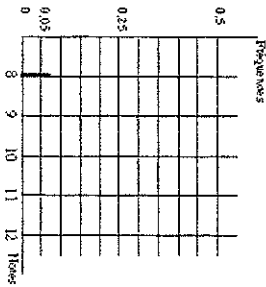
a) Complétez le tableau des effectifs par note :

Notes	8	9	10	11	12
Nombre de copies	2				

b) Complétez le tableau des fréquences des notes :

Notes	8	9	10	11	12
Fréquences	0,07				

c) Représentez ces fréquences par un diagramme en bâtons.



- EVAPM3/92 - 134 :
- a) R.E. : 2 - 5 - 11 - 7 - 3 ; R = 87 % NR = 08 %
 b) R.E. : 0,07 - 0,18 - 0,39 - 0,25 - 0,11
 ou autres valeurs approchées correctes,
 ou réponses en cohérence avec le tableau a) : R = 42 % NR = 32 %
 c) diagramme en cohérence avec le tableau a) : R = 51 % NR = 36 %

C161029

Origine : APMEP
Sens : Objet
Type d'activité : Calculatoire
Forme : Fermée
Durée estimée : De 5 à 10 min
Facilité : 50
Complexité : A4
Mots-clés : Diagramme; Fréquence; Graphique; Statistique;
Capacités : S011; S012; S014;
Âges : 14 ; 16;

Les 20 élèves d'une club ont donné leur âge en années. Leur liste, par ordre alphabétique, donne le relevé des âges suivants :

13 ; 15 ; 15 ; 14 ; 16 ; 13 ; 15 ; 14 ; 16 ; 16 ;
15 ; 14 ; 15 ; 13 ; 15 ; 14 ; 14 ; 15 ; 14 ; 16

a) Faites un tableau des effectifs par âges

(complétez le tableau ci-dessous.)

Âges en années	13 ans	14 ans	15 ans	16 ans
Nombre d'élèves

b) Calculez les fréquences des quatre âges

(complétez le tableau ci-dessous.)

Âges en années	13 ans	14 ans	15 ans	16 ans
Fréquences

c) Représentez ces fréquences par un diagramme en bâtons.

(Choisissez l'échelle de vos bâtons.)

- EVAPM3/90 - D28 :
- tableau a) R.E. 3 ; 6 ; 7 ; 4 ; R = 89 % NR = 07 %
 R = 85 % (orientés en LP)
 tableau b) R.E. 0,15 ; 0,30 ; 0,35 ; 0,20 ; R = 31 % NR = 33 %
 R = 13 % (orientés en LP)
 bâtons de longueurs proportionnelles aux effectifs
 avec indication des fréquences : R = 24 % NR = 35 %
 R = 08 % (orientés en LP)
 EVAPM3/95 - C19 :
- tableau a) R.E. 3 ; 6 ; 7 ; 4 ; R = 94 % NR = 02 %
 tableau b) R.E. 0,15 ; 0,30 ; 0,35 ; 0,20 ; R = 48 % NR = 02 %
 bâtons de longueurs proportionnelles aux effectifs
 avec indication des fréquences : R = 33 % NR = 02 %

Près de 90% des élèves remplissent sans aucune erreur un tableau qui demande, malgré sa simplicité, un minimum d'analyse. Les difficultés commencent lorsqu'il faut calculer les fréquences, moins d'un tiers répondent correctement. Le mot même de "fréquence" semble servir de repoussoir. A noter des réponses telles que 13ans/58ans (voire 13ans/20ans, ce qui montrerait que "20 ans" est vraiment un âge de référence!) Les 35 % de non-réponses à la question c) proviennent essentiellement de celles déjà observées aux questions précédentes. Mais, parmi les réponses, moins de la moitié sont correctes. Le fait de ne pas donner les axes et leur graduation paralyse et il y a eu des erreurs dans la graduation des fréquences : dès lors les longueurs des bâtons ne sont plus proportionnelles aux fréquences.

Ref : Brochure 3/1990 p83

Clé10071

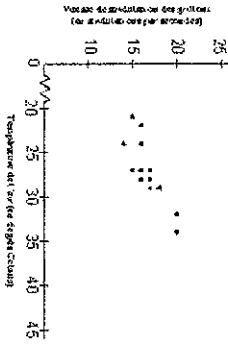
Origine : APMEP
Sens : Objet
Type d'activité : Traductif
Forme : Semi-ouverte
Durée estimée : De 5 à 10 min
Facilité : 80
Complexité : A3
Mots-clés : Statistique;
Capacités : ;
Ages : 17;

Ces scientifiques ont observé que les grillons hognent leur ailes plus vite par temps chaud que par temps froid. En notant la rapidité des stridulations du grillon, il est possible d'estimer la température de l'air. Voici un graphique montrant 13 observations des stridulations d'un grillon par seconde et la température de l'air associée.

Dans cet exercice, aucune copie aide n'est demandée.

- 1° Sur le graphique ci-contre, dessiner la droite qui semble ajuster au mieux ces données.
- 2° En utilisant cette droite, donner une estimation de la température de l'air quand on entend 22 stridulations de grillons par seconde.

Estimation de la température de l'air :



Clé10072

Origine : APMEP
Sens : Objet
Type d'activité : Calculatoire
Forme : Fermée
Durée estimée : De 5 à 10 min
Facilité : 70
Complexité : A4
Mots-clés : Statistique;
Capacités : ;
Ages : 17;

Pour les six dernières années, une grande surface de distribution a enregistré les dépenses publicitaires et les ventes suivantes (valeurs en dizaines de millions de francs)

Expenses publicitaires : x	26	27	29	31	32	35
Ventes : y	4500	4800	4950	5100	5250	5400

- 1° Calculer le coefficient de corrélation entre ces deux variables.
Peut-on estimer qu'un ajustement linéaire soit justifié ?

Répondre sur votre copie (code : B1)

- 2° On admet qu'un ajustement linéaire est intéressant. Déterminer une équation de la droite de régression de y en x obtenue par la méthode des moindres carrés.

Répondre sur votre copie (code : B2)

- 3° Pour l'année suivante, le budget publicitaire est de 350 000 F. À l'aide de l'ajustement précédent, déterminer le montant des ventes que l'on peut espérer atteindre.

Répondre sur votre copie (code : B3)

- 4° Ce résultat paraît peu réaliste : l'entreprise espère plutôt arriver à un montant des ventes de 56 000 000 F.

Toujours à l'aide de l'ajustement précédent, déterminer le montant des dépenses publicitaires que l'on peut envisager pour espérer atteindre ce nouveau montant des ventes.

Répondre sur votre copie (code : B4)

CI610073

Origine : APMEP
Sens : Objet
Type d'activité : Calculatoire
Forme : Fermée
Durée estimée : De 5 à 10 min
Facilité : 30
Complexité : C1
Mots-clés : Pourcentages; Statistique;
Capacités : ;
Âges : 17;

Un lycée compare des séries d'enseignement général et des séries d'enseignement technologique. Dans les séries d'enseignement général, le taux de réussite au Bac est le même pour les filles et pour les garçons. Dans les séries d'enseignement technologique, le taux de réussite au Bac est aussi le même pour les filles et pour les garçons, mais pas forcément le même que dans l'enseignement général.

- Pourriez-vous affirmer que dans l'ensemble du lycée - enseignement général et enseignement technologique réunis - le taux de réussite au Bac est le même pour les filles et pour les garçons ?
- OUI Justifier votre réponse.
 NON Donner un contre-exemple.

Répondez sur votre copie (code : A)

CI610075

Origine : APMEP
Sens : Objet
Type d'activité : Critique
Forme : Fermée
Durée estimée : De 5 à 10 min
Facilité : 40
Complexité : C1
Mots-clés : Moyenne; Statistique;
Capacités : ;
Âges : 17;

La répartition des salaires mensuels (en francs) dans une grosse entreprise de 900 salariés est la suivante (Ho désigne le nombre de salariés) :

Salaires	Ho	Salaires	Nh	Salaires	Nh	Salaires	Nh	Salaires	Nh	Salaires	Nh	Salaires	Nh
7500	7	8750	60	9510	40	9880	15	10100	3	10700	3		
7650	6	8800	30	9600	37	9810	10	10120	5	10800	2		
7700	9	8920	30	9650	38	9820	10	10200	7	10900	2		
8050	8	9100	25	9680	45	9835	20	10250	6	11050	5		
8150	10	9350	20	9720	45	9900	13	10280	5	11200	3		
8210	50	9360	35	9725	44	9950	12	10360	3	11700	1		
8250	40	9380	35	9750	20	9980	5	10310	10	12050	2		
8720	45	9700	34	9750	18	10050	2	10360	5	12700	1		

1°) Sachant que le salaire mensuel moyen n'est pas une des quatre valeurs ci-dessus, indiquez laquelle en expliquant pourquoi vous éliminez les autres. Cochez la bonne réponse :

- 8705 F 9205 F 10305 F 11305 F

Expliquez votre réponse (code : C19)

2°) L'entreprise envisage de augmenter les salaires de 10% pour les salariés à 9700 F de salaire mensuel. Sans effectuer le calcul, indiquez comment vous feriez pour calculer, au centime près, le nouveau salaire moyen dans l'entreprise.

Répondez sur votre copie (code : C27)

2) Documents relatifs à la séance 13

a. Plan de la séance

- I- Bibliographie Laurence Cassou, Marie Noëlle Fourmy, Monique Novak
- II- Moyenne et médiane Anne Allard, Christine Postic, Sylvie Mendes, Emmanuel Rouault
..... Catherine Bauduin, Yann Dartois, Corinne Hecht, Hélène Radzynski, Catherine Turlier, Françoise Verdier
- III- Un scénario de formation..... Sylvine Auclair, Martine Jacquin, Etienne Lakritz, Anne-Marie Sanchez

b. Bibliographie commentée

Cette bibliographie serait donnée au cours d'un stage sur les statistiques, mais il nous a semblé intéressant de le présenter sous la forme :

« LES STATISTIQUES OUI ...mais ... » afin de sensibiliser les stagiaires sur leurs interprétations délicates voire erronées .

Il nous a paru intéressant de classer notre bibliographie en trois parties :

Trois articles d'informations générales qui

- montrent l'énorme potentiel des statistiques,
- posent le problème du choix d'une bonne lecture d'une série statistique et d'une analyse adéquate,
- mettent en garde contre l'interprétation des données chiffrées ,
- montrent l'impact visuel subjectif de certaines représentations .

- « **Les statistiques, le calcul des probabilités et l'enseignement professionnel** » article de **Jean-Louis Piednoir**, paru dans le bulletin APMEP n° 413 de Décembre 1997.

L'objectif de cet article est de proposer une approche de la statistique et du calcul des probabilités proche des préoccupations de l'enseignement professionnel. Il explique l'intérêt des statistiques et pose la question de l'interprétation des résultats d'une enquête statistique suivant les critères retenus. Il expose la démarche statistique et soulève le problème du choix du résumé numérique d'une série statistique quantitative suivant les besoins de l'utilisateur. Il fait ensuite la distinction entre statistique descriptive et statistique inductive. La conclusion est un plaidoyer pour l'enseignement des statistiques en lycée(professionnel).

- « **Les statistiques : de leur utilisation dans le domaine public en remontant vers leurs sources** » article de **René Mulet- Marquis**, paru dans Repères- IREM n°55 d'Avril 2004. APMEP n°451 Mars-Avril 2 004.

L'auteur René Mulet-Marquis, professeur au collège de Corbas, IREM de Lyon, part d'un exemple tiré du débat national sur l'Ecole. Par une réflexion faite en 3 temps (interprétation proposée, remontée aux sources et examen des sources des données) , l'auteur conclut que chaque fois que l'on présente, utilise ou consulte des données chiffrées, il faut faire preuve d'une extrême vigilance en particulier lorsqu'elles sont utilisées en dehors et parfois très loin de leur contexte de production.

- « **A propos de « camemberts » ou ...l'art de manipuler l'information** » article de **François Gouletquer**, paru dans IREM n°48. (DIJON).

L'objectif de cet article est d'attirer l'attention des professeurs et des élèves sur l'impact visuel et subjectif des représentations en 3D.

Il s'agit de la présentation des résultats d'une élection sous la forme de « camembert », les pourcentages n'étant pas indiqués mais la proportionnalité des pourcentages et des aires des secteurs respectée. La perception des résultats n'étant pas, dans l'ensemble, conforme aux vraies proportions et dépendant du graphique présenté, l'auteur émet l'hypothèse que l'œil est plus sensible aux mesures des angles au centre qu'aux aires des secteurs. Cet article est une petite leçon de manipulation de l'information qui montre que l'enseignement des statistiques développe un esprit critique et concourt à l'éducation à la citoyenneté.

Deux articles qui s'adressent plus particulièrement aux collègues de collège qui

- illustrent ce que l'étude des statistiques au collège peut apporter à la formation générale, à la préparation de l'étude des probabilités et à la formation du citoyen,
- montrent les pièges de la moyenne.

- « **Pourquoi il ne faut pas laisser de côté les chapitres de statistiques au collège** » de **Jean-Claude Girard**, paru dans REPERE n°23 d'avril 1996.¹

L'idée de l'article part du constat de négligence des chapitres de statistiques au collège.

Il illustre dans notre titre: « Les statistiques: OUI ». D'une lecture facile, il montre la diversité

des contenus de cet enseignement en liaison avec le domaine mathématique et l'éducation à la citoyenneté, à tous les niveaux. Il insiste sur l'impact des statistiques sur le développement de l'esprit critique au collège et sur l'intérêt de cet enseignement pour faciliter l'étude des probabilités au lycée.

Trois parties sont développées pour illustrer l'importance de ces chapitres délaissés:

- les graphiques
- le calcul des paramètres en liaison avec l'idée de distribution statistique
- le conceptuel avec l'idée de hasard et variabilité des résultats.

- « **les pièges de la moyenne** » article de **Francis Casiro**, paru dans TANGENTE n°77 d'Octobre-Novembre 2000.

Au cours d'une brève présentation étymologique et historique de la moyenne son double statut y est présenté : A savoir un premier aspect lié à la théorie des erreurs en dualité avec un second lié à la recherche d'un paramètre de position. L'interprétation de la moyenne, terme passé dans le langage courant peut engendrer une confusion avec la médiane ou le mode.

Dans ce même ordre d'idée, les pièges de la moyenne aboutissent à un même constat :

La confusion entre les types de moyenne(harmonique, géométrique, quadratique ou arithmétique.)Et dans quel type de situation, la moyenne est-elle un critère pertinent ?

Trois articles s'adressant plus particulièrement aux collègues de lycée qui proposent

- une information sur l'introduction de la démarche statistique dans les nouveaux programmes de lycée,
- un apport historique sur les notions de moyenne, médiane et écart-type.

- Dossier : Statistiques et Probabilités « **A propos de l'introduction de l'enseignement de la statistique dans les lycées** » article de **Claudine Robert**, paru dans le bulletin APMEP n°425 de Novembre-Décembre 1999.

L'objectif de cet article est une information pour expliciter l'introduction dans les programmes de mathématiques des premiers éléments d'une démarche statistique.

En effet, jusqu'à présent, il s'est développé une statistique propre à l'enseignement secondaire où on se limitait au traitement numérique et graphique de l'information chiffrée. D'où l'idée d'introduire dans les nouveaux programmes les premiers éléments d'une démarche statistique, plus proche de celle pratiquée par les statisticiens, notamment la fluctuation d'échantillonnages et des activités Random de la calculatrice.

- « **Moyenne, Médiane, Ecart-Type** » article de **Anne Bové et Marie-Céline Comairas**, paru dans Perères-IREM n°48 de Juillet 2002.

L'objectif de cet article est de donner du sens aux notions de moyenne et médiane : il apporte des informations qui conduisent à une réflexion sur l'enseignement des statistiques.

Il présente les définitions actuelles des mots médiane, moyenne, variance et écart-type proposées dans les programmes et dans les manuels scolaires, puis il nous trace un historique rapide de ces notions à partir de la théorie des erreurs, datant de la fin du 18ème siècle.

Il conclut sur les répercussions possibles de cette étude sur l'enseignement des statistiques dans le secondaire et propose des activités en classe, non expérimentées.

¹ Cet article fait l'objet d'un résumé détaillé et d'une critique.

.c. Réflexion à propos d'un stage "sur site" (plan du document)

Enseigner des statistiques descriptives
Comment dynamiser une équipe de professeurs ?
Réflexions autour d'un mini-scénario

Introduction	2
<i>Première séance : la démarche statistique</i>	4
Document 1 : rapport envergure / taille	7
<i>Deuxième séance : autour de la tendance centrale</i>	8
Document 2 : 4 parcours possibles.....	12
Document 3 : étude de g.....	13
Document 4 : fiche de lecture	14
<i>Troisième séance : rapport Kahane, programmes,</i>	15
Document 5 : fiche de lecture de l'article « A BAS LA MOYENNE »	18
<i>Quatrième séance : les glissements</i>	19
<i>Cinquième séance : les boîtes de distribution</i>	20
<i>Sixième séance : bilans, perspectives</i>	22
<i>Références des ouvrages utilisés</i>	23

1) Activités analysées lors de la séance 1

- TD n°10, Déclic 2000, Pile ou face, évolution des fréquences.....page 63
- Activité 2, Bréal 2000, Tirages dans une urne de composition connue.....page 64
- TD n°7, Déclic 2000, Fluctuation d'échantillonnagepage 65



10 PILE OU FACE : évolution des fréquences

On se propose de regarder comment évolue la fréquence d'apparition de PILE au fur et à mesure des lancers d'une pièce de monnaie.

1 Expérience réelle

a) Lancer une pièce vingt fois de suite en complétant à chaque lancer le tableau suivant.

numéro du lancer	1	2	3	4	20
PILE (1) ou FACE (0)										
nombre de PILE obtenu depuis le premier lancer										
fréquence de PILE depuis le début (à 0,001 près)										

b) Dans un repère, placer les points dont l'abscisse est le numéro du lancer et l'ordonnée est la fréquence de PILE obtenus depuis le premier lancer.

2 Simulation sur la calculatrice

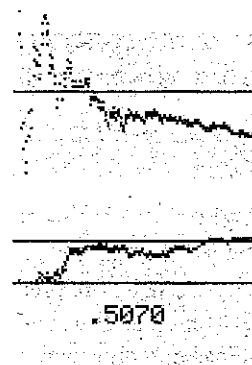
► Voir rabat de couverture, p. V

Ce programme PILE simule le lancer d'une pièce 300 fois de suite, calcule à chaque lancer la fréquence de PILE obtenus depuis le début, puis il place le point correspondant dans un repère, compare la fréquence à 0,5 et affiche la fréquence finale.

TI 80, 82(*), 83	Casio-GRAPH	commentaires
: 0 → Xmin : 300 → Xmax : 100 → Xscl : 0.4 → Ymin : 0.6 → Ymax : 0.1 → Yscl : DrawF 0.5 : 0 → P	ViewWindow 0, 300, 100, 0.4, 0.6, 0.1 ↓ Graph Y=0.5 ↓ 0 → P ↓	Définition de la fenêtre d'affichage et des graduations sur les axes. Tracé de $y = 0,5$. 0 dans le compteur de PILE.
: For (K, 1, 300) : randInt(0, 1) → A	For 1 → K To 300 ↓ Int(2 * Ran#) → A ↓	Début de boucle Lancer de la pièce
: If A=1: P+1 → P	if A=1 Then P+1 → P ↓	Si c'est PILE, ajouter 1 au compteur.
: P / K → F : Pt-On(K, F) : End : Disp F	P / K → ↓ Plot K, F ↓ Next F ▲	Calcul de la fréquence. Affichage du point (K ; F). Fin de boucle. Sortie de la fréquence finale

- Sur TI : Xmin... par **[VARS]** 1:Window...
DrawF par **[DRAW]** DRAW
Pt-on(par **[DRAW]** POINTS
- Sur Casio : ViewWindow par **[V-Window]** V-Win
Graph Y = par **[Sketch]** **[GRPH]** Y=
Plot par **[Sketch]** **[PLOT]** Plot

- Exécuter le programme une première fois.
- Exécuter une seconde fois.
(Sur TI, le graphique précédent est conservé : l'effacer par **[DRAW]** **[ClrDraw]**.)
(Ci-contre deux exemples obtenus avec TI 83.)
- Exécuter plusieurs fois ce programme, et noter la fréquence finale. Commenter les résultats obtenus.
- Collecter les fréquences obtenues dans la classe et donner les valeurs extrêmes des fréquences obtenues.
- Reprendre le programme en remplaçant 300 par 500.



ACTIVITÉ 2

Tirages dans une urne de composition connue

Objectifs :

Par la présentation du principe du sondage, par l'expérimentation de tirages avec remise dans une urne de composition connue puis la simulation :

– introduire la notion de fluctuation d'échantillonnage ;
– faire constater que les différentes valeurs de la proportion observée diffèrent d'autant moins de la vraie proportion que le nombre d'expériences ou de simulations est plus grand.

Le fichier *Tirage.xls* permet de simuler cette expérience à l'aide d'Excel.

Une urne contient des boules, dont 30 % sont rouges. Les boules sont invisibles de l'extérieur et indiscernables au toucher. On peut par exemple mettre dans une boîte 3 boules rouges et 7 boules d'autres couleurs que rouge.

On extrait au hasard successivement plusieurs fois une boule, on note sa couleur et on la remet dans l'urne : de la sorte, la proportion de boules rouges dans l'urne reste toujours la même.

- 1) Effectuez un tel tirage de 40 boules, calculez et notez la proportion de boules rouges.
 - a) Comparez votre résultat à celui de vos camarades. Qu'observez-vous ?
 - b) Cumulez vos observations (en groupes de 2, puis de 3). Comparez à nouveau les résultats. Qu'observez-vous ?
- 2) L'expérience d'un grand nombre de tirages étant longue et fastidieuse, on a recours à sa **simulation** à l'aide d'un ordinateur. Faites la simulation d'un tirage de 100 boules. Comparez les résultats avec ceux de vos camarades. Qu'observez-vous ?
- 3) Faites la simulation de 100 tirages de 100 boules. Observez la proportion obtenue à chaque simulation de 100 tirages et la proportion obtenue pour la simulation de 100, 200, 300, ..., 10 000 tirages. Que pouvez-vous conclure ?

Breal p147 - 2000

Annexe page 64



7 Fluctuation d'échantillonnage

On se propose de mener une enquête statistique complète pour répondre à la question suivante : « Quelle est la distribution des fréquences d'apparition des voyelles a, e, i, o, u et y dans la langue française ? »

Il est clair qu'une telle étude ne peut être faite que sur un échantillon et il est préférable qu'il soit non biaisé : éviter une page de « *La disparition* » de G. PÉREC ! ou d'un livre sur Pythagore qui risque de contenir beaucoup de y.

Chaque élève choisit comme échantillon les dix premières lignes d'une page du livre étudié actuellement en français, dans les pages 100 à 299.

Il n'osa pas dormir dans la maison. Il prit tout un tas d'outils, un chaudron, un hachoir, un tamis, un allumoir, un baril d'alcool, puis il gagna, non loin, dans un taillis, un abri qu'il avait auparavant choisi. Il y bricola tant qu'il put, donnant jour par jour à son installation un tour plus sûr. Il chassait ; il tua un lapin ; il attrapa un jour au lasso un agouti : il fit du lard, du saindoux, du jambon, du boudin.

Un mois passa. La mousson arriva. L'azur s'obnubila ; l'on vit s'amassant à l'horizon strato, nimbo, puis cirro-cumulus. Un haut courant arrivait du bas fond. Un flux montant ...

Extrait de « *La disparition* » de G. PÉREC, éd. Gallimard.

1 Tirage au sort d'un échantillonnage pour chaque élève

a) Lancer une pièce de monnaie :

- si on obtient PILE, on choisit dans les pages de 100 à 199 ;
- si on obtient FACE, on choisit dans les pages de 200 à 299 .

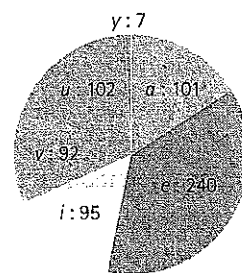
b) Prendre dix cartons où l'on écrit les dix chiffres de 1 à 9 et 0 (ou dix jetons de loto de 1 à 9 et 10 pour le 0).

- On tire au hasard un carton (ou un jeton) : le chiffre sera celui des unités du numéro de page.
- On remet le carton (ou le jeton), puis de nouveau on tire un carton : le chiffre sera celui des dizaines. (Si la page est blanche, ou ne comporte pas 10 lignes, on prendra la page suivante.)

2 Collecte des données

a) Pour chaque voyelle, relever l'effectif. Calculer la taille de l'échantillon, c'est-à-dire le nombre total de voyelles, puis la fréquence de chaque voyelle.

b) Représenter sous forme d'un diagramme circulaire la distribution des fréquences. (Ci-contre, celui de la distribution des effectifs pour le texte de cette page, sauf l'extrait de « *La disparition* ».)



3 Comparaison d'échantillons et regroupement

a) Sur toute la classe, noter l'étendue des fréquences de chaque voyelle.

b) Former des groupes de cinq élèves. Regrouper les cinq relevés pour ne constituer qu'un échantillon. Donner le nouvel effectif de chaque voyelle, et en déduire la distribution des fréquences pour ce nouvel échantillon.

c) Regrouper tous les résultats de chaque groupe et, de la même façon, déterminer la distribution des fréquences pour la totalité de l'enquête.

► Voir Problème 34 et Programme ECHANT, p. 194

■ Pour aller plus loin

Au CDI ou sur Internet, rechercher la distribution des fréquences de chaque lettre en français. Comparer à chacun des échantillons ci-dessus.

Rechercher une raison à la distribution des lettres dans le jeu de Scrabble®.

Annexe page 65

2) **Bilan de la séance 3**

Compte rendu des retours sur les analyses de chapitre de manuels

(consigne : manuel de 2^{de}, chapitre simulation et fluctuation, mathématiques cachées, la lois est-elle sensée être connue ou non, type de situation, questions dégagées)

16 présents

Premier groupe **Transmath 2000**

Il y a deux chapitres de statistiques dont un sur simulation et fluctuation d'échantillonnages. Ce dernier a une structure non typique : il contient essentiellement des activités. Toutes les activités sont sur le même modèle : description et test d'une expérience, regroupement de données, simulation, explication (arbres, tableaux, dénombrements...)

Les exercices s'appuient sur les activités (activités non découpées, très difficiles à comprendre ou à mettre en œuvre, dont les objectifs sont soit vérifier que... soit du type constatation).

Tous les diagrammes présentés sont des diagrammes en bâton. Quelques pages sur les thèmes. Le manuel est très proche des programmes.

Les calculatrices sont très présentes.

Questions :

- quel lien stats/proba ? Pourquoi simuler ? pourquoi stat avant proba ? les élèves ont-ils compris l'intérêt de la simulation ?
- Comment évaluer ce type de travail ?
- Quelle est l'importance des stats ?
- Quel est l'objectif de tout ça ?

Second groupe

Situation du lièvre et de la tortue (Bordas 2000, Hachette 2000, hatier ,programmes)

Il n'y a pas de loi connue (pour les élèves)

Questions :

- On regroupe des échantillons mais comme ça diminue le nombre d'échantillons, que peut-on réellement en tirer ?
- Quel est la valeur de cet échantillon de 300 (30 fois 10 tirages regroupés) ?
- Comment crée-t-on les nombres aléatoires (tables de nombres, calculatrices, ordinateurs...) ?
- Quel est le statut de la programmation nécessaire à la simulation ? Comment l'enseigne-t-on ? Quel temps y passe-t-on ?
- Quels objectifs du programme : préparation aux proba? mode ? donner du sens aux sondages du quotidien ? donner du sens à ce qu'est un modèle , une simulation ? qu'est-ce qu'une compétence dans "concevoir et simuler dans des situations simples " ?
- Problème de gestion, de modalité ; problème d'articulation entre théorie et pratique.

3ème groupe

Abscisse 2000 Magnard (ou 2004 ?)

Questions :

- qu'est-ce qu'un échantillon fiable (exemple d'un sondage avant les élections de délégué : 10 personnes sur 30 sont interrogées)
 - Comment fabrique-t-on un échantillon ? Comment supprime-t-on les biais ?
- [Jacqueline précise tout de suite qu'il s'agit là de techniques « bricolées », statistiques très « appliquées », méthodes complexes]

– Est-ce la taille relative de l'échantillon qui est importante ou bien la taille de la population n'intervient pas ?

Repère 2000, Hachette

On simule pour appréhender. Quelle est la population ? Quel est l'échantillon ? quelle conclusion apporte-t-on aux élèves à la fin de l'activité ?

Dimathème 2000

Qu'est-ce qu'un échantillon suffisamment grand ? Un bon échantillon ?

4ème groupe

Modulo 2004, Didier

Exemple de confusion expérience/simulation : la loi n'est pas connue (répartition filles/garçons à la naissance) et on simule (donc à partir d'une loi connue : 50/50), comparaison d'une série réelle à une loi "plaquée" (loi de Benford)

Quel est l'intérêt de simuler un lancer de pièce ?

Simulation de la situation de deux personnes qui ont le même anniversaire dans la classe.

Les conclusions ne sont jamais demandées

Questions :

– Echantillons : quelle est la différence entre 10 échantillons de 100 et 100 échantillons de 10 ? la fluctuation d'échantillonnage est-elle vue sur le nombre d'échantillons ou sur la taille ?

– Le hasard construit par l'homme l'est comment ? comment rendre acceptable le hasard construit par l'homme ?

3) Documents relatifs à la séance 12

.a. Plan de la séance

- I. BIBLIOGRAPHIE. Odile VIEGAS. Isabelle BREUILLARD. Caroline RAMDANI.
- II. INTRODUCTION et PROGRAMME AU COLLEGE ET EN SECONDE. Dominique HALPERIN
- III. PROGRAMME EN PREMIERE ET EN TERMINALE Catherine STAROSWIECKI
- IV. EXPERIMENTER.MODELISER.SIMULER. Hélène CLAPIER. Géraldine BONAL. Brigitte .YVERT. Dominique BAROUX. Françoise HERAULT .Suzanne GALLAND .Didier GENET.
- V. REFLEXION SUR LA SIMULATION Géraldine NOTTER
- VI. COMMENT JUSTIFIER LA SIMULATION ? Ghislaine SIMON
- VII. ADEQUATION A UN MODELE. Brigitte SOTURA

.b. Bibliographie commentée²

- Sur quelques conceptions du hasard. Bernard COURTEBRAS enseignant et membre de la commission inter -IREM - Chapitre IV d'une thèse de Doctorat en Sociologie des mathématiques.

Ce chapitre présente le hasard de façon historique ; il est découpé en six parties : les cinq premières de façon étimologique et la sixième d'un point de vue mathématiques.

- 1) l'origine du mot hasard
- 2) la pensée qui nie le hasard avec les oppositions primitif/civilisé et enfant/adulte.
- 3) Le hasard est le destin
- 4) Le hasard et ses différents courants de pensée
- 5) Le hasard est une réalité objective avec les différents point de vue philosophiques (Démocrite, Aristote et Cournot)

² Les articles non commentés et présents dans la bibliographie de la brochure ont été supprimés de la liste originale.

6) Catégorisation dans l'approche mathématique contemporaine avec d'autres types d'exemples du hasard (tirage au sort, bénin, sauvage, lent et formel)

- *Les nombres pseudo-aléatoires* Michel JANVIER de l'équipe informatique à l'IREM de Montpellier
L'utilisation des calculatrices (touche Random) et des ordinateurs (fonction aléa pour Excell) permet d'obtenir des nombres dits pseudo aléatoires. Cet article expose une méthode pour trouver ces nombres et montre, grâce à l'inégalité de Bienaymé Tchebichev, qu'il faut un grand nombre de répétition pour obtenir une précision satisfaisante. Les nombres pseudo aléatoires permettent d'effectuer des simulations d'expériences aléatoires.

- *Statistique au Lycée Bernard PARZYSZ co Edition APMEP ADIREM (livre à paraître)*
L'introduction des probabilités au lycée : un processus de modélisation comparable à celui de la géométrie
Michel HENRY de l'IREM de Franche-Comté - REPERE-IREM n°36 -1999.
La première partie constitue un rappel historique des différentes conceptions des probabilités et permet à l'auteur de montrer l'analogie du point de vue épistémologique, entre les probabilités et la géométrie, toutes deux issues de la pratique sociale.
La deuxième et la troisième partie concerne les choix des programmes alors en vigueur et leurs conséquences didactiques (pas inintéressant mais les programmes ont changé...)

- *La simulation en statistique* Philippe DUTARTE, professeur en lycée Technologique – REPERE-IREM n°47 Avril 2002³
Cet article donne des définitions et des éclaircissements mathématiques sur différents thèmes : simulation, modélisation, hasard, générateur de nombre aléatoire, comment justifier la simulation, ce que peut apporter la simulation à l'enseignement de la statistique, fourchette ...

- *Simulation et modélisation* Joëlle FONTANA et Maryse NOGUES- REPERE-IREM n° 46 - année 2002

- *Simulation et modélisation d'un problème de bouteilles* François HUGUET de l'IUFM de Quimper - REPERE-IREM n°36 – juillet 1999
L'auteur a sans doute été motivé par ses lectures personnelles, et en particulier par les idées d'Engel. L'article est complété par les remarques et observations de Régis Gras, prof de l'université et de l'IREM de Rennes.
Il expose un exemple de simulation et modélisation d'un problème de bouteilles, problème très riche, adaptable à plusieurs niveaux scolaires, il donne des idées et des pistes de travail mais en aucun cas une activité clé en main.

- *Probabilités-suites numériques et programmation* Michel Bourguet de l'IREM de Strasbourg - REPERE-IREM n°41 - Octobre 2000
L'auteur Michel Bourguet de l'IREM de Strasbourg décrit dans cet article trois activités pour une durée de six à dix heures, activités qu'il a pratiquées en classe de première STI, BTS et en terminale L.
Sont proposées deux situations de jeux, dé et jeu de l'oie, et une situation où l'aléatoire est moins visible (problème d'anniversaires). Ces situations font très naturellement débat dans la classe ; elles permettent la rencontre des suites et des probabilités ; l'envie de connaître la réponse aux questions posées incite à l'écriture de programmes de calcul effectif. L'article est enrichi de quelques réflexions pédagogiques de l'auteur, sur la modélisation et sur les processus d'abstraction en probabilités.
Article intéressant agréable à lire ; de bonnes idées de séquences d'enseignement en première et terminale qui sont encore d'actualité avec les programmes 2000.

- *Des statistiques à la pensée statistique* Université de Montpellier II - IREM de Montpellier 2001
Cet ouvrage est constitué de quatre parties :
· La première partie traite l'approche historique de « cette pensée statistique » et aborde des questions tel que, qu'est ce que le hasard, qu'est ce qu'un échantillon, que veut dire modéliser, simuler....
· La deuxième partie traite un point de vue didactique, elle donne un éclairage sur le choix et la façon d'aborder les nouveaux programme de seconde.
· La troisième partie présente des exemples de modélisation et de simulation.
· La quatrième partie constitue une biblio, c'est un outil précieux pour les enseignants avec quelques activités clé en main.

Publimath.irem.univ-mrs.fr/

³ Cet article fait l'objet d'un résumé détaillé et d'une critique.

4) *Expérience mise en oeuvre pendant la séance*

a. Descriptif

Scénario Spaghettis

D'après l'article de Joëlle Fontana et Maryse Nogues
Simulation et modélisation
Etude d'un exemple
IREM de Montpellier

Rappel

- On choisit avant de couper l'emplacement des 2 coupures → simulation 1
- On coupe un premier morceau, et on recoupe le second → simulation 2
- On coupe un premier morceau et on recoupe le plus grand des deux → simulation 3
- On coupe un premier morceau et on recoupe le plus petit des deux → simulation 4

Simulation 1.

On considère que couper un spaghetti en trois, c'est découper un segment, que l'on peut supposer de longueur 1, en trois morceaux, en plaçant deux points M et N d'abscisse x et y sur ce segment. Il s'agit donc de trouver deux nombres x et y , compris entre 0 et 1 qui sont les abscisses des deux coupures et qui déterminent les longueurs des trois morceaux, notés a , b et c . On obtient :

$$\begin{aligned}a &= \min(x, y); \\b &= \max(x, y) - \min(x, y); \\c &= 1 - (a + b).\end{aligned}$$

Les morceaux qui permettent la construction d'un triangle sont donc les morceaux de longueurs a , b et c qui satisfont les trois inégalités triangulaires. Ceci conduit à la condition :

$$\max(a, b, c) < \frac{1}{2}$$

Preuve : Supposons que a soit le maximum de a , b et c . On doit avoir $a < b + c$, soit $a < 1 - a$, ce qui donne donc $a < 1/2$.

Le même raisonnement s'applique si le maximum est b ou c . La condition $\max(a, b, c) < 1/2$ est bien satisfaite. \square

Calcul de la valeur théorique

Approche graphique

Soient x et y deux nombres compris entre 0 et 1. L'ensemble des points $M(x, y)$ possibles est donc le carré de côté 1 construit sur les unités (cf. figure ci-après).

En raison de la symétrie du problème, on peut supposer que x est le plus petit des deux nombres.

On obtient ainsi comme ensemble de points possibles le triangle « supérieur » du carré obtenu.

Le maximum des trois nombres x , $y - x$ et $1 - y$ doit être inférieur à $1/2$. Chacun d'eux doit donc l'être. On a ainsi trois inéquations :

$$\begin{cases} x < \frac{1}{2} \\ y < x + \frac{1}{2} \\ y > \frac{1}{2} \end{cases}$$

Par régionnement du plan, on obtient le triangle foncé dont l'aire est égale à $1/8$. Le rapport entre l'aire du triangle foncé et celle du triangle hachuré étant, lui de $1/4$.

Approche géométrique du problème

Si ABC est un triangle équilatéral, et M un point quelconque à l'intérieur du triangle, on a :

$$AH = MI + MJ + MK$$

Preuve :

Pour un triangle XYZ , on note A_{XYZ} son aire.

$$A_{AMC} = \frac{AC \times MJ}{2}; \quad A_{BMC} = \frac{BC \times MI}{2}; \quad A_{AMB} = \frac{AB \times MK}{2};$$

Or, ABC est équilatéral, donc

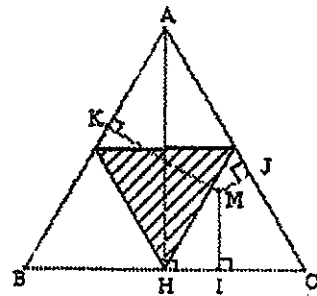
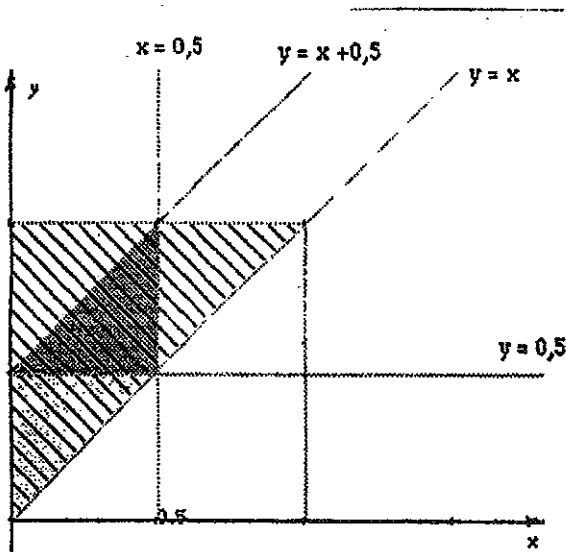
$$A_{ABC} = \frac{AC \times AH}{2} = \frac{AC(MJ + MI + MK)}{2}$$

D'où

$$AH = MI + MJ + MK$$

Couper un spaghetti de longueur AH correspond donc à placer un point M de façon quelconque dans le triangle. On reporte ensuite les trois longueurs MI , MJ puis MK sur la longueur AH pour obtenir les trois morceaux.

Ces trois longueurs permettront de construire un triangle si chacune d'elles est inférieure à la moitié de AH . Ainsi le point M doit se trouver dans le triangle des « milieux » dont l'aire est bien égale au quart de l'aire du triangle ABC .



Simulation 2.

On peut aussi considérer que couper un spaghetti en trois, supposé toujours de longueur 1, c'est couper un premier morceau de longueur a , $a \in [0, 1]$, puis un second morceau de longueur b , $b \in [0, 1 - a]$, la troisième longueur étant donnée par $c = 1 - (a + b)$.
Pour vérifier que les trois segments obtenus forment un triangle, le test reste le même :

$$\max(a, b, c) < \frac{1}{2}$$

Calcul de la valeur théorique

Approche graphique

$x \in [0, 1]$ Soit x le premier nombre aléatoire ($x \in [0, 1]$) et soit y le second ($y \in [0, 1]$), les longueurs des segments sont :

$$a = x \quad b = (1 - x)y \quad c = 1 - (a + b).$$

La construction du triangle est possible lorsque les trois segments ont des longueurs inférieures à 0,5, soit :

$$\begin{cases} x < \frac{1}{2} \\ (1 - x)y < \frac{1}{2} \\ x + (1 - x)y > \frac{1}{2} \end{cases}$$

ou encore :

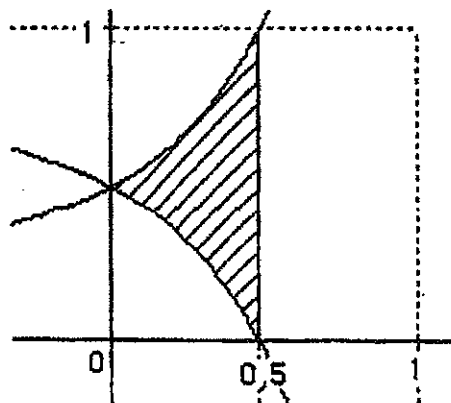
$$x < \frac{1}{2} \quad y < \frac{1}{2(1-x)} \quad y > \frac{1-2x}{2(1-x)}$$

C'est à dire lorsque les couples (x, y) sont les coordonnées des points de la zone hachurée ci-dessous.
L'aire du domaine est égale à : $\ln(2) - 0,5$ valeur qui à 10^{-3} près vaut 0,193.

Preuve :

Soit \mathcal{A} l'aire du domaine hachuré.

$$\begin{aligned} \mathcal{A} &= \int_0^{0,5} \left(\frac{1}{2(1-x)} - \frac{1-2x}{2(1-x)} \right) dx \\ &= \int_0^{0,5} \frac{x}{1-x} dx \\ &= \int_0^{0,5} \frac{x-1}{1-x} + \frac{1}{1-x} dx \\ &= \left[-x \right]_0^{0,5} - \left[\ln(1-x) \right]_0^{0,5} \\ &= \ln(2) - 0,5 \end{aligned}$$



Simulation 3.

Approche graphique

Lorsque le premier nombre aléatoire x est inférieur à $1/2$, la situation est identique à celle de la simulation 2. En appelant y le second nombre aléatoire, les trois morceaux ont pour longueur :

$$a = x \quad b = (1-x)y \quad c = 1 - (a+b)$$

Ce qui conduit au même système d'inéquations.

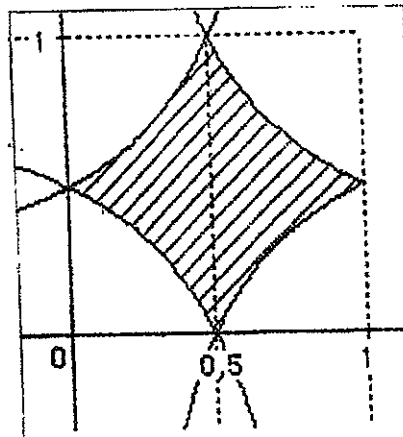
Lorsque le premier nombre aléatoire x est supérieur à $1/2$, on a :

$$a = 1 - x \quad b = xy \quad c = 1 - (a+b)$$

Il est alors possible d'obtenir un triangle lorsque :

$$x > 1/2 \quad y < 1/(2x) \quad y > 1/(2x)$$

Les points du domaine hachuré de la figure ci-dessous, ont des coordonnées $(x; y)$ qui permettent de construire un triangle. Par symétrie avec le cas précédent, cette aire est $2(\ln(2) - 0.5)$ ce qui en valeur approchée à 10^{-3} près donne : 0,386.



.b. Programmes des simulations

SPAGHETI	SPAGHET2	SPAGHET3
IMPUT « N= » , N Seq(X,X,1,N,1) → L1 Seq (rand,X,1,N,1) → L2 Seq (rand,X,1,N,1) → L3 Min (L2 , L3) → L4 Max (L2, L3) – L4 → L5 1 – (L4 + L5) → L6 Max (Max (L4, L5) , L6) → L7 1 – Int(L7 + 0,5) → L8 Disp(Sum(L8)/N).	IMPUT « N= » , N Seq(X,X,1,N,1) → L1 Seq (rand,X,1,N,1) → L2 Seq (rand,X,1,N,1) → L3 L3 × (1 – L2) → L4 1 – (L2 + L4) → L5 Max (L2, Max(L4 , L5) → L6 1 – Int(L6 + 0,5) → L7 Disp(Sum(L7)/N).	IMPUT « N= » , N Seq (rand,X,1,N,1) → L1 Min(L1, 1 – L1) → L2 Seq (rand,X,1,N,1) → L3 L3 × (1 – L2) → L4 1 – (L2 + L4) → L5 Max (L2, Max(L4 , L5) → L6 1 – Int(L6 + 0,5) → L7 Disp(Sum(L7)/N).

Pour tout renseignement sur les publications diffusées par notre IREM,

Vous pouvez soit :

Consulter notre site WEB

<http://www.ccr.jussieu.fr/iremParis7/welcome.html>

Demander notre catalogue en écrivant à

IREM Université Paris 7

Case 7018

2 place Jussieu

75251 Paris cedex 05

TITRE :

Former à...enseigner la statistique au lycée
perspectives en formation de formateurs et en formation d'enseignants

AUTEUR :

C. Hache, J. Mac Aleese, A. Robert

RESUME :

Dans ce cahier, nous allons montrer des choix effectivement faits en formations de formateurs et d'enseignants en décrivant trois scénarios de formation de formateurs et un stage de formation continue inscrit au PAF. Nous présenterons d'abord quelques réflexions sur la statistique et son enseignement au lycée, puis nous indiquerons un certain nombre de modalités assez générales mises en œuvre dans les formations ainsi que des activités précises proposées aux participants. Nous terminerons en indiquant les chroniques des formations concernées. Des annexes permettent de présenter certains des documents utilisés.

MOTS CLES :

didactique, enseignement, formation, hasard, lycée, statistique, statistique descriptive,

Editeur : IREM
Université PARIS 7-Denis Diderot
Directeur responsable de la
publication : R.CORI
Case 7018 - 2 Place Jussieu
75251 PARIS Cedex 05
Dépôt légal : 2005
ISBN : 2-86612-268-2