Références au programme

Programme officiel

Contenus	Compétences exigibles	Commentaires
Statistique		Il s'agit essentiellement, d'une part de faire acquérir aux élèves les premiers outils de comparaison de séries statistiques, d'autre part de les habituer à avoir une attitude de lecteurs responsables face aux informations de nature statistique.
Caractéristiques de position d'une série statistique	Une série statistique étant donnée (sous forme de liste ou de tableau, ou par une représentation graphique), proposer une valeur médiane de cette série et en donner la signification.	On repère, en utilisant effectifs ou fréquences cumulées, à partir de quelle valeur du caractère on peut être assuré que la moitié de l'effectif est englobée. Les exemples ne devront soulever aucune difficulté au sujet de la détermination de la valeur de la médiane.
Approche de caractéristiques de dispersion d'une série statistique	Une série statistique étant donnée, déterminer son étendue ou celle d'une partie donnée de cette série.	L'étude de séries statistiques ayant même moyenne permettra l'approche de la notion de dispersion avant toute introduction d'indice de dispersion. On introduira l'étendue de la série ou de la partie de la partie de la série obtenue après élimination des valeurs extrêmes. On pourra ainsi aborder la comparaison de deux séries en calculant quelques caractéristiques de position et de dispersion, ou en interprétant des représentations graphiques données.
Initiation à l'utilisation de tableurs-grapheurs en statistique		Les tableurs que l'on peut utiliser sur tous les types d'ordinateurs permettent, notamment en liaison avec l'enseignement de la technologie, d'appliquer de manière rapide à des données statistiques les traitements étudiés.

Textes d'accompagnement

La place des statistiques dans la vie courante et leur utilisation dans de nombreuses disciplines demandent que la formation du futur citoyen se poursuive en ce domaine : on le fait en abordant le problème de la comparaison de séries statistiques, avec une première approche de la notion de dispersion.

Le contenu et les commentaires du programme concernant la statistique constituent un prolongement de ceux des classes antérieures, l'objectif de l'enseignement de statistique descriptive au collège étant indiqué dans le document d'accompagnement des programmes du cycle central.

En classe de $3^{\text{ème}}$, il s'agit d'aider les élèves à franchir une nouvelle étape dans le développement de leur autonomie de jugement à propos d'informations qui peuvent être nombreuses. Dans le cas d'un regroupement en classes, les choix effectués peuvent avoir des effets sur les résultats numériques ou les représentations graphiques et leurs interprétations. En classe de $4^{\text{ème}}$, on a pu observer que " la moyenne d'une population dont les éléments sont rangés par ordre croissant ne sépare pas ceux-ci, en général, en deux parties de même effectif ", ce qui justifie l'introduction de la médiane en classe de $3^{\text{ème}}$. Les élèves disposent alors de deux indicateurs de la tendance centrale d'une population, leur position relative pouvant faire l'objet d'une interprétation dans des situations appropriées.

La nécessité de distinguer deux séries statistiques de même tendance centrale justifie l'intérêt de la notion de dispersion. Dans ce premier contact, le programme se limite à l'étendue d'une série statistique ou à l'étendue d'une partie donnée de celle-ci ; cela permet, sans difficulté technique, de familiariser les élèves avec une démarche habituelle en statistique : procéder à une synthèse de l'information sous la forme de nombres mesurant respectivement la position et la dispersion de la série étudiée.

Choix de la représentation d'une série statistique, interprétation des résultats obtenus et comparaison de deux séries statistiques peuvent être conduits, sans répétitions inutiles ni pertes de temps, en utilisant des tableurs-grapheurs ou en répartissant le travail au sein de la classe. De plus, outre son intérêt spécifique, l'enseignement des statistiques contribue au développement des compétences en mathématiques, notamment celles liées au calcul et à la construction, la lecture et l'utilisation de graphiques ; toutes les capacités correspondantes peuvent être mises en oeuvre au cours d'activités interdisciplinaires.

1 Caractéristiques de position d'une série statistique

1.1 Moyenne d'une série statistique

- Série représentée en liste
  Une classe de $5^{\text{ème}}$ a fait un contrôle de mathématiques. Voici la liste des notes obtenues : 8, 11, 5, 12, 2, 17, 7, 8, 19, 2, 10, 4, 7, 7, 11, 14, 7, 8, 12, 5, 10, 8, 5, 10, 11, 8, 12, 10, 14, 12.
  Si l'on veut calculer la moyenne de ce contrôle, on additionne les notes et on divise par le nombre de notes :

$$\frac{8+11+5+\ldots+10+14+12}{30}=\frac{276}{30}=9,2$$

- Série regroupée par valeurs dans un tableau
  On peut aussi regrouper ces nombres dans le tableau suivant selon les valeurs des notes obtenues :

Notes	2	4	5	7	8	10	11	12	14	17	19
Effectif	2	1	3	4	5	4	3	4	2	1	1

  La moyenne est alors **une moyenne pondérée** par les effectifs ; on la calcule en multipliant chaque valeur par l'effectif correspondant, en faisant la somme de ces produits et en divisant par l'effectif total, soit :

$$\frac{2\times 2+4\times 1+5\times 3+\ldots+14\times 2+17\times 1+19\times 1}{30}=\frac{276}{30}=9,2$$

– Série regroupée par classe dans un tableau
  On peut enfin regrouper ces nombres en classes dans le tableau suivant :

Notes	$0 \leq n < 5$	$5 \leq n < 10$	$10 \leq n < 15$	$15 \leq n \leq 20$	TOTAL
centre de la classe	2,5	7,5	12,5	17,5	
Effectif	3	12	13	2	30

  On considère alors que le centre de chaque classe représente la classe donc la moyenne de la série ainsi regroupée en classesest égale à :

$$\frac{2,5\times 3+7,5\times 12+12,5\times 13+17,5\times 2}{30}=\frac{295}{30}\approx 9,83$$

Remarques :
– La moyenne et la moyenne pondérée par les effcetifs sont égales.
– Le regroupement en classes permet des calculs plus rapides mais ne permet pas d'obtenir la valeur exacte de la moyenne.

1.2 Médiane d'une série statistique

Définition La **médiane** $\mathcal{M}$ d'une série statistique est la valeur qui partage le groupe étudié en deux sous-groupes de même effectif chacun tels que :
– tous les éléments du premier groupe on des valeurs inférieures ou égales à $\mathcal{M}$;
– tous les éléments du deuxième groupe ont des valeurs supérieures ou égales à $\mathcal{M}$.

Application :
- **Détermination de la médiane d'une série statistique**
  – **A partir d'un tableau d'effectifs cumulés ou de fréquences cumulées**
    Exemple :
    Sur une population de 75 feuilles de platane, on étudie la longueur en mm de la grande nervure. On obtient le tableau statistique suivant :

Longueurs	102	112	122	132	142	152	162	172	182
Effectif	1	6	6	10	13	19	10	8	2
Effectif cumulé croissant	1	7	13	23	36	55	65	73	75

Les feuilles étant rangées par longueurs croissantes, la case grisée indique que de la $37^{\text{ème}}$ à la $55^{\text{ème}}$, les feuilles ont pour longueur 152 mm.
Or $75 = 37 + 1 + 37$ donc la médiane est la longueur de la $38^{\text{ème}}$ feuille c'est-à-dire 152 mm.

$$\begin{array}{ll} \text{Rang :} & 1^{re}\ \ 2^{e}\ \ \ldots\ 36^{e}\ 37^{e}\ 38^{e}\ 39^{e}\ \ldots\ 75^{e} \\ \text{Longueur :} & \underbrace{102\ 112\ \ldots\ 142\ 152}_{37\text{ feuilles}}\ {\color{red}152}\ \underbrace{152\ \ldots\ 182}_{37\text{ feuilles}} \end{array}$$

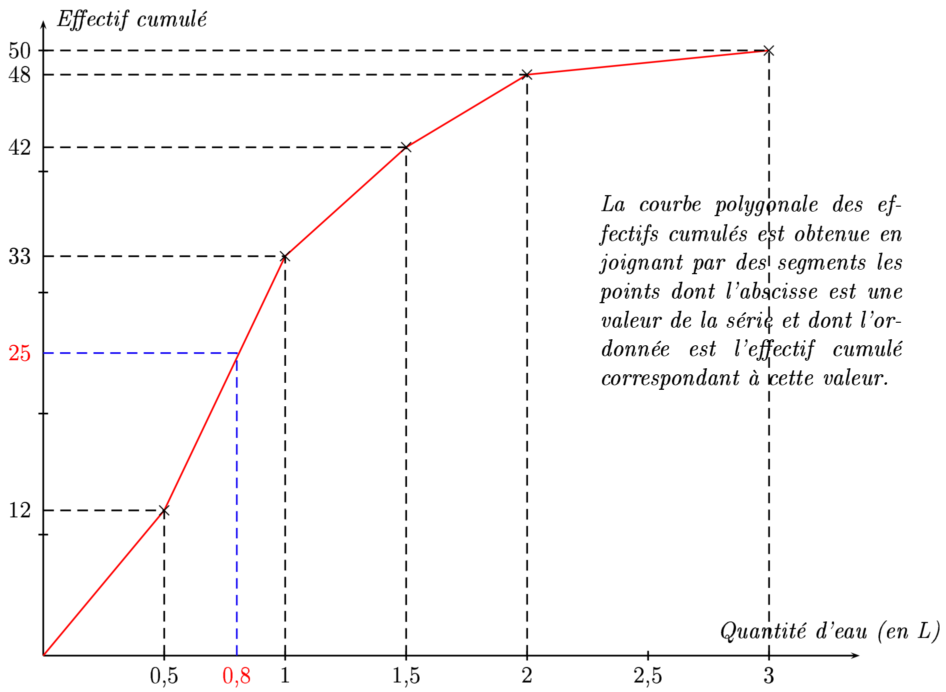
D'autre part, la longueur moyenne de ces feuilles est de :

$$\bar{l} = \frac{102 \times 1 + 112 \times 6 + 122 \times 6 + 132 \times 10 + 142 \times 13 + 152 \times 19 + 162 \times 10 + 172 \times 8 + 182 \times 2}{75}$$

$$\bar{l} = \frac{10920}{75} = 145,6$$

donc la médiane et la moyenne sont (en général) **différentes**.

- **À partir d'une représentation graphique**
  Une valeur approchée de la médiane peut être obtenue à l'aide de la courbe polygonale des effectifs cumulés croissants (ou des fréquences cumulées) en lisant la valeur correspondant à la moitié de l'effectif total (ou à une fréquence cumulée égale à 50%) :
  À la question "Quelle quantité d'eau buvez-vous par jour ?", les cinquante personnes interrogées ont donné des réponses qui ont permis de tracer le polygone des effectifs cumulés croissants suivant :



$\mathcal{M}$ est environ égal à 0,8 L : en effet, la moitié des personnes interrogées consomme moins de 0,8 L par jour (ou la moitié des personnes interrogées consomme plus de 0,8 L par jour).

2 Une caractéristique de dispersion : l'étendue

Définition **L'étendue** d'une série statistique est égale à la différence entre la plus grande et la plus petite valeur prises par cette série. Elle mesure la "dispersion" de cette série.

Les critères de dispersion indiquent de quelle façon les valeurs du caractère sont groupées (plus ou moins resserrées) autour des valeurs "centrales" : l'étendue est le plus simple de ces paramètres. En général, lorsque l'étendue est élevée, la dispersion est grande.
Exemple : pour la série avec les platanes, l'étendue est de $182 - 102 = 80$ mm.

3 Notions de quatrième

3.1 Effectifs et fréquences

Définition Dans un tableau statistique, l'**effectif** est le nombre de réponses associées à chaque valeur.
L'ensemble des valeurs et des effectifs forme une **série statistique**.
En divisant l'effectif d'une valeur par l'effectif total, on obtient la **fréquence**.
En multipliant par 100 la fréquence, on obtient la **fréquence en pourcentage**.

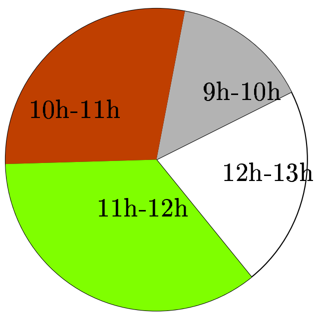
La standardiste d'une radio FM a noté le nombre d'appels téléphoniques reçus par tranches d'heures au cours d'une matinée. Elle obtient les résultats suivants :

Tranches horaires	9h-10h	10h-11h	11h-12h	12h-13h	**Total**
Effectifs (nombres d'appels)	19	37	46	28	**130**
Mesures d'angles au centre en degrés	52,6	102,5	127,4	77,5	**360**

1. Compléter le tableau.
2. Représenter cette série statistique par un diagramme circulaire

Réponse à la question 2 :
Pour construire un diagramme circulaire, il faut calculer les mesures des angles de chaque valeur. Les angles au centre du disque sont proportionnels aux effectifs donc il s'agit de compléter un tableau de proportionnalité. Sachant qu'un tour complet mesure 360° et que celui-ci doit représenter l'effectif total 130, il faut compléter le tableau suivant :

On obtient alors le diagramme suivant :



3.2 Effectifs et fréquences cumulés

Définition Lorsque les valeurs sont rangées dans l'ordre croissant, on obtient **l'effectif cumulé croissant** d'une valeur en additionnant son effectif à ceux qui le précèdent (on additionne à partir de la gauche du tableau).
De la même manière, les **fréquences cumulées croissantes** s'obtiennent en divisant l'effectif cumulé croissant par l'effectif total.
Pour obtenir les effectifs ou les fréquences **cumulés décroissants**, on additionne à partir de la droite du tableau.

Remarque : les effectifs cumulés croissants indiquent quel est l'effectif de la série dont la valeur est inférieure à une valeur donnée.
Exemple : On reprend l'exemple précédent :

Tranches horaires	9h-10h	10h-11h	11h-12h	12h-13h
Effectifs (nombres d'appels)	19	37	46	28
Effectifs **cumulés croissants**	**19**	19 + 37 = **56**	19 + 37 + 46 = **102**	**Total des appels : 130**

Tranches horaires	9h-10h	10h-11h	11h-12h	12h-13h
Fréquences en %	14,6	28,5	35,4	21,5
Fréquences **cumulées décroissantes**	**Total des appels : 100**	28,5 + 35,4 + 21,5 = **85,4**	35,4 + 21,5 = **56,9**	21,5

3.3 Moyenne

Définition La **moyenne** d'une série statistique est le quotient de la somme de toutes les valeurs par l'effectif total de cette série :

$$\text{moyenne} = \frac{\text{somme de toutes les valeurs}}{\text{effectif total de la série}}$$

Exemple : Dans une usine, sept employés calculent le salaire moyen (en €) des salaires de leur atelier.

$$\frac{760 + 825 + 915 + 990 + 1065 + 1160 + 1296}{7} \approx 1002$$

Le salaire moyen des employés de cet atelier s'élève environ à 1002 €.

3.4 Moyenne pondérée

Définition La **moyenne pondérée** d'une série statistique est le quotient de la somme des valeurs, affectées chacune de leur coefficient, par la somme totale des coefficients.

Exemple : Dans une classe de 28 élèves, les notes à un devoir se répartissent de la manière suivante :

Notes	6	7	8	9	10	11	12	13	15	17
Effectifs	2	1	2	4	6	3	3	3	2	2

Nous représentons cette série par un diagramme en barres (FIG. 1).

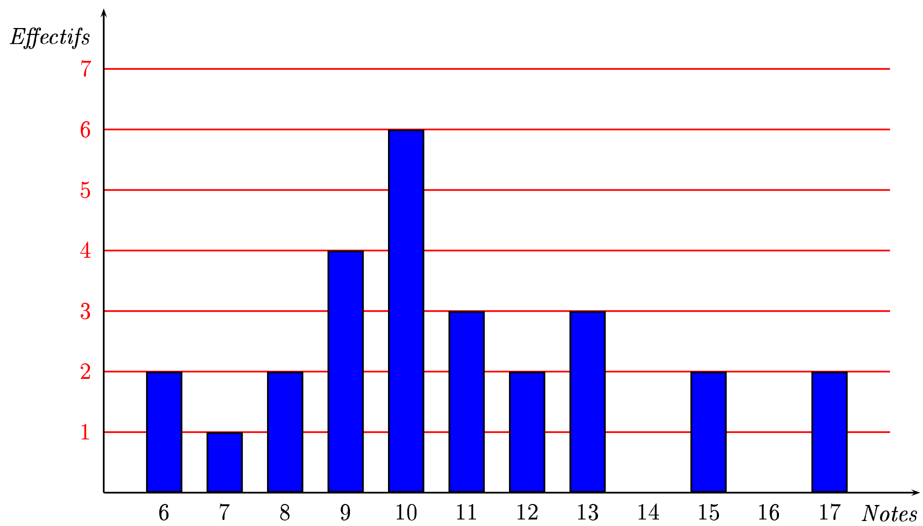


FIG. 1 – **Diagramme en barres :** La hauteur des barres est égale aux effectifs

Pour calculer la moyenne, on effectue le calcul suivant :

$$\frac{6\times 2+7\times 1+8\times 2+9\times 4+10\times 6+11\times 3+12\times 3+13\times 3+15\times 2+17\times 2}{2+1+2+4+6+3+3+3+2+2}=\frac{303}{28}\approx 10,8$$

Pour calculer la moyenne pondérée d'une série de valeurs, il faut :
- calculer les produits de chaque valeur par leur coefficient ;
- calculer la somme de ces produits ;
- puis diviser le résultat par la somme des coefficients.

3.5 Répartition en classes et moyenne

Il arrive dans certains cas qu'une série statistique soit répartie en classes, c'est-à-dire que l'on prend des intervalles de valeurs :
Exemple : Le tableau ci-dessous présente la répartition de 2000 adultes suivant leur taille :

Tailles en cm	[140 ; 150[	[150 ; 160[	[160 ; 170[	[170 ; 195[
Effectifs	48	397	913	642

Il est impossible a priori de calculer la moyenne de cette série puisqu'on ne connait pas les valeurs des tailles et leurs effectifs. On considère alors que la valeur au centre de la classe va représenter la classe. On calcule alors la moyenne pondérée pour obtenir une valeur approchée de la moyenne de la série.

Tailles en cm	[140 ; 150[	[150 ; 160[	[160 ; 170[	[170 ; 195[
Centre de la classe	145	155	165	182,5
Effectifs	48	397	913	642

Pour calculer la taille moyenne, on effectue le calcul suivant :

$$\frac{145\times 48+155\times 397+165\times 913+145\times 642}{2000}\approx 168$$

La taille moyenne de ce groupe d'adultes est d'environ 168 cm.

4 Eléments historiques et culturels

Les activités statistiques remontent à la plus haute Antiquité (en Égypte vers -1700 et en Chine encore avant peut-être) et correspondent à des recueils de données chiffrées dont le but était de fournir des renseignements sur la population. Le mot "statistique" appartient au langage administratif et apparaît en 1664 dans les premières enquêtes lancées par Colbert, ministre de Louis XIV, pour recenser la population, les fermages et les animaux sur toute la France. Connaître et expliquer les phénomènes sociaux et économiques nécessite le besoin de représenter pour visualiser, d'où diverses formes de graphiques, en barres ou en secteurs, mais aussi le besoin de résumer la masse de données recueillies. Plusieurs paramètres sont alors créés : le "mode" (valeur apparaissant le plus fréquemment) et des "moyennes".
C'est à l'astronomie que l'on doit l'origine de ces concepts. Les Babyloniens et les Grecs avaient repéré la position de certaines étoiles et la périodicité de leur mouvement après un grand nombre de mesures et calculs. Tycho Brahé (1546-1601) fait, lui aussi, de multiples observation d'une même quantité pour en estimer une valeur la plus juste possible. Il cherche à éliminer le maximum d'erreurs dans ses observations et calculs et, pour estimer la valeur qui lui paraît la plus juste possible, il utilise la moyenne arithmétique.
Les sondages n'arriveront que plus tard au début du XIX$^{\text{ème}}$ lorsque des méthodes de relevés partiels permettront de rendre compte aussi précisément des phénomènes.
De nos jours, les méthodes statistiques sont utilisées en médecine et en pharmacie (efficacité d'un traitement ou d'un médicament), en politique (sondages avant élections), dans l'industrie (contrôle de qualité, gestion des stocks, marketing) et dans de très nombreux domaines.