Leçon d'oral du CAPES de mathématiques

Statistique descriptive à une variable

Niveau : lycée et approfondissements

Table des matières

1 Introduction **3**

2 Prérequis et objectifs **3**
2.1 Prérequis . . . 3
2.2 Objectifs . . . 3
2.3 Objectifs didactiques . . . 4

3 Vocabulaire et organisation des données **4**
3.1 Population, individu et caractère . . . 4
3.2 Série statistique et effectifs . . . 4
3.3 Effectifs et fréquences cumulés . . . 5

4 Représentations graphiques **5**
4.1 Diagramme en bâtons . . . 5
4.2 Diagramme circulaire . . . 5
4.3 Histogramme . . . 5
4.4 Diagramme en boîte . . . 5

5 Indicateurs de position **6**
5.1 Mode . . . 6
5.2 Moyenne arithmétique . . . 6
5.3 Médiane . . . 6
5.4 Quartiles et quantiles . . . 7

6 Indicateurs de dispersion **7**
6.1 Étendue . . . 7
6.2 Écart interquartile . . . 7
6.3 Variance . . . 7
6.4 Écart-type . . . 8
6.5 Propriétés de la variance . . . 8

7 Exemple détaillé **8**
7.1 Tableau statistique . . . 8
7.2 Moyenne . . . 9
7.3 Médiane et quartiles . . . 9
7.4 Étendue . . . 9
7.5 Variance et écart-type . . . 9
7.6 Bilan de l'exemple . . . 9

8 Comparaison de distributions **10**

9 Mise en œuvre avec un tableur **10**

10 Aspects didactiques et pédagogiques **10**
10.1 Objectifs d'apprentissage 10
10.2 Difficultés fréquentes 11
10.3 Proposition de démarche pédagogique 11

11 Ouverture et limites **11**

12 Conclusion **11**

1 Introduction

La statistique descriptive regroupe les méthodes permettant de recueillir, organiser, résumer et représenter des données observées sur une population ou un échantillon.

L'étude d'une variable consiste à examiner un caractère statistique associé aux individus d'une population : âge, taille, note obtenue à une évaluation, durée d'attente, nombre de livres lus, etc.

L'objectif est de répondre à plusieurs questions :

— Quelle est la valeur centrale de la distribution ?

— Comment les observations sont-elles réparties ?

— Quelle est l'importance de leur dispersion ?

— Quelle représentation graphique permet de comprendre rapidement les données ?

La statistique descriptive se distingue de la statistique inférentielle : elle décrit les données disponibles, tandis que la seconde cherche notamment à tirer des conclusions sur une population à partir d'un échantillon.

Problématique de la leçon : comment résumer une série statistique à une variable sans perdre les informations essentielles sur sa position, sa dispersion et sa forme ? “‘latex

2 Prérequis et objectifs

2.1 Prérequis

La compréhension de cette leçon nécessite les connaissances et compétences suivantes :

— maîtriser les calculs numériques usuels : fractions, pourcentages, sommes et moyennes ;

— savoir lire, organiser et exploiter un tableau de données ;

— connaître les notions de proportionnalité, de fréquence et de pourcentage ;

— savoir ordonner des nombres réels et comparer des valeurs ;

— maîtriser les calculs de carrés et de racines carrées ;

— savoir lire un graphique et interpréter des informations numériques.

2.2 Objectifs

Cette leçon vise à présenter les outils fondamentaux de la statistique descriptive à une variable et à en préciser les conditions d'utilisation.

À l'issue de l'étude, on doit être capable de :

— définir une population statistique, un individu, un caractère, une modalité et un effectif ;

— distinguer les caractères qualitatifs et quantitatifs, discrets et continus ;

— construire et interpréter un tableau d'effectifs, de fréquences et d'effectifs cumulés ;

— choisir une représentation graphique adaptée à la nature des données ;

— calculer et interpréter les indicateurs de position : mode, moyenne, médiane et quartiles ;

— calculer et interpréter les indicateurs de dispersion : étendue, écart interquartile, variance et écart-type ;

— comparer plusieurs séries statistiques en mobilisant conjointement des indicateurs de position et de dispersion ;

— utiliser une calculatrice ou un tableur pour traiter une série statistique et contrôler la cohérence des résultats ;

— exercer un regard critique sur les conclusions tirées d'un résumé statistique.

2.3 Objectifs didactiques

Sur le plan didactique, il s'agit de faire comprendre que les indicateurs statistiques sont des outils de synthèse dont le choix dépend de la nature des données et de la question posée.

L'élève doit notamment comprendre que la moyenne peut être sensible aux valeurs extrêmes, que la médiane décrit la position centrale d'une distribution et que plusieurs indicateurs sont souvent nécessaires pour en rendre compte.

L'enjeu est de passer du calcul mécanique à l'interprétation des résultats dans leur contexte, en développant la capacité à organiser, représenter, comparer et analyser des données. "'

3 Vocabulaire et organisation des données

3.1 Population, individu et caractère

On appelle :

— **population statistique** l'ensemble des individus étudiés ;

— **individu** un élément de cette population ;

— **caractère statistique** la propriété observée sur chaque individu ;

— **modalité** une valeur ou une catégorie prise par le caractère ;

— **effectif total** le nombre d'individus étudiés, noté N.

On distingue trois grands types de caractères :

— qualitatif nominal : couleur des yeux, catégorie socioprofessionnelle ;

— qualitatif ordinal : niveau de satisfaction, classement ;

— quantitatif : âge, masse, durée, nombre d'objets.

Un caractère quantitatif est discret lorsque ses valeurs sont isolées, comme un nombre d'enfants. Il est souvent dit continu lorsqu'il peut prendre toutes les valeurs d'un intervalle, comme une durée mesurée.

3.2 Série statistique et effectifs

Considérons une série quantitative dont les valeurs distinctes sont

$$x_1 < x_2 < \cdots < x_p.$$

On note n_i l'effectif associé à la valeur x_i. L'effectif total est

$$N = \sum_{i=1}^{p} n_i.$$

La fréquence associée à x_i est

$$f_i = \frac{n_i}{N}.$$

On a les propriétés

$$0 \leq f_i \leq 1, \qquad \sum_{i=1}^{p} f_i = 1.$$

Les fréquences peuvent être exprimées en pourcentage en multipliant leur valeur par 100.

3.3 Effectifs et fréquences cumulés

Pour une série ordonnée, l'effectif cumulé croissant jusqu'à la valeur x_i est

$$N_i = \sum_{j=1}^{i} n_j.$$

La fréquence cumulée croissante est

$$F_i = \sum_{j=1}^{i} f_j = \frac{N_i}{N}.$$

Ces quantités permettent de déterminer les rangs et de répondre à des questions telles que : « Quelle proportion des observations est inférieure ou égale à une valeur donnée ? »

4 Représentations graphiques

Le choix d'une représentation dépend de la nature du caractère et de l'objectif poursuivi.

4.1 Diagramme en bâtons

Pour un caractère quantitatif discret, on représente les différentes valeurs par des points ou des bâtons. La hauteur des bâtons est proportionnelle aux effectifs ou aux fréquences.

4.2 Diagramme circulaire

Pour un caractère qualitatif ou une variable discrète comportant peu de modalités, un diagramme circulaire peut représenter les proportions.

L'angle correspondant à une modalité de fréquence f_i est égal à

$$\alpha_i = 360^\circ f_i.$$

4.3 Histogramme

Pour une variable quantitative regroupée en classes, on utilise souvent un histogramme.

Si les classes ont des amplitudes différentes, ce sont les **aires** des rectangles qui doivent être proportionnelles aux effectifs ou aux fréquences. La hauteur représente alors une densité d'effectif ou une densité de fréquence.

Pour une classe $[a_i, b_i[$ d'effectif n_i, la densité d'effectif est

$$d_i = \frac{n_i}{b_i - a_i}.$$

Il faut éviter de confondre hauteur et effectif lorsque les amplitudes des classes diffèrent.

4.4 Diagramme en boîte

Le diagramme en boîte, ou boîte à moustaches, représente généralement :

— le minimum ;

— le premier quartile Q_1 ;

— la médiane ;

— le troisième quartile Q_3 ;

— le maximum.

La boîte s'étend de Q_1 à Q_3 et contient au moins la moitié centrale des observations selon les conventions usuelles de détermination des quartiles.

Ce graphique facilite la comparaison de plusieurs distributions, notamment leur position et leur dispersion.

5 Indicateurs de position

5.1 Mode

Le mode est une valeur dont l'effectif est maximal. Une série peut avoir plusieurs modes.

Le mode est utilisable pour des variables qualitatives comme quantitatives. Pour des données continues regroupées en classes, on parle plutôt de classe modale.

5.2 Moyenne arithmétique

Pour les observations $x_1, \ldots, x_N$, la moyenne arithmétique est

$$\overline{x} = \frac{1}{N} \sum_{j=1}^{N} x_j.$$

Lorsque les valeurs distinctes x_i sont associées aux effectifs n_i, on obtient

$$\boxed{\overline{x} = \frac{1}{N} \sum_{i=1}^{p} n_i x_i = \sum_{i=1}^{p} f_i x_i.}$$

La moyenne est un indicateur de position qui utilise toutes les observations. Elle est sensible aux valeurs extrêmes.

Propriétés

Pour deux constantes a et b, si $y_j = ax_j + b$, alors

$$\overline{y} = a\overline{x} + b.$$

La somme des écarts à la moyenne est nulle :

$$\sum_{j=1}^{N} (x_j - \overline{x}) = 0.$$

La moyenne peut être interprétée comme le centre de gravité des observations lorsque chaque observation porte la même masse.

5.3 Médiane

La médiane, notée M_e, est une valeur qui partage la série ordonnée en deux groupes : au moins la moitié des observations sont inférieures ou égales à la médiane, et au moins la moitié lui sont supérieures ou égales.

Pour une série ordonnée

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(N)},$$

on peut choisir :

$$M_e = \begin{cases} x_{((N+1)/2)}, & \text{si } N \text{ est impair,} \\ \dfrac{x_{(N/2)} + x_{(N/2+1)}}{2}, & \text{si } N \text{ est pair.} \end{cases}$$

La seconde convention est couramment utilisée au lycée. Plus généralement, la définition par partage de la distribution admet parfois plusieurs médianes.

La médiane est moins sensible que la moyenne aux valeurs extrêmes. Elle est particulièrement pertinente pour décrire des distributions asymétriques, par exemple des revenus ou des durées d'attente.

5.4 Quartiles et quantiles

Le premier quartile Q_1 est défini comme la plus petite valeur observée telle qu'au moins 25 % des observations lui soient inférieures ou égales.

Le troisième quartile Q_3 est la plus petite valeur observée telle qu'au moins 75 % des observations lui soient inférieures ou égales.

Dans le cas d'une série de N observations ordonnées, une convention courante au lycée consiste à prendre :

$$Q_1 = x_{(\lceil N/4 \rceil)}, \qquad Q_3 = x_{(\lceil 3N/4 \rceil)},$$

où $\lceil u \rceil$ désigne le plus petit entier supérieur ou égal à u.

Les conventions de calcul des quartiles peuvent différer selon les logiciels et les ouvrages. Il convient donc d'annoncer la convention retenue.

Plus généralement, un quantile d'ordre q, avec $0 < q < 1$, est une valeur permettant de situer une proportion q des observations en dessous ou à cette valeur.

6 Indicateurs de dispersion

6.1 Étendue

L'étendue est la différence entre la valeur maximale et la valeur minimale :

$$\boxed{E = x_{\max} - x_{\min}.}$$

Elle est simple à calculer mais ne tient compte que des deux valeurs extrêmes.

6.2 Écart interquartile

L'écart interquartile est défini par

$$\boxed{IQR = Q_3 - Q_1.}$$

Il mesure la largeur de l'intervalle contenant la moitié centrale des observations selon la convention retenue.

Contrairement à l'étendue, il est peu sensible aux valeurs extrêmes isolées.

6.3 Variance

La variance mesure la dispersion autour de la moyenne. Elle est définie par

$$\boxed{V = \frac{1}{N} \sum_{j=1}^{N} (x_j - \overline{x})^2.}$$

Pour une série donnée par valeurs et effectifs :

$$V = \frac{1}{N} \sum_{i=1}^{p} n_i (x_i - \overline{x})^2.$$

En développant le carré, on obtient la formule de calcul :

$$\boxed{V = \frac{1}{N}\sum_{j=1}^{N} x_j^2 - \overline{x}^{\,2}.}$$

La variance est toujours positive ou nulle. Elle est nulle si et seulement si toutes les observations sont égales.

6.4 Écart-type

L'écart-type est la racine carrée de la variance :

$$\boxed{\sigma = \sqrt{V}.}$$

Il s'exprime dans la même unité que les observations, contrairement à la variance.

Un écart-type faible indique que les observations sont relativement regroupées autour de la moyenne. Un écart-type élevé traduit une dispersion plus importante, sans toutefois suffire à décrire toute la forme de la distribution.

6.5 Propriétés de la variance

Pour $Y = aX + b$, au sens des transformations appliquées aux données,

$$V(Y) = a^2 V(X), \qquad \sigma(Y) = |a|\sigma(X).$$

L'ajout d'une constante ne modifie donc pas la dispersion, tandis qu'une multiplication par a multiplie l'écart-type par $|a|$.

La variance est particulièrement sensible aux valeurs extrêmes, en raison de l'élévation au carré des écarts.

7 Exemple détaillé

On étudie le nombre de livres lus au cours d'un mois par dix élèves. Les observations sont :

$$2,\ 4,\ 3,\ 5,\ 4,\ 1,\ 3,\ 4,\ 6,\ 3.$$

7.1 Tableau statistique

Après classement des valeurs, on obtient :

$$1,\ 2,\ 3,\ 3,\ 3,\ 4,\ 4,\ 4,\ 5,\ 6.$$

Valeur x_i	1	2	3	4	5	6
Effectif n_i	1	1	3	3	1	1
Fréquence f_i	0,1	0,1	0,3	0,3	0,1	0,1
Effectif cumulé	1	2	5	8	9	10

L'effectif total est $N = 10$.

7.2 Moyenne

La somme des observations vaut

$$2 + 4 + 3 + 5 + 4 + 1 + 3 + 4 + 6 + 3 = 35.$$

Ainsi,

$$\boxed{\overline{x} = \frac{35}{10} = 3,5.}$$

En moyenne, les élèves ont lu $3,5$ livres au cours du mois étudié.

7.3 Médiane et quartiles

Les observations centrales sont celles de rangs 5 et 6, soit respectivement 3 et 4. La médiane vaut donc

$$\boxed{M_e = \frac{3+4}{2} = 3,5.}$$

Avec la convention de rang annoncée précédemment :

$$Q_1 = x_{(\lceil 10/4 \rceil)} = x_{(3)} = 3,$$

et

$$Q_3 = x_{(\lceil 30/4 \rceil)} = x_{(8)} = 4.$$

Par conséquent,

$$\boxed{IQR = 4 - 3 = 1.}$$

Au moins un quart des élèves ont lu au plus 3 livres, et au moins trois quarts ont lu au plus 4 livres.

7.4 Étendue

Le minimum est 1 et le maximum est 6. Ainsi,

$$\boxed{E = 6 - 1 = 5.}$$

7.5 Variance et écart-type

La moyenne des carrés des observations est

$$\frac{1^2 + 2^2 + 3^2 + 3^2 + 3^2 + 4^2 + 4^2 + 4^2 + 5^2 + 6^2}{10} = \frac{145}{10} = 14,5.$$

On en déduit

$$V = 14,5 - 3,5^2 = 14,5 - 12,25 = 2,25.$$

L'écart-type est donc

$$\boxed{\sigma = \sqrt{2,25} = 1,5.}$$

7.6 Bilan de l'exemple

La moyenne et la médiane sont égales à $3,5$ livres. La moitié centrale des observations est comprise entre 3 et 4 livres, tandis que l'étendue vaut 5 livres.

Cet exemple illustre la complémentarité des indicateurs : la moyenne et la médiane décrivent la position, alors que l'étendue, l'écart interquartile et l'écart-type renseignent sur la dispersion.

8 Comparaison de distributions

Deux séries peuvent avoir la même moyenne et des dispersions très différentes.
Considérons :

$$A = (4, 5, 5, 6), \qquad B = (1, 5, 5, 9).$$

Les deux moyennes valent

$$\overline{x}_A = \overline{x}_B = 5.$$

Pour la série A :

$$V_A = \frac{(4-5)^2 + (5-5)^2 + (5-5)^2 + (6-5)^2}{4} = \frac{1}{2}.$$

Pour la série B :

$$V_B = \frac{(1-5)^2 + (5-5)^2 + (5-5)^2 + (9-5)^2}{4} = 8.$$

Ainsi, les deux séries ont la même moyenne, mais la série B est beaucoup plus dispersée.

Cet exemple montre qu'un indicateur isolé ne suffit pas toujours pour décrire une distribution. Il est souvent nécessaire d'associer un indicateur de position et un indicateur de dispersion, voire une représentation graphique.

9 Mise en œuvre avec un tableur

Un tableur ou une calculatrice permet de traiter rapidement des séries de grande taille.
Les élèves peuvent :

— saisir et trier les données ;

— calculer les effectifs et les fréquences ;

— construire un diagramme en bâtons ou un histogramme ;

— déterminer la moyenne et l'écart-type ;

— comparer plusieurs séries ;

— vérifier les calculs réalisés à la main.

Il faut cependant distinguer la compréhension mathématique de la simple utilisation d'un outil. L'élève doit pouvoir expliquer la signification des indicateurs obtenus et vérifier que le graphique choisi est adapté aux données.

On veillera également à identifier la convention utilisée par le logiciel pour les quartiles et l'écart-type.

10 Aspects didactiques et pédagogiques

10.1 Objectifs d'apprentissage

À l'issue de la séquence, les élèves doivent savoir :

— identifier la population, les individus et le caractère étudié ;

— organiser une série dans un tableau d'effectifs et de fréquences ;

— calculer et interpréter la moyenne, la médiane, les quartiles et les indicateurs de dispersion ;

— choisir et lire une représentation graphique ;

— comparer deux distributions en justifiant les indicateurs utilisés ;

— utiliser un logiciel de calcul statistique en contrôlant la cohérence des résultats.

10.2 Difficultés fréquentes

Plusieurs erreurs doivent être anticipées :

— calculer la médiane sans ordonner les données ;

— confondre moyenne et médiane ;

— considérer les quartiles comme nécessairement uniques quelle que soit la convention de calcul ;

— confondre variance et écart-type ;

— représenter des classes d'amplitudes différentes par des rectangles dont les hauteurs sont les effectifs ;

— interpréter une moyenne comme une valeur nécessairement observée dans la série.

Une autre difficulté consiste à interpréter les indicateurs dans leur contexte. Une moyenne seule peut masquer une distribution très dissymétrique ou la présence de valeurs extrêmes.

10.3 Proposition de démarche pédagogique

Une séance peut s'organiser en quatre étapes :

1. **Situation-problème :** recueillir les temps de trajet des élèves et demander comment résumer les résultats.
2. **Organisation :** construire un tableau d'effectifs et de fréquences, puis une représentation graphique adaptée.
3. **Calcul et interprétation :** déterminer moyenne, médiane et indicateurs de dispersion.
4. **Synthèse :** comparer deux groupes et discuter les limites d'un résumé statistique.

Les données réelles permettent de donner du sens aux calculs et de développer l'esprit critique face aux graphiques et aux résumés statistiques.

11 Ouverture et limites

La statistique descriptive à une variable constitue une base pour des études plus approfondies :

— statistique descriptive à deux variables ;

— étude d'une liaison entre deux caractères ;

— régression et ajustement affine ;

— probabilités et lois de probabilité ;

— statistique inférentielle et estimation.

Il convient de rappeler que les indicateurs résument une distribution sans la décrire complètement. Deux séries différentes peuvent partager la même moyenne, la même médiane et le même écart-type. L'analyse graphique et le contexte de collecte restent donc essentiels.

12 Conclusion

La statistique descriptive à une variable vise à résumer et à représenter une série de données pour en faciliter l'interprétation.

Les indicateurs de position, comme la moyenne et la médiane, permettent de situer la distribution. Les indicateurs de dispersion, comme l'étendue, l'écart interquartile et l'écart-type, mesurent des aspects complémentaires de l'étalement des observations.

Le choix des outils dépend de la nature des données, de la question étudiée et de la sensibilité des indicateurs aux valeurs extrêmes. L'objectif pédagogique ne se limite donc pas à appliquer des formules : il consiste à apprendre à choisir, interpréter et confronter plusieurs résumés statistiques.

Références

[1] INSEE, *Moyenne*, Définitions, Institut national de la statistique et des études économiques. https://www.insee.fr/fr/metadonnees/definition/c1970

[2] INSEE, *Médiane*, Définitions, Institut national de la statistique et des études économiques. https://www.insee.fr/fr/metadonnees/definition/c1376

[3] INSEE, *Quartiles*, Définitions, Institut national de la statistique et des études économiques. https://www.insee.fr/fr/metadonnees/definition/c1844

[4] Ministère de l'Éducation nationale, *Programme d'enseignement de mathématiques de la classe de seconde générale et technologique*, Bulletin officiel de l'Éducation nationale, 2 avril 2026. https://www.education.gouv.fr/bo/2026/Hebdo14/MENE2602914A

[5] Université de Rennes, *Introduction à la statistique descriptive*, support de cours. https://capes-math.univ-rennes1.fr/cours-pdf/StatDesc.pdf