Information chiffrée et statistiques

2de

I Proportions

Définition On considère un ensemble E (appelé également population) contenant une partie A (ou sous-population). On note n_E et n_A respectivement le nombre d'individus dans l'ensemble E et dans l'ensemble A. La proportion de A dans E est le nombre $p = \frac{n_A}{n_E}$.

Remarque
On peut l'écrire sous forme décimale, fractionnaire ou en pourcentage.

Exemples
1) Dans une classe de seconde de 32 élèves, 18 élèves ont choisi la spécialité mathématiques en fin d'année.
La proportion des élèves ayant choisi l'option mathématiques dans la classe est :

$$a = \frac{18}{32} = \frac{9}{16} = 0{,}5625 \text{ soit } a = \frac{56{,}25}{100}$$

C'est-à-dire 56,25% des élèves de la classe ont choisi l'option mathématiques.

2) Parmi les 480 élèves de seconde, 15 % ont choisi l'option grec ou latin.

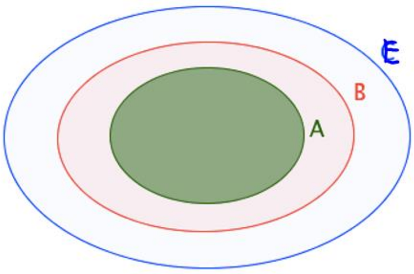
15 % de 480 ont choisi l'option grec ou latin, soit :

15 % x 480 = $\frac{15}{100}$ x 480 = 72 élèves.
Pensez au double de 5 : 2×5 ; au triple de 5 : 3×5.

Propriété Soit E un ensemble et A, B des parties de E tels que $A \subset B \subset E$.
Soit p_1 la proportion de A dans B et soit p_2 la proportion de B dans E.
Alors, $p = p_1 \times p_2$ est la proportion de A dans E.

Démonstration en exercice

Figure



Exemple
Sur 67 millions d'habitants en France, 66 % de la population est en âge de travailler (15-64 ans). La population active représente 70 % de la population en âge de travailler.
Soit F est la population française, T la population en âge de travailler et A la population active.
La proportion de A dans T est 70 % et la proportion de T dans F est 66 %.
Donc la proportion de A dans F est égale à :70 % x 66 % = 0,7 x 0,66 = 0,462 = 46,2 soit 46,2 % des français sont actifs.
46,2 % de 67 = 0,462 x 67 = 30,954. La France compte environ 31 millions d'actifs

II Lien entre une évolution et un pourcentage

Une grandeur numérique positive (un prix, un nombre d'habitants, un salaire, ….) évolue au cours du temps.

Définition

Remarque
Le numérateur $V_f - V_i$ est appelé variation absolue.

Définition

Remarques
1) Si t>0, alors c'est un pourcentage de hausse.
2) Si t<0, alors c'est un pourcentage de baisse.

Exemple

Remarque *Signe d'un taux*
Dans les situations concrètes, l'usage veut qu'un taux puisse être positif même si l'évolution est une réduction.
On peut ainsi dire :
"des soldes de -50%" ou "des soldes de 50%". "une baisse de -3% ou "une baisse de 3%". "un taux de réduction de -0,6" ou "un taux de réduction de 0,6".
Cela ne pose pas de problème dans la mesure ou on sait que l'évolution est une réduction.

Propriété et définition

Démonstration en exercice

On peut retenir cette propriété de la façon suivante :
- Augmenter une valeur de *p* % revient à la multiplier par
- Diminuer une valeur de *p* % revient à la multiplier par

Remarques
1) Si le coefficient multiplicateur est inférieur à 1, alors c'est une baisse.
2) Si le coefficient multiplicateur est supérieur à 1, alors c'est une hausse.

Exemples

III Evolutions

3.1 Evolutions successives

Attention, **les pourcentages d'évolutions successives ne s'additionnent pas**. En particulier une hausse de *t* % n'est pas compensée par une baisse de *t* %.

En effet, par exemple : Le prix initial d'un article est de 250 €. Le prix de cet article a successivement augmenté puis diminué de 12%.

Propriété

Démonstration en exercice

Exemple

3.2 Evolution réciproque

Propriété

Démonstration en exercice

Exemple

IV Statistiques

Dans un monde de plus en plus quantifié, les statistiques constituent un enjeu essentiel pour la formation du citoyen, en lui fournissant des outils pour maîtriser l'information chiffrée et ainsi comprendre et participer au débat public. Par ailleurs, les outils de la statistique sont fréquemment utilisés par de nombreuses disciplines, tant scientifiques qu'économiques ou sociales. Le rôle de la statistique est de présenter une masse de données sous forme lisible. Puis, si possible, de la résumer par quelques nombres caractéristiques (moyenne, médiane, quartiles…).

4.1 Ecart interquartile

On rappelle que la médiane M_e d'une série statistique permet de partager en deux une série rangée dans l'ordre croissant.

Méthode pour déterminer la médiane d'une série statistique
Dans un 1^er^ temps, on classe la série par ordre croissant.

- Si la série comporte un nombre N impair de termes, la médiane est alors le $\frac{N+1}{2}$ème terme.
- Si la série comporte un nombre N pair de termes, la médiane est la moyenne des termes situés au $\frac{N}{2}$ème rang et $\frac{N}{2}+1$ème rang (car entre le $\frac{N}{2}$ème et $\frac{N}{2}+1$ème, il n'y a pas de terme !).

Exemples

1) Soit la série statistique suivante :

3	4	4	5	7	9	11	13	15	16	18

N=11 est impair donc la médiane est située au $\frac{N+1}{2}$ème rang de la série soit le 6ème terme. Donc $Me = 9$.

2) Soit la série statistique suivante :

2	5	7	8	8	12	12	15	15	16

N=10 est pair la médiane est la moyenne des termes situés au $\frac{N}{2}$ème rang et $\frac{N}{2}+1$ème rang soit les 5ème et 6ème termes.

$Me = \frac{8+12}{2} = 10$.

Pour partager la série statistique en quatre, on utilise les quartiles.

Définitions
1) Le 1er quartile Q_1 est la plus petite valeur de la série telle qu'au moins 25 % des données soient inférieures ou égales à Q_1.
2) Le 3ème quartile est la plus petite valeur de la série telle qu'au moins 75 % des données soient inférieures ou égales Q_3.

Méthode pour déterminer les quartiles
Dans un 1er temps, on classe la série par ordre croissant.

- Pour déterminer le 1er quartile, on calcule $\frac{N}{4}$. Si $\frac{N}{4} \in \mathbb{N}$, alors Q_1 est le $\frac{N}{4}^{\text{ème}}$ terme de la série sinon on tronque le nombre décimal à l'entier supérieur (2,25-> 3).
- Pour déterminer le 3ème quartile, on calcule $\frac{3N}{4}$ et on procède de la même façon que précédemment.

Exemple
Soit la série statistique suivante :

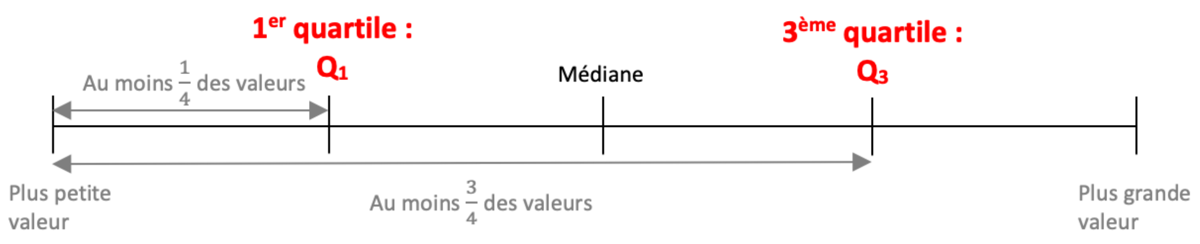
2	5	7	8	8	12	12	15	15	16

$\frac{N}{4} = \frac{10}{4} = \frac{5}{2} =$ 2,5-> on prend la 3ème valeur, donc $Q_1 = 7$. $\frac{3N}{4} = \frac{15}{2} =$ 7,5-> on prend la 8ème valeur, donc $Q_3 = 15$.

Pour estimer la dispersion des valeurs d'une série statistique on peut utiliser son étendue ou, mieux, son écart interquartile.

Définitions
1) L'étendu d'une série statistique, notée E, est la différence entre la plus grande et la plus petite des valeurs de cette série : $E = Max - Min$.
2) L'intervalle intervalle interquartile d'une série statistique est l'intervalle $I_Q = [Q_1\,;Q_3]$.
3) L'écart interquartile d'une série statistique, notée E_Q, est tel que $\boldsymbol{E_Q = Q_3 - Q_1}$.

Remarque
L'intervalle interquartile contient 50% des valeurs (centrales) de la série. L'écart interquartile n'est pas sensible aux valeurs extrêmes.



Exemple
Soit la série statistique précédente.
Me=10, Q1=7 et Q3=15. Donc $I_Q = [7; 15]$ et $E_Q = 15 - 7 = 8$.

4.2 Moyenne, variance et écart-type

Dans ce paragraphe, on va mettre en place d'autres paramètres qui vont nous permettre de dire si les valeurs de la série sont regroupées ou au contraire dispersées autour de sa moyenne et entre elles.
On considère dans la suite, la série statistique définie par le tableau ci-dessous.

Valeur	x_1	x_2	…	x_p
Effectif	n_1	n_2	…	n_p

L'effectif total est $N = n_1 + n_2 + \cdots + n_p$.

Définition *Moyenne*
La moyenne de la série statistique est le nombre réel, noté $\bar{x}$, défini par :
$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \cdots + x_p n_p}{N}$$

Remarque La moyenne est un paramètre de tendance centrale.

Exemple Soit la série suivante :

2	5	7	8	8	12	12	15	15	16

On a donc $\bar{x} = \frac{2+5+7+2\times 8+2\times 12+2\times 15+16}{10} = 10$.

Propriétés *Linéarité de la moyenne*
(i) Si l'on multiplie toutes les valeurs d'une série par une même constante a, sans en changer les effectifs, alors la moyenne est multipliée par cette constante a.
(ii) Si on ajoute une même constante b à toutes les valeurs d'une série statistique, sans en changer les effectifs, alors la moyenne est augmentée de cette valeur b.

Démonstrations
(i) On a : $\bar{y} = \frac{an_1x_1 + an_2\, x_2 + \cdots an_p x_p}{N} = a\bar{x}$.

(ii) On a : $\bar{y} = \frac{n_1(x_1+b) + an_2(\, x_2+b) + \cdots an_p(x_p+b)}{N} = \bar{x} + b$.

Exemple
Soit la série précédente. Si l'on décide d'ajouter 1 à chaque valeur, alors la nouvelle moyenne est de $\bar{y} = 11$.

Définition *Variance*
La variance de la série statistique est le nombre réel, noté V, défini par :
$$V = \frac{n_1 \times (x_1 - \bar{x})^2 + n_2 \times (x_2 - \bar{x})^2 + \cdots + n_p \times (x_p - \bar{x})^2}{N}$$

Exemple
Si on reprend la série précédente,
$V = \frac{(2-10)^2+(5-10)^2+(7-10)^2+2\times(8-10)^2+2\times(12-10)^2+2\times(15-10)^2+(16-10)^2}{10} = 18$.

Définition *Ecart-type*
L'écart-type de la série statistique est le nombre réel, noté σ, défini par :

$$\boldsymbol{\sigma = \sqrt{V}}$$

Interprétation
La variance est la moyenne des carrés des écarts par rapport à la moyenne. Elle mesure donc la dispersion autour de la moyenne. Elle n'est pas très parlante car elle s'exprime dans le carré de l'unité du caractère.
L'écart-type a l'avantage de s'exprimer dans la même unité que le caractère. Il permet de comparer la dispersion de la série de valeurs (par rapport à la moyenne) et l'intérêt réside dans la comparaison de valeurs. Contrairement à l'écart interquartile, il tient compte de l'ensemble de la population (les valeurs extrêmes influencent l'écart-type).

Exemple
Voici les notes de mathématiques d'Anna et Brahim obtenues lors des 4 derniers DS :
Anna : 8 – 9 – 11 – 9
Brahim : 1 - 14 – 3 – 19
Comparons ces deux séries de notes :

- Anna : $\overline{x_1} = 9{,}25$; $V_1 = 1{,}19$, $\sigma_1 = 1{,}09$
- Brahim : $\overline{x_2} = 9{,}25$; $V_2 = 56{,}19$, $\sigma_2 = 7{,}5$

On a : $\overline{x_1} = \overline{x_2}$ mais $\sigma_1 < \sigma_2$.

Les notes de Brahim sont plus dispersées par rapport à leur moyenne.