Statistiques

Denis Vekemans [*]

1 Introduction

Les statistiques proposent différents indicateurs qui permettent de résumer, ou de rendre apparentes certaines propriétés d'une population que l'on veut étudier.

Population : ensemble de ce qui est étudié, cela peut être des personnes ou des choses.

2 Série statistique discrète

Les valeurs du caractère sont discrètes.

Les éléments qui composent la population sont appelés des individus. L'effectif total est le nombre d'individus de la population. Ce qui est étudié de ces individus est appelé un caractère. Un caractère peut être qualitatif ou quantitatif. Une série statistique est l'ensemble des valeurs d'un caractère pour chacun des individus d'une population donnée. Si le caractère étudié est quantitatif, ce sera une série de nombres.

Exemple de série statistique.

On a demandé aux 28 élèves d'une classe de CM1 leur nombre de frères et soeurs. Voici les réponses obtenues :

2 ; 1 ; 0 ; 1 ; 4 ; 0 ; 3 ; 2 ; 0 ; 1 ; 2 ; 5 ; 1 ; 2 ; 2 ; 3 ; 0 ; 0 ; 1 ; 3 ; 1 ; 2 ; 4 ; 3 ; 0 ; 4 ; 1 ; 2.

Population : les élèves de la classe de CM1.

Caractère : le nombre de frères et soeurs.

Valeurs du caractère : 0, 1, 2, 3, 4, 5.

Effectif total (le nombre d'élèves de la classe) : 28.

Effectif et fréquence.

L'effectif d'une valeur est le nombre d'individus de la population a qui est attribué cette valeur.

La fréquence d'une valeur est le quotient de l'effectif de cette valeur par l'effectif total.

Reprenons l'exemple précédent. On peut présenter les résultats à l'aide d'un tableau :

Nombre frères et soeurs	0	1	2	3	4	5
Effectif	6	7	7	4	3	1

La fréquence de la valeur 2 est donc $\frac{1}{4}$ ou $0,25$.

[*]Laboratoire de mathématiques pures et appliquées Joseph Liouville ; 50, rue Ferdinand Buisson BP 699 ; 62 228 Calais cedex ; France

La fréquence d'une valeur correspond à la probabilité de cette valeur.

La moyenne.

La moyenne d'une série statistique est le quotient de la somme de toutes les valeurs pondérées par de cette série par l'effectif total. Exemple de série statistique

Nombre frères et soeurs x_i	0	1	2	3	4	5
Effectif n_i	6	7	7	4	3	1
Fréquence f_i	$\frac{6}{28}$	$\frac{7}{28}$	$\frac{7}{28}$	$\frac{4}{28}$	$\frac{3}{28}$	$\frac{1}{28}$

le nombre de valeurs du caractère est $p = 6$.

Ici, l'effectif total est $n = \sum_{k=1}^{p} n_i$: $n = 6$.

La moyenne m est $m = \dfrac{\sum_{k=1}^{p} n_i \times x_i}{n} = \displaystyle\sum_{k=1}^{p} f_i \times x_i$: $m = \dfrac{\frac{6}{28}\times 0 + \frac{7}{28}\times 1 + \frac{7}{28}\times 2 + \frac{4}{28}\times 3 + \frac{3}{28}\times 4 + \frac{1}{28}\times 5}{28} = \frac{56}{28} = 2$.

Médiane.

Une valeur médiane $Mé$ partage les valeurs d'une série statistique en deux groupes de manière à ce que

— 50% (ou plus) des valeurs sont inférieures ou égale à $Mé$;

— 50% (ou plus) des valeurs sont supérieures ou égale à $Mé$.

Remarque : dans le cas de médianes multiples, il est d'usage de choisir le moyenne de toutes les médianes.

Dans l'exemple précédent, la seule médiane est 2 car il y a 20 enfants qui ont 2 frères/soeurs ou moins et 15 enfants qui ont 2 frères/soeurs ou plus.

L'étendue.

L'étendue d'une série statistique est la différence entre la valeur la plus grande et la valeur la plus petite.

Dans l'exemple précédent, l'étendue est $x_6 - x_5 = 5 - 0 = 5$.

Quartiles.

Un premier quartile Q_1 partage les valeurs d'une série statistique en deux groupes de manière à ce que

— 25% (ou plus) des valeurs sont inférieures ou égale à Q_1 ;

— 75% (ou plus) des valeurs sont supérieures ou égale à Q_1.

Un troisième quartile Q_3 partage les valeurs d'une série statistique en deux groupes de manière à ce que

— 75% (ou plus) des valeurs sont inférieures ou égale à Q_3 ;

— 25% (ou plus) des valeurs sont supérieures ou égale à Q_3.

Remarque : dans le cas de quartiles multiples, il est d'usage de choisir le moyenne de tous les quartiles.

Lorsque Q_1 et Q_3 sont déterminés, la différence $Q_3 - Q_1$ s'appelle l'écart inter-quartile.

Dans l'exemple précédent,

— le seul premier quartile est 1 car il y a 13 enfants qui ont 1 frères/soeurs ou moins et 22 enfants qui ont 1 frères/soeurs ou plus ;

— le seul troisième quartile est 3 car il y a 24 enfants qui ont 3 frères/soeurs ou moins et 8 enfants qui ont 3 frères/soeurs ou plus ;

— l'écart inter-quartile est $3 - 1 = 2$.

3 Série statistique classée

Les valeurs du caractère sont classées (des intervalles).

L'effectif d'une classe est le nombre de valeurs comprises dans cette classe. La fréquence d'une classe est le quotient de l'effectif de cette classe par l'effectif total. L'amplitude d'une classe est la différence entre la plus grande valeur et la plus petite valeur de cette classe.

Exercice 1 Dans le but d'étudier la loi de survie d'un certain type de matériel, une entreprise s'est livrée sur 600 machines identiques à des observations résumées dans le tableau suivant :

Année de mise en réforme	0	1	2	3	4	5	6	7	8	9	10	11	12
Nombre de machines encore en service à la fin de l'année	600	592	564	508	391	267	155	87	34	17	7	2	0

— Déterminer la durée médiane d'existence du type de matériel utilisé, en justifiant votre résultat. Donner une interprétation du résultat obtenu.

— Déterminer sa durée moyenne d'existence. On admettra que les mises à la réforme s'effectuent de façon uniforme dans le courant de l'année. On arrondira le résultat à 10^{-2} près.

Durée de vie des machines en années	$[0,1]$	$]1,2]$	$]2,3]$	$]3,4]$	$]4,5]$	$]5,6]$	$]6,7]$	$]7,8]$	$]8,9]$	$]9,10]$	$]10,11]$	$]11,12]$
Effectif des machines	8	28	56	117	124	112	68	53	17	10	5	2
Effectifs cumulé des machines	8	36	92	209	333	445	513	566	583	593	598	600

1. La moitié de l'effectif est 300, la médiane est donc une valeur comprise dans l'intervalle $]4,5[$.

   Le calcul exact de cette valeur s'effectue par interpolation linéaire en cherchant l'abscisse du point d'ordonnée 300 de la droite qui passe par les points de coordonnées $(4;209)$ et $(5;333)$. Cette droite représente graphiquement la fonction linéaire $f:\mathbb{R}\to\mathbb{R}; x\mapsto a\times x+b$ vérifiant $f(4)=209$ et $f(5)=333$, ce qui donne $f(x)=124\times x-287$.

   $f(x)=300$ donne $x=\frac{587}{124}$. La médiane est donc de $\frac{587}{124}$ années, ce qui est proche de $4,7$ années.
2. Le calcul de la moyenne m s'effectue en utilisant les centres des classes

$$m=\frac{8\times 0,5+28\times 1,5+56\times 2,5+117\times 3,5+124\times 4,5+112\times 5,5+68\times 6,5+53\times 7,5+17\times 8,5+10\times 9,5+5\times 10,5+2\times 11,5}{600}=\frac{2924}{600}.$$

   $m=4,87$ arrondi à 10^{-2} près.

4 Exercices non corrigés

Exercice 2 Dans deux classes de 24 élèves chacune, on demande aux collégiens qui utilisent tous l'autobus, combien de temps ils passent dans ce moyen de transport pour se rendre à leur collège.

(a) Reproduire et compléter la première colonne du tableau suivant qui représente les résultats de cette enquête, en sachant que tous les élèves ont donné une réponse.

Temps en minutes	Effectif	Fréquences f_i
$0 < t < 15$	6	
$15 \leq t < 30$	24	
$30 \leq t < 45$		
$45 \leq t < 60$	3	

(b) Quel est l'effectif d'élèves passant au moins 30 minutes dans l'autobus pour se rendre au collège ?

(c) Déterminer les valeurs maximales et minimales de la variable étudiée.

(d) Compléter la colonne des fréquences correspondant à cette étude statistique.

Exercice 3 Dans une entreprise, les salaires, en euros, se répartissent de la façon suivante :

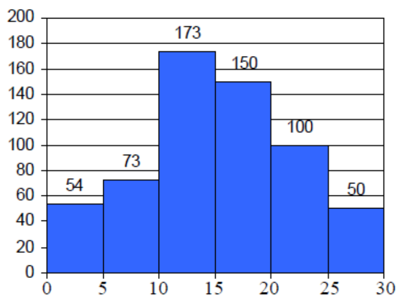
Classes	Effectifs	Classes	Effectifs
$[1000; 1200[$	12	$[1600; 1800[$	18
$[1200; 1400[$	20	$[1800; 2000[$	6
$[1400; 1600[$	40	$[2000; 2200[$	4

(a) Faites un histogramme des effectifs.

(b) Quel est le salaire médian dans cette entreprise ?

(c) Quel est le salaire moyen dans cette entreprise ?

Exercice 4 Une enquête sur l'argent de poche mensuel de 600 jeunes a donné les résultats regroupés sous la forme de l'histogramme suivant :



(a) Compléter le tableau suivant :

Somme d'argent en euros x_i	Effectif n_i	Fréquence f_i	Angle en degrés
$[0; 5[$			
$[5; 10[$			
$[10; 15[$			
$[15; 20[$			
$[20; 25[$			
$[25; 30[$			
Total	600	1	360

(b) Regrouper les données sous la forme d'un diagramme à secteurs.

(c) Calculer le montant moyen.

Exercice 5 On considère un ensemble de notes : $12; 4; 16; 16; 16; 7; 9; 12; 9; 12$.

(a) Faire un tableau d'effectifs et établir un diagramme en bâtons.

(b) On répartit les notes en quatre sous-groupes selon qu'elles appartiennent à l'intervalle : $[0; 5[$, $[5; 10[$, $[10; 15[$ ou $[15; 20[$.

(c) Faire un tableau d'effectifs et établir un histogramme.

Exercice 6 Trouver le 1er et le 3e quartile de la série : $27; 12; 4,5; 16; 25; 18; 7; 15; 12,5; 26; 18,5; 11$.

Exercice 7 Indiquer si les affirmations suivantes sont vraies ou fausses.

Si on ajoute 2 à toutes les valeurs d'une série statistique, on augmente

(a) la médiane de 2.

(b) la moyenne de 2.

(c) l'étendue de 2.

(d) le premier quartile de 2.