

Séries statistiques à une variable (Partie 1)

Concepts fondamentaux

F. Wlazinski

Licence d'économie

1 Introduction

Définition 1.1

Le *caractère*, appelé aussi l'*attribut* ou la *variable*, est l'objet de l'étude sur les *individus* d'une *population*. Les valeurs prises par la variable sont aussi appelées *modalités* ou *score*.

L'*effectif total* ou la *taille* souvent noté N ou n est le nombre d'individus de cette population.

Exemple 1.2

Remarque 1.3

Définition 1.4

On classe le caractère d'une population en quatre catégories :

- les variables *quantitatives discrètes* qui prennent des valeurs numériques isolées.
- les variables *quantitatives continues* qui prennent des valeurs numériques sous forme d'intervalles.
- les variables *qualitatives ordonnées* qui prennent des valeurs non numériques qui peuvent être classées dans un certain ordre.
- les variables *qualitatives (non ordonnées)* qui prennent des valeurs non numériques qui ne rentrent pas dans la catégorie précédente.

Remarque 1.5

2 Traitements des données : Classes, fréquences et graphiques

Remarque 2.1

Suivant la taille des données, on doit souvent en faire un classement.

Même si cela n'est pas nécessaire lorsque pour une série statistique de petite taille, il est souvent plus simple d'essayer de résumer les informations concernant une variable en comptant le nombre d'individus qui en possèdent la même valeur. C'est ce que l'on appelle la *distribution des effectifs*. Cette distribution peut se faire avec ou sans perte d'informations.

Exemples 2.2

Ces exemples seront repris dans la suite du cours.

Exemple A : On considère la fréquentation de l'après-midi d'une cafétéria universitaire pendant la semaine :

Jour	Lundi	Mardi	Mercredi	Jeudi	Vendredi
Effectifs	18	37	4	16	12

Exemple B : On considère les notes de deux étudiants sur un semestre qui sont :

6; 7; 8; 9; 11; 12; 13 et 14 pour le premier et 0; 2; 4; 6; 14; 16; 18 et 20 pour le second.

Exemple C : On considère les âges d'étudiants d'un groupe de Licence :

17; 18; 18; 18; 18; 18; 18; 18; 18; 18; 18; 19; 19; 19; 19; 19; 19; 19; 20; 20; 20 et 21

Exemple D : On considère les notes d'un étudiant sur un semestre qui sont coefficientés :

7 coefficient 3, 9 coefficient 1 et 12 coefficient 4.

Exemple E : On considère les notes d'un groupe d'étudiants pour un devoir qui sont :

0;0;0;0;0;1;1;2;3;3;5;5;5;5;5;7;7;7;7;7;8;8;8;9;9;10;10;10;10;10;10;10;10;11;11;11;12;12;12;12;12;14;14;16;17;17;18 et 20

Exemple F : On considère l'exemple E mais avec un classement avec perte d'information.

Intervalles	[0;5[[5;10[[10;15[[15;20]
n_i	10	16	17	5

Exemple G : On considère à nouveau l'exemple E mais avec un autre classement avec perte d'information.

Intervalles	[0;8[[8;12[[12;18[[18;20]
n_i	21	15	10	2

Définition 2.3

On appelle *fréquence (absolue)* le quotient de l'effectif d'une classe par rapport à l'effectif total.

Remarques 2.4

Exemples 2.5

Remarque 2.6

3 Mode

Définition 3.1

Le *mode* d'une série statistique est la valeur la plus fréquente du caractère.

Remarques 3.2

- Attention : Il existe des séries qui comportent plusieurs modes (distribution bi-modale ou multi-modale).
- Déterminer le mode dans le cas discret ne pose aucun problème.
- Dans le cas continu, la classe possédant le plus grand effectif est appelée la *classe modale*.

Exemple 3.3

Remarque 3.4

4 Fréquences cumulées, médiane et quantiles

A partir de cette section, nous travaillerons essentiellement avec des séries statistiques d'une variable quantitative (discrète ou continue).

Définition 4.1

Dans le cas d'une variable discrète, l'effectif cumulé croissant (resp. décroissant) associé à une valeur a est la somme des effectifs de toutes les modalités inférieures ou égales (resp. supérieures ou égales) à a dans la série statistique.

Dans le cas d'une variable continue, l'effectif cumulé croissant (resp. décroissant) de la classe $[a; b[$ est la somme des effectifs de cette classe et des classes qui précèdent (resp. qui suivent).

Exemples 4.2

Définition 4.3

Dans le cas d'une variable discrète, la fréquence cumulée croissante (resp. décroissante) associée à une valeur a est la somme des fréquences de toutes les modalités inférieures ou égales (resp. supérieures ou égales) à a dans la série statistique.

Dans le cas d'une variable continue, la fréquence cumulée croissante (resp. décroissante) de la classe $[a; b[$ est la somme des fréquences de cette classe et des classes qui précèdent (resp. qui suivent).

Remarque 4.4

Exemples 4.5

Remarques 4.6

Définition 4.7

Une médiane M_e d'une série statistique *ordonnée* est une valeur qui partage la population en deux groupes de même effectif.

Remarques 4.8

Définition 4.9

On peut étendre le principe de détermination de la médiane, c'est-à-dire le fractionnement en deux parties équivalentes de la distribution ordonnée, en fractionnant celle-ci en plusieurs parties équivalentes. Les valeurs ainsi obtenues sont des quantiles qui sont appelées quartiles pour 4 parties; déciles pour 10 parties et centiles pour 100 parties.

En particulier, les premier, second et troisième quartiles notés Q_1 , Q_2 et Q_3 sont les trois valeurs qui partagent en quatre la série de données ordonnées. Q_1 est approximativement la première valeur qui permet de dépasser les 25% de fréquences cumulées, Q_2 est égale à la médiane et Q_3 est approximativement la première valeur qui permet de dépasser les 75% de fréquences cumulées.

Remarques 4.10

Exemples 4.11

Remarque 4.12