Séries statistiques à une variable (Partie 2)

Caractéristique d'une distribution

F. Wlazinski

Licence d'économie

1 Moyenne arithmétique et variance

1.1 Cas discret

Définition 1.1

La *moyenne arithmétique*, notée $\overline{x}$ ou $E(x)$, d'une variable quantitative discrète x qui prend n valeurs x_1, x_2,..., x_n est $\overline{x} = \dfrac{x_1 + x_2 + \ldots + x_n}{n} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$.

Exemple 1.2

Rappelons les données de l'exemple B :

6; 7; 8; 9; 11; 12; 13 et 14 pour l'étudiant 1 et 0; 2; 4; 6; 14; 16; 18 et 20 pour l'étudiant 2.

Définition 1.3

Soit une variable quantitative x qui prend les valeurs $\underbrace{x_1, \ldots x_1}_{n_1\,fois}$, $\underbrace{x_2, \ldots x_2}_{n_2\,fois}$, ..., $\underbrace{x_p, \ldots x_p}_{n_p\,fois}$

avec $n_1 + n_1 + \ldots n_p = n$

Alors $\overline{x} = \dfrac{n_1 \times x_1 + n_2 \times x_2 + \ldots + n_p \times x_p}{n} = \dfrac{1}{n}\sum_{i=1}^{p} n_i x_i$.

Il s'agit d'une moyenne arithmétique dite *pondérée*.

Exemples 1.4

- Rappelons les données de l'exemple C :

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

Remarque 1.5

Avec les notations de la définition 1.3 et si on note $f_i \left(= \dfrac{n_i}{n}\right)$ la fréquence d'apparition de x_i, on peut obtenir la moyenne en fonction des fréquences : $\overline{x} = f_1x_1 + f_2x_2 + \ldots + f_px_p = \sum_{i=1}^{p} f_i x_i$.

Exemple 1.6

Rappelons les données de l'exemple C :

Age	17	18	19	20	21
Effectifs	1	10	7	3	1
Fréquences	$0,045$	$0,455$	$0,318$	$0,136$	$0,045$

Remarques 1.7

- Dans la suite du cours, on travaillera essentiellement avec des séries pondérées.

Lorsque ce n'est pas le cas, on peut trivialement attribuer un poids de 1 à toutes les valeurs.

- La somme de tous les écarts entre les valeurs x_i et la moyenne $\overline{x}$ est nulle i.e. $\displaystyle\sum_{i=1}^{p} n_i(x_i - \overline{x}) = 0$.
- La somme de tous les carrés écarts entre les valeurs x_i et la moyenne $\overline{x}$ est un minimum

i.e. $\displaystyle\sum_{i=1}^{p} n_i(x_i - \overline{x})^2 = \min_{\alpha\in\mathbb{R}} \sum_{i=1}^{p} n_i(x_i - \alpha)^2$.

Remarque 1.8

Remarque 1.9

On s'intéresse maintenant à l'etude de la *fluctuation* autrement appelée la *dispersion* autour d'une valeur centrale (mode, médiane ou moyenne) d'une série.

Définition 1.10

L'*étendue* d'une série est la différence entre la plus grande et la plus petite valeur de celle-ci.

Définition 1.11

La *variance* d'une série est la moyenne arithmétique des carrés des écarts à la moyenne arithmétique.
Autrement dit, avec les notations de la définition 1.3, on a :

$$V(x) = \frac{n_1\left(x_1 - \overline{x}\right)^2 + n_2\left(x_2 - \overline{x}\right)^2 + \cdots + n_p\left(x_p - \overline{x}\right)^2}{n} = \frac{1}{n}\sum_{i=1}^{p} n_i\left(x_i - \overline{x}\right)^2$$

L'*écart type* σ de la série est la racine carrée de la variance c'est-à-dire $\sigma_x = \sigma(x) = \sqrt{V(x)}$.

Remarque 1.12

Avec les notations de la définition 1.3, on a :

$$E(x^2) = \overline{x^2} = \frac{n_1 \times x_1^2 + n_2 \times x_2^2 + \ldots + n_p \times x_p^2}{n} = \frac{1}{n}\sum_{i=1}^{p} n_i x_i^2.$$

Ou encore :

$$E(x^2) = \overline{x^2} = \sum_{i=1}^{p} f_i x_i^2.$$

Propriété 1.13

La variance est égale à la moyenne (arithmétique) des carrés moins le carré de la moyenne (arithmétique) c'est-à-dire $V(x) = \overline{x^2} - \overline{x}^2 = E(x^2) - E(x)^2$.

Exemples 1.14

Remarque 1.15

Dans le cadre de test et lorsque l'on travaille avec un petit échantillon de la population, on corrige ces valeurs par *le biais de petit échantillon.*

La *variance corrigée* est alors $S_x^2 = V(x) \times \dfrac{n}{n-1}$.

Il s'en suit que l'*ecart-type corrigée* est $S_x = \sqrt{S_x^2} = \sigma_x \times \sqrt{\dfrac{n}{n-1}}$.

Remarque 1.16

Définition 1.17

Le *coefficient de variation* noté CV_x d'une série x est une mesure de dispersion relative et sa valeur est $CV_x = \dfrac{\sigma_x}{\overline{x}}$.

Remarque 1.18

Exemples 1.19

Propriété 1.20

Soient x_1 et x_2 deux séries quantitatives d'effectifs respectifs n_1 et n_2, de moyennes $E(x_1)$ et $E(x_2)$ et de variances $V(x_1)$ et $V(x_2)$.
La moyenne de la série x conjointe de x_1 et x_2 est $E(x) = \dfrac{n_1 \times E(x_1) + n_2 \times E(x_2)}{n_1 + n_2}$.
La variance de x est $V(x) = \dfrac{n_1(V(x_1) + (E(x) - E(x_1))^2) + n_2(V(X_2) + (E(x) - E(x_2))^2)}{n_1 + n_2}$.

Exemple 1.21

Définition 1.22

On utilise les notations de la définition 1.3.
On appelle série des *masses* de la série x la série des $n_i \times x_i$.

Exemple 1.23

Définition 1.24

La *mediale* de la série x est la médiane de la série des masses de x.

Exemple 1.25

1.2 Cas continu

Définition 1.26

Soient $a < b$ deux réels. Le centre de tout intervalle de a à b est la moyenne $\dfrac{a+b}{2}$.

Définition 1.27

Soit x une variable quantitative continue dont les valeurs sont dans les intervalles $[x_i; x_{i+1}[$ pour i compris entre 1 et $p-1$. On appelle *centres des classes* de x les centres des intervalles $[x_i; x_{i+1}[$. Ils sont généralement noté c_i pour i de 1 à $p-1$.

Exemple 1.28

Rappelons les données de l'exemple G

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

Principe

Définition 1.29

La moyenne arithmétique, notée $\overline{x}$ ou $E(x)$, d'une variable quantitative continue x continue dont les valeurs sont dans les les intervalles $[x_i; x_{i+1}[$ de centre c_i pour i de 1 à $p-1$ et dont les effectifs respectifs sont n_1, $n_2, \ldots, n_p$ est $\overline{x} = \dfrac{n_1 \times c_1 + n_2 \times c_2 + \ldots + n_p \times c_p}{n_1 + n_2 + \ldots + n_p} = \dfrac{1}{n} \sum_{i=1}^{p} n_i c_i$.

Exemple 1.30

Définition 1.31

Avec les notations de la définition 1.29, la variance de x et l'écart type sont :

$$V(x) = \frac{n_1 (c_1 - \overline{x})^2 + n_2 (c_2 - \overline{x})^2 + \dots n_p (c_p - \overline{x})^2}{n} = \frac{1}{n} \sum_{i=1}^{p} n_i (c_i - \overline{x})^2$$

$\sigma_x = \sigma(x) = \sqrt{V(x)}$.

Remarque 1.32

Exemple 1.33

Définition 1.34

On utilise les notations de la définition 1.29.
On appelle série des *masses* de la série x la série des $n_i \times c_i$.

Exemple 1.35

Définition 1.36

La définition de la médiale d'une série x quantitative continue est la même que pour une série discrète : c'est la médiane de la série des masses de x.

Exemple 1.37

2 Courbe de concentration et indice de Gini

Définition 2.1

Soit x une variable statistique quantitative. La courbe obtenue en reliant les points dont les abscisses sont les fréquences cumulées croissantes et les ordonnées les fréquences cumulées croissantes des masses est appelée *courbe de concentration* ou *courbe de Lorentz*.
Cette courbe est inscrite dans un carré de côté 1 appelé *carré de Gini*.

Remarques 2.2

Exemple 2.3

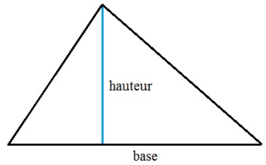
Définition 2.4

L'*aire de concentration* est l'aire de la surface située entre la courbe de concentration et la première bissectrice.
L'*indice de Gini* est un indice de concentration qui vaut deux fois l'aire de concentration.

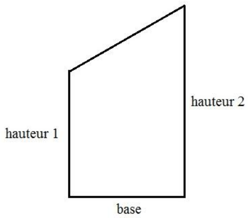
Remarques 2.5

Exemple 2.6

Rappels



L'aire du triangle est $\dfrac{\text{base} \times \text{hauteur}}{2}$.



L'aire du trapèze est $\dfrac{\text{hauteur 1} + \text{hauteur 2}}{2} \times \text{base}$.

Exemple 2.7

3 Les différentes moyennes

Nous utilisons dans cette section, les notations de la définition 1.3. Il suffit de remplacer x_i par c_i pour étendre les définitions qui suivent à une variable quantitative continue.

Définition 3.1

La moyenne *quadratique* est définie par $Q(x) = \sqrt{\dfrac{1}{n}\displaystyle\sum_{i=1}^{p} n_i x_i^2}$.

Remarque 3.2

Définition 3.3

La moyenne *géométrique* est définie par $G(x) = \left(x_1^{n_1} \times x_2^{n_2} \times \ldots \times x_p^{n_p}\right)^{1/n}$.

Exemple 3.4

Définition 3.5

La moyenne *harmonique* est définie par $H(x) = \dfrac{n}{\displaystyle\sum_{i=1}^{p} \dfrac{n_i}{x_i}}$.

Exemple 3.6