S4 2019-2020

Nathan TONNELLE

Statistiques

MINFO0401

Table des matières

[Statistique descriptive 2](#_Toc30494219)

[Vocabulaire 2](#_Toc30494220)

[Population 2](#_Toc30494221)

[Echantillon 2](#_Toc30494222)

[Caractère et modalité 2](#_Toc30494223)

[Représentation graphique 3](#_Toc30494224)

[Variables qualitatives 3](#_Toc30494225)

[Variable quantitative 3](#_Toc30494226)

[La fonction cumulative 3](#_Toc30494227)

# Statistique descriptive

## Vocabulaire

### Population

Une population est l'ensembe des infividus ou objets sur lesquels portent une études statistique, on le note P

#### Exemple

1. Logement d'une ville
2. Personnel d'une enteprise
3. Animaux d'un parc naturel

### Echantillon

est une partie de la population e étudier sur laquelle porte l"tude statistique.

Une étude statistique portant sur un echantillon est appelé sondage.

On appelle étude associative ou ressenssement si elle porte sur l'ensemble de la population.

### Caractère et modalité

Une étude statistique porte sur 1 ou plusieurs caractères communs à tout les individus de la population à étudier. Un caractère est aussi appelé variable.

#### Exemple

1. Surface de logement
2. Age, ancienneté, revenu
3. Régime alimentaire, espèces

#### Modalité

Les modalités d'un caractère sont les différentes valeurs que peut prendre se caractère sur les individu de la population étudiée

##### Exemple

1. R+
2. [16,10)
3. Herbivore, carnivore, omnivore

On distingue 2 types de caractères :

* Caractère qualitatif :

Les modalités sont des attributs qualitatif (ex: régime alimentaire)

* Caractère quantitatif

Les modalités sont des quantités numériques (ex : age)

Les caractères quantitatifs sont de 2 types

* Variable discrète :

Les modalités de la variable appartiennent à un ensemble discret tel N, Z, N²

* Variable continue :

Les modalitées de la variable prennent des valeurs dans un ensemble continu tel que R

#### Remarque :

Pour étudier une variable ontinue on constitu des classes de valeurs possibles, ces classes sont des intervals d'amplitude égale ou inégale est constitué alors de nouvelles modalité ou des caractères.

#### Attention :

Le découpage en classes peut influer sur les resultats et les interprétations que l'on peut faire.

Sil est trop important, il risque de faire apparaitre des irrégularités artificiellles car les effectifs des classes seront trop faibles.

Sil est trop grossier, il conduira à une perte d'information.

#### Effectif et fréquence

L'effectif d'une modalité ou d'une classe de modalité est le nombre d'individu de la population correspondant à cette modalité ou à cette classe de modalités. On note ni pour la ième modalité.

La fréquence de la ième modalité (ou classe) est donnée par le rapport de son effectif sur l'effectif total de la population noté n.

On a alors fi=ni/n

#### Remarque :

Si la variable X possède K modalités x1,..,xk alors

i=n i

## Représentation graphique

Il existe différentes façons de représenter graphiquement des variables

### Variables qualitatives

#### Diagramme circulaire

C'est un disque dans lequel chaque modalité est représenté par un secteur angulaire proportionnel à sa fréquence.

Pour 1<=i<=f **α** i=360xfi

#### Diagramme en tuyaux d'orgue

C'est un diagramme formé de rectangles tous de même largeur et dont les hauteurs sont proportionnelles aux fréquences des modalités

### Variable quantitative

Il y a 2 sortes de représentation graphique des variables quantitatives :

* Diagramme différentiel
* Diagramme intégrale

#### Variables discrètes

Le diagramme différentiel utilisé dans le cas d'une variable discrète est le diagramme à bâton.

Où chaque bâton est de longueur ou de hauteur proportionnel à la fréquence de la modalité correspondante

##### Exemple :

Nombre d'enfants dans un échantillon de 50 familles.

Nombre d'enfant ni fi

0 9 0.18

1 7 0.14

2 12 0.24

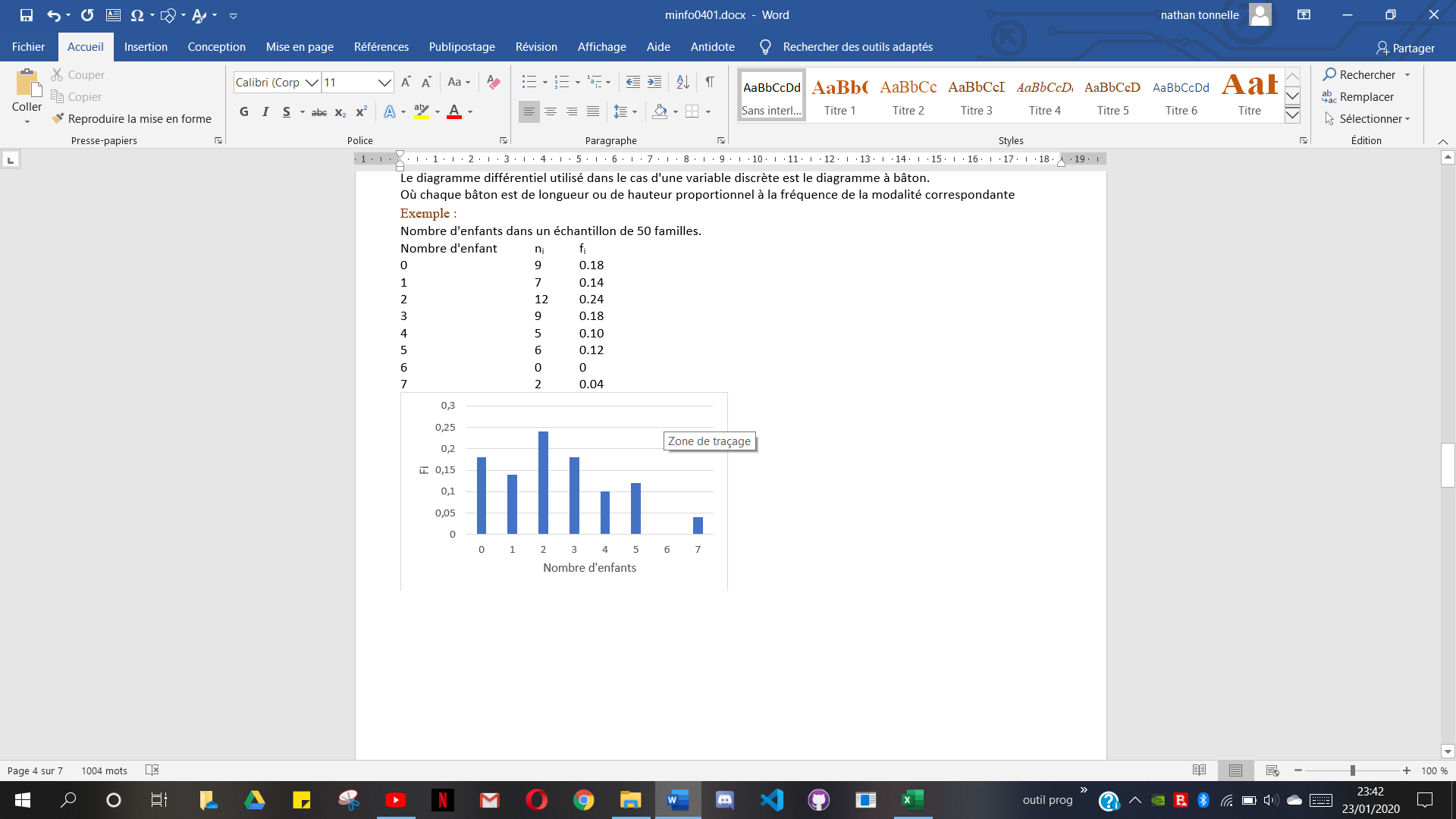
3 9 0.18

4 5 0.10

5 6 0.12

6 0 0

7 2 0.04



### La fonction cumulative

La fonction cumulative d'une variable X prise en un point x , noté F(x) est définie comme la proportion de la population pour laquelle la variable X prend des valeurs >= à x

Si les modalités de X sont x1,x2,…xk alors pour xi<=x<xi+1, F(x)=

Exemple :

Si x<0 F(x)=0

0<=x<1 F(x)=0.18

1<=x<2 F(x)=0.32

2<=x<3 F(x)=0.56

3<=x<4 F(x)=0.74

4<=x<5 F(x)=0.84

5<=x<7 F(x)=0.86

x>=7 F(x)=1

## Variables continues

Si ei et ei+1 sont les extrémités de la classe n°i, noté [ei,ei+1], on notera ci soit milieu et ai son amplitude, 1<=i<=k

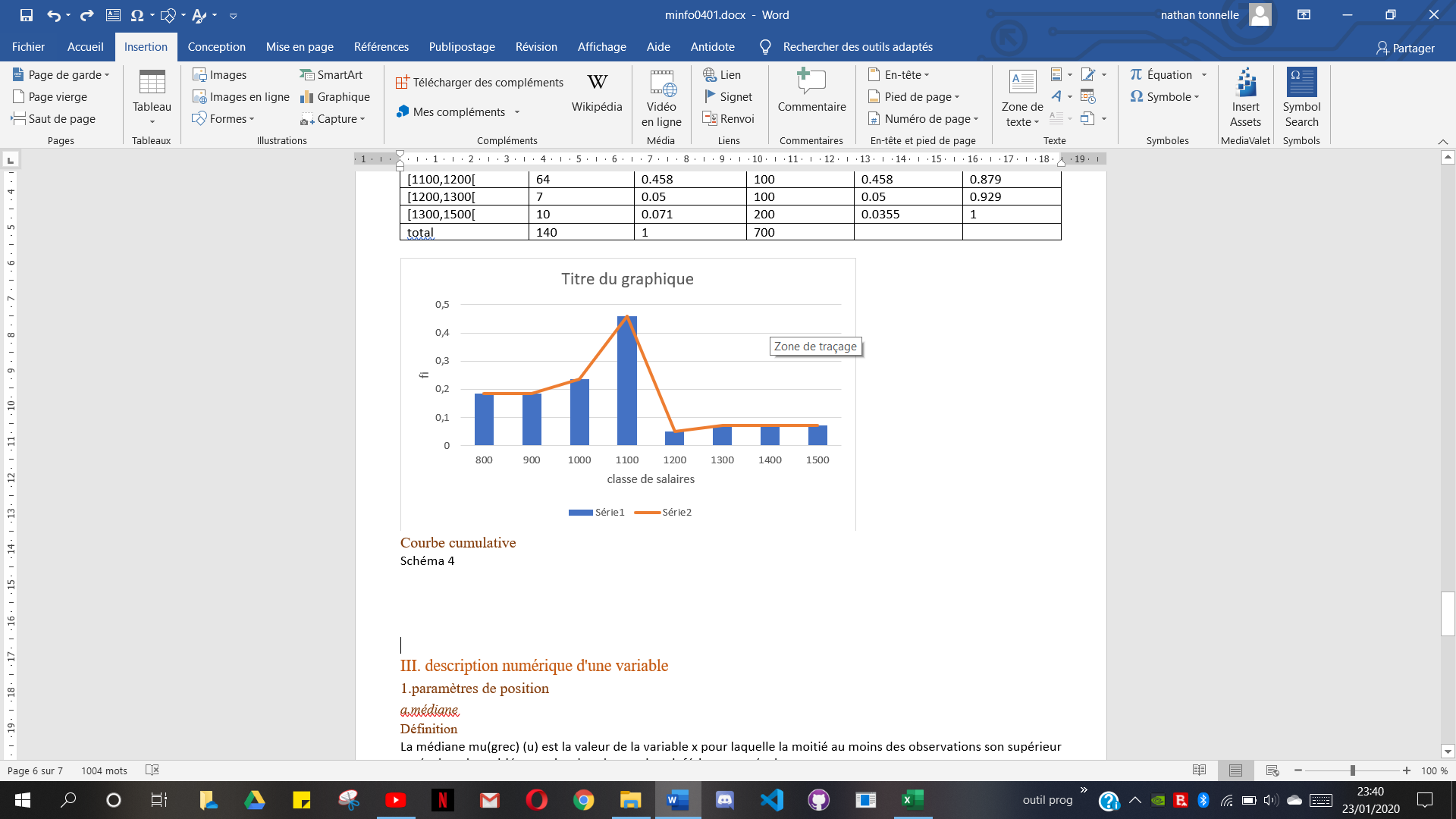
### Définition (représentation graphique différentielle)

Un histogramme est une représentation graphique où chaque classe est représenté par un rectangle de base proportionnelle à son amplitude et de surface proportionnelle à sa fréquence. Ainsi la hauteur de la classe n°i est hi=fi/ai. Un polygone statistique est un polygone reliant le milieu des bases supérieurs des rectangles de l'histogramme.

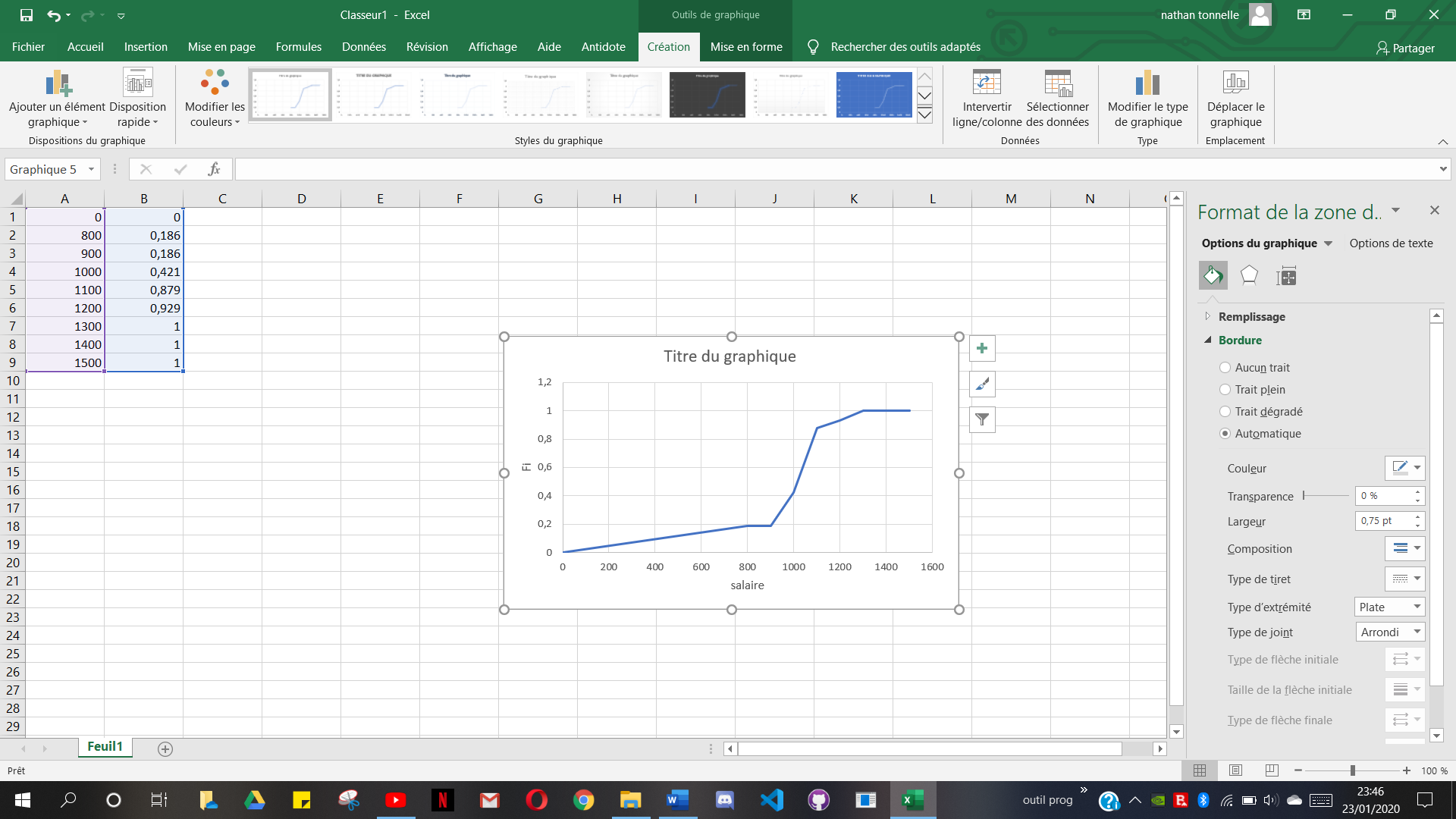
### Exemple :

Salaire mensuel net des ouvriers d'un établissement industriel.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Classes de salaire | ni | fi | ai | hi=fi/ai (echelle x100) | Fi |
| [800,1000[ | 26 | 0.186 | 200 | 0.096 | 0.186 |
| [1000,1100[ | 33 | 0.235 | 100 | 0.235 | 0.421 |
| [1100,1200[ | 64 | 0.458 | 100 | 0.458 | 0.879 |
| [1200,1300[ | 7 | 0.05 | 100 | 0.05 | 0.929 |
| [1300,1500[ | 10 | 0.071 | 200 | 0.0355 | 1 |
| total | 140 | 1 | 700 |  |  |



### Courbe cumulative



## III. description numérique d'une variable

### 1.paramètres de position

#### a.médiane

##### Définition

La médiane mu(grec) (u) est la valeur de la variable x pour laquelle la moitié au moins des observations son supérieur ou égale et la moitié au moins des observations inférieures ou égales

##### Remarque

Dans le cas continue on détermine d'abords la classe médiane avant de calculer le point le médian par la méthode d'interpellation linéaire.

##### Exemple

1. (nombre d'enfants par famille) u=2 cara (F2=0.32 et F3=0.56)
2. (salaires des ouvriers)

Classe médiane = [1100,1200[

Calcul de la médiane par la méthode d'interpellation linéaire

#### b.le mode

##### définition

le mode est la valeur de la variable x ayant la plus grande fréquence

##### remarque

certaines séries statistiques peuvent avoir plusieurs modes. Dans le cas continu on parle de classe modale, on veillera cependant a tenir compte de l'amplitude des classe. La classe modale correspond à la calsse ayant la plus grande hauteur hi.

##### exemple

1. (salaire des ouvriers)

Classe modale =[1100,1200[

*Equation (M M2)*

*Equation(M3 M4)*

#### c.la moyenne

soit x une variable prenant les valeurs x1,…,xk avec les effectifs n1,…,nk respectivement (avec les fréquences f1,…,fk). alors la moyenne de la variable x est donnée par

Où

dans le cas continu, la moyenne d'une variable x est définie par

où les Ci sont les milieux des classes [ei,ei+1[

##### Propriétés

1. Linéarité

Si on considère la transformation Y=aX+b

Alors la moyenne de Y est = a +b

1. Si on définit la fonction

Alors L(c) prend son minimum pour c=

1. Si P=P1UP2

Où la moyenne de X sur P1 est , et l'effectif est n1; la moyenne de X sur P2 est 2 et l'effectif de P2 est n2, alors la moyenne de X sur P est :

### 2. paramètre de dispersion

#### a. Etendu

l'etendu est la différence entre la plus grande et la plus petite valeur de la série statistique.

#### b. l'ecart moyen obsolu

on considère une série de n observations dont les modalités sont x1,…,xk.

on note Xbarre sa moyenne empirique.

Alors l'écart-type moyen absolu est défini par

Schéma 1

#### c. variance (ecart quadratique moyen)

si x1,…,xn ston les modalités d'une variance X observée n fois, alors la variance de X est définie par **schéma 2**

propriété :

schéma 3

#### d. ecart type

l'ecart type est défini par **schéma 4**

intérêt :

X est en Km => en Km²

L'écart type exprime la dispersion dans la même unité de mesure que la variable X

### 3. coefficient de variation

Le coefficient de variation est défini par **schéma 5**

Intérêt:

Ce coefficient est indépendant de l'unité de mesure. Il permet alors de comparer les dispersions de sens statistiques exprimés dans des unités de mesure différents.

### 4. les moments

Définition

On appelle moment d'ordre t (t appartient à N) par rapport à une constante a d'une variable statistique x

**Schéma 6**

Les moments non centrés correspondent à **schema 7**

Les moments centrés correspondent à a=Xbarre **schéma 8**

Remarque :

**Schéma 9**

### Caractéristiques de forme

#### a. coefficient d'asymétrie

le coefficient d'asymétrie de Fisher est défini par **schéma 10**

si A=>B alors Bbarre => Abarre

remarque :

si la distribution est symétrique, alors **schéma 11**

#### b. coefficient d'aplatissement (Kertosis)

**schéma 12**

il compare l'aplatissement de courbe statistique à la courbe de la loi N(0,1).

**Schéma 13**

Box-plot(boite à moustache)

Definition :

Premier quartile q1 est la valeur de la variable x pour laquelle au moins ¼ des observations lui sont inférieur ou égales et au moins ¾ des observations supérieures ou égales.

Le troisième quartile q3 est la valeur de la variable x pour laquelle au moins ¾ des observations lui sont inférieur ou égales et au moins ¼ des observations supérieures ou égales.

Remarque :

Dans le cas d'une variable continue, on détermine d'abord les classes contenant le premier quartile q1 et le 3eme quartile q3 avant de procéder par la méthode de l'interpellation linéaire au calcul de q1 et q3

**Schéma 14**