Université sidi Med Ben Abdellah
Faculté des sciences Juridiques, Economiques et Sociales
FES

Année Universitaire 2011-12

# STATISTIQUE DESCRIPTIVE

SEMESTRE 1

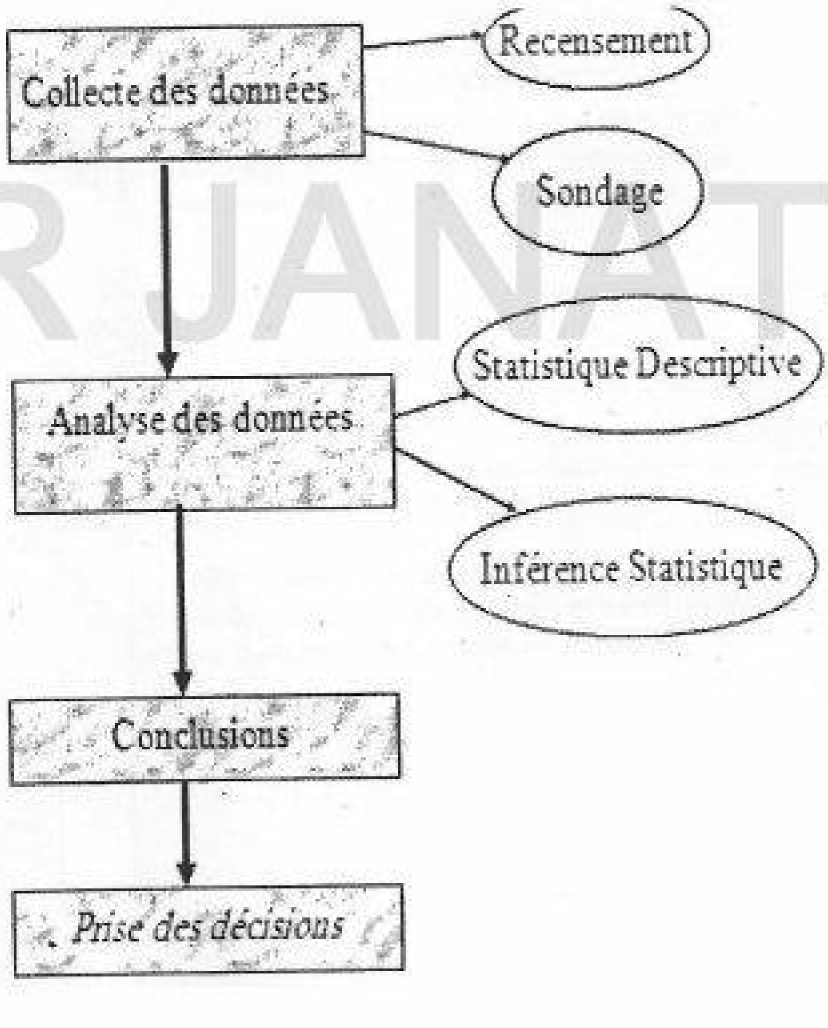
*Sections A & B*

MODULE : *Méthodes Quantitatives I*

ELÉMENT : *Statistique I*

D.TOUIJAR





Sections A &B

Année Universitaire 2011-12

Semestre S1

Enseignant: Driss TOUIJAR

**Filière** : Sciences Economiques et Gestion

**Module** : Méthodes Quantitatives I

**Elément :**

**STATISTIQUE I**

## Attention !

*N'essayez pas de comprendre le cours en lisant tout (e) seul(e) Ce document.*

*Par contre, je vous recommande vivement d'assister à toutes les séances en espérant mieux cerner le programme de statistique I*

# INTRODUCTION GENERALE

## la Statistique et Les Statistiques

– Le mot statistique désigne à la fois un ensemble de données d'observations et l'activité qui consiste dans leur recueil, leur traitement et leur interprétation.

–D'une façon plus précise :

- On désigne par Les statistiques, un ensemble de données ou d'informations relatives à un phénomène ou à un processus donné;

  exemple : la population marocaine en 2007, les naissances au Maroc en 2006, l'évolution des entreprises, des emplois...

- Par contre La statistique est en général, un ensemble de méthodes scientifiques qui servent à décrire et à analyser des données. Elles nous permettent aussi de tirer des conclusions et de prendre des décisions et aussi de faire des prévisions.

✓ En ce qui concerne ce semestre, on se contentera d'étudier la méthode descriptive : c'est une méthode qui vise à décrire des ensembles nombreux ; d'où l'appellation:

**statistique descriptive**.

Elle a pour objet de présenter les données sous forme de tableaux et de graphiques et de les résumer en quelques valeurs numériques appelée

**caractéristiques.**

## (b) *Changement d'origine et d'échelle*

- ***Propriété :*** Soit $X$ une variable statistique de moyenne arithmétique $\bar{x}$. Si $Y$ est une variable statistique telle que $Y=aX+b$, où $a$ et $b$ sont des réels quelconques, alors la moyenne arithmétique de $Y$ est :

$$\boxed{\bar{y} = a\bar{x} + b}$$

Démonstration :

$$\bar{y} = \sum_{i=1}^{k} f_i y_i = \sum_{i=1}^{k} f_i (ax_i + b)$$

$$= a\underbrace{\sum_{i=1}^{k} f_i x_i}_{\bar{x}} + b\underbrace{\sum_{i=1}^{k} f_i}_{1} = a\bar{x} + b$$

## (C) *Propriétés algébriques de la moyenne arithmétique*

- *i)* la moyenne des écarts à la moyenne arithmétique est nulle :

$$\sum_{i=1}^{k} f_i (x_i - \bar{x}) = 0$$

- *ii)* La moyenne des carrés des écarts à une constante a est minimale pour $a = \bar{x}$

$$\min_{a \in \blacklozenge} \sum_{i=1}^{k} f_i (x_i - a)^2 = \sum_{i=1}^{k} f_i (x_i - \bar{x})^2$$

**Démonstration :**

$$\sum_{i=1}^{k} f_i (x_i - a)^2 = \sum_{i=1}^{k} f_i \left[(x_i - \bar{x}) + (\bar{x} - a)\right]^2$$

$$= \sum_{i=1}^{k} f_i (x_i - \bar{x})^2 + \sum_{i=1}^{k} f_i (\bar{x} - a)^2$$

$$+ 2\sum_{i=1}^{k} f_i (x_i - \bar{x})(\bar{x} - a)$$

$$= \sum_{i=1}^{k} f_i (x_i - \bar{x})^2 + (\bar{x} - a)^2 \underbrace{\sum_{i=1}^{k} f_i}_{=1}$$

$$+ 2(\bar{x} - a)\underbrace{\sum_{i=1}^{k} f_i (x_i - \bar{x})}_{=0}$$

$$= \sum_{i=1}^{k} f_i (x_i - \bar{x})^2 + (\bar{x} - a)^2$$

## (d) *Propriété de l'agrégation*

- *Soit une population* P *de taille $n$, composée de m sous populations* $P_1, P_2, \ldots, P_m$ *; de tailles respectives $n_1, n_2, \ldots n_m$ et de moyennes respectives $\bar{x}_1, \bar{x}_2, \ldots \bar{x}_m$. Alors la moyenne arithmétique $\bar{x}$ de la population* P *est donnée par :*

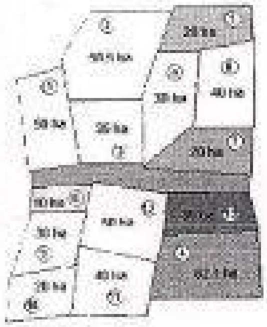
$$\boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_i \bar{x}_i}$$

- **Exemple :**

Le salaire moyen des cadres dans l'entreprise E est de 4000 DH.

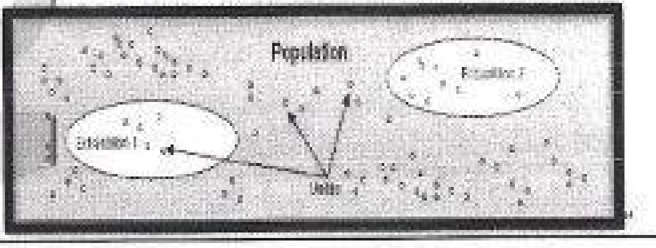
- Le salaire moyen des cadres masculins est de 4200 DH.
- Le salaire moyen des cadres féminins est de 3000 DH.

1. Quelle est la répartition hommes - femmes des cadres ?



Population de n exploitations agricoles

- ***Définition 2** : Les éléments qui composent une population sont appelés des Individus (ou unités Statistiques). Un sous-ensemble d'une population est appelé Échantillon.*




- *Définition 3 : La taille d'une population est le nombre d'individus qui la composent.*
- *Définition 4 : Un caractère est un critère relatif auquel on observe les individus d'une population.*

- A chaque individu, on attribut un ou plusieurs caractères qui peuvent être soit **quantitatifs** (s'ils sont mesurables; exemple : salaire, nb d'enfants par ménage...) ou **qualitatifs** (sinon; exemple : sexe, état matrimonial... ).

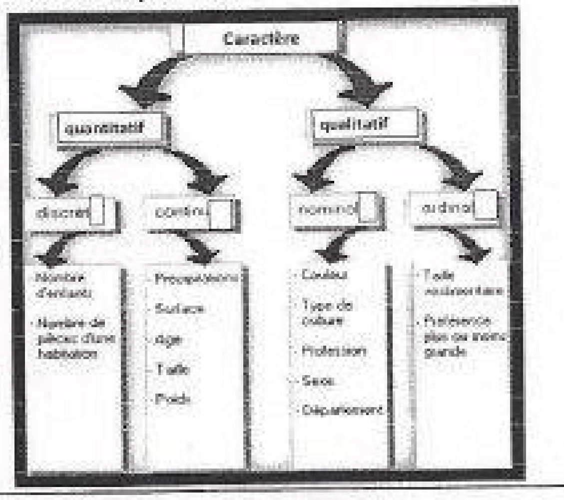
- Une valeur que peut prendre un caractère s'appelle **modalité**.

- Un caractère **qualitatif** peut être soit :
  - **Ordinal** : si ses modalités peuvent être naturellement ordonnées exemple : satisfaction plus ou moins grande après l'achat d'un produit.
  - **Nominal** : si ses modalités **ne** peuvent être naturellement ordonnées exemple : état matrimoniale.

- On appelle **variable statistique**, un caractère **quantitatif**.
  - On distingue deux sortes de variables statistiques:

- **Les variables statistiques discrètes** (notées: **v.s.d.**) : se sont des variables dont l'ensemble des modalités est un ensemble discret (la variable ne peut prendre que des valeurs isolées d'un intervalle).
  - Exemple : Pour le nombre d'enfants par ménage l'ensemble des modalités peut être {0, 1, 2, 3, 4}.
- **Les variables statistiques continues** (**v.s.c.**) : dans ce cas, l'ensemble des modalités est continue; la variable peut prendre toutes les valeurs d'un intervalle.
  - **Exemple :** Salaire, âge, taille, poids ...etc.

- Tableau Récapitulatif



- **Exemple :** On observe, au cours d'une semaine, 20 machines selon le nombre de pièces défectueuses produites :
- **8, 16, 9, 33, 14, 5, 3, 7, 10, 7, 9, 9, 3, 8, 3, 3, 5, 14, 8, 7.**

→On l'appelle série brutes.
→L'effectif total de la population est donc

***n =20***

- En classant ces nombres par ordre croissant, on obtient la série ordonnée :
- **3, 3, 3, 3, 5, 5, 7, 7, 7, 8, 8, 8, 9, 9, 9, 10, 14, 14,16, 33.**
- On obtient les **K** modalités après regroupement des observations :
- **3, 5, 7, 8, 9, 10, 14, 16, 33 ⇨ (k=9) < (n=20)**

k = nombre de modalités
n : taille de la population
$K \leq n$

| | | | |
|---|---|---|---|
| → la modalité | 3 | a pour effectif | $n_1=4$ |
| → la modalité | 5 | a pour effectif | $n_2=2$ |
| → la modalité | 7 | a pour effectif | $n_3=3$ |
| → la modalité | 8 | a pour effectif | $n_4=3$ |
| → la modalité | 9 | a pour effectif | $n_5=3$ |
| → la modalité | 10 | a pour effectif | $n_6=1$ |
| → la modalité | 14 | a pour effectif | $n_7=2$ |
| → la modalité | 16 | a pour effectif | $n_8=1$ |
| → la modalité | 33 | a pour effectif | $n_9=1$ |

- *Remarque : 4+2+3+3+3+1+2+1+1=20*

- ***Définition*** *: L'effectif $n_i$ d'une modalité $x_i$ est le nombre d'individus ayant cette modalité. L'effectif total (ou taille) d'une population, noté **n** , est le nombre d'individus qui composent cette population.*
- *On a donc :*

n : effectif total
ni : effectif partiel

$$\boxed{\sum_{i=1}^{k} n_i = n} \quad (\overbrace{n_1 + n_2 + \dots + n_k}^{k \text{ éléments}})$$

- **Définition** : *On appelle fréquence de la modalité $x_i$, la proportion des individus présentant cette modalité. On écrit :* $\boxed{f_i = \frac{n_i}{n}}$ $; i = 1, \dots, k$
- Remarque :

$$\boxed{\sum_{i=1}^{k} f_i = 1}$$

- En %:

$$\boxed{f_i\% = f_i \times 100 \Rightarrow \sum_{i=1}^{k} f_i\% = 100}$$

$$\sum_{i=1}^{K} f_i = \sum_{i=1}^{K} \left(\frac{n_i}{n}\right) = \frac{1}{n} \sum_{i=1}^{K} n_i = \frac{1}{n} \times n = 1$$

- ***Exemple :***

$$\boxed{f_1 = \frac{n_1}{n} = \frac{4}{20} = 0{,}20 \text{ et } f_6 = \frac{n_6}{n} = 0{,}05}$$

- **Commentaire :** La proportion des machines ayant produit 3 pièces défectueuses est de 20%; et celle des machines ayant produit 10 pièces défectueuses est de 5%.
- ***Définition***: *On appelle distribution d'un caractère **X**, l'ensemble de couples* $\{(x_1,n_1),(x_2,n_2),\dots,(x_k,n_k)\}$

•**Remarque :**

–En terme de fréquence, la distribution de $X$, s'écrit aussi :

$$\boxed{\{(x_1, f_1), (x_2, f_2), \ldots, (x_k, f_k)\}}$$

*Exemple* : *La distribution des défectuosités des 20 machines est :*

$$\{(03;4),(5;2),(7;3),(8;3),(9;3),(10;1),(14;2),(16;1),(33;1)\}$$

•*Ou* $\{(3;0{,}20),(5;0{,}10),(7;0{,}15),(8;0{,}15)\ldots\}$

25

– *Pour une meilleure exploitation de la distribution, on a intérêt à la représenter par un tableau statistique.*

## • II) TABLEAUX STATISTIQUES

• *Exemple Introductif :* Supposons que l'on ait fait une enquête auprès de 20 femmes selon 9 caractères : Prénom, nom, jour de naissance, mois de naissance, années de naissance, nombre d'enfants, revenu annuel du ménage, ville natale, opinion sur la qualité d'un produit alimentaire pour bébé.

– Ces données ont été reportées sur un bordereau, sous forme d'une matrice de 20 éléments qui comprennent chacun les 9 données concernant une femme.

26

Cette matrice (série selon plusieurs variables) peut être représentée comme suit :

|  | Prénom | Nom | Jour | Mois | Année | Nombre d'enfants | Revenu | Ville | Opinion |
|---|---|---|---|---|---|---|---|---|---|
|  |  | Date de naissance |  |  |  |  |  |  |  |
| 1 | Aline | Michaud | 23 | 1 | 1953 | 4 | 30000 | Paris | Mauvaise |
| 2 | Mathilde | Crepineau | 13 | 10 | 1953 | 0 | 80000 | Montpellier | Très bonne |
| 3 | Bénédicte | Anselin | 8 | 3 | 1953 | 1 | 5000 | Rouen | Mauvaise |
| 4 | Henriette | Tafie | 3 | 2 | 1953 | 1 | 13000 | Paris | Mauvaise |
| 5 | Danielle | Crice | 16 | 4 | 1953 | 2 | 40000 | Marseille | Moyenne |
| 6 | Ludivine | Lagorce | 15 | 5 | 1953 | 3 | 40000 | Marseille | Moyenne |
| 7 | Agnès | Roche | 2 | 5 | 1953 | 2 | 10000 | Nice | Bonne |
| 8 | Rita | Mona | 5 | 5 | 1953 | 3 | 45000 | Paris | Moyenne |
| 9 | Andrée | Laminal | 22 | 6 | 1953 | 3 | 80000 | Nancy | Passable |
| 10 | Pauline | Zaël | 23 | 1 | 1953 | 4 | 30000 | Nice | Moyenne |
| 11 | Zoé | Forec | 9 | 9 | 1953 | 2 | 60000 | Nice | Passable |
| 12 | Lola | Marseille | 7 | 7 | 1953 | 3 | 55000 | Marseille | Bonne |
| 13 | Priscille | Loumad | 3 | 2 | 1953 | 3 | 85000 | Montpellier | Bonne |
| 14 | Violaine | Tyrk | 16 | 4 | 1953 | 5 | 60000 | Nice | Moyenne |
| 15 | Christine | Dardue | 15 | 6 | 1953 | 2 | 40000 | Rouen | Bonne |
| 16 | Fabielle | Coule | 2 | 6 | 1953 | 1 | 10000 | Nancy | Passable |
| 17 | Noëlle | Gore | 5 | 11 | 1953 | 3 | 120000 | Nice | Mauvaise |
| 18 | Rachel | Mel | 22 | 2 | 1953 | 3 | 80000 | Paris | Moyenne |
| 19 | Sabine | Eboum | 23 | 10 | 1953 | 4 | 18000 | Paris | Très bonne |
| 20 | Jeanne | Rivière | 9 | 12 | 1953 | 5 | 90000 | Marseille | Passable |

Dans cette 1[ère] Partie, on ne considérera qu'un seul caractère à la fois.

### Représentation de la distribution d'un caractère X par un Tableau

**1) Cas d'un caractère qualitatif :**

– Soit la distribution d' un caractère qualitatif $X$ étudié sur une population de $n$ individus :

$$\{(x_1, n_1), (x_2, n_2), \ldots, (x_k, n_k)\}$$

28

• Sa représentation par tableau est alors comme suit :

| Modalité $x_i$ | Effectif $n_i$ | Fréquence $f_i$ |
|---|---|---|
| $x_1$ | $n_1$ | $f_1$ |
| $x_2$ | $n_2$ | $f_2$ |
| ⋮ | ⋮ | ⋮ |
| $x_i$ | $n_i$ | $f_i$ |
| ⋮ | ⋮ | ⋮ |
| $x_k$ | $n_k$ | $f_k$ |
| Total | $n=\sum n_i$ | $1=\sum f_i$ |

29

• **Exemple :** On reprend les 20 femmes selon l'opinion «caractère ordinal» :

• Ma, Tb, Ma, Ma, Mo, Mo, Bo, Mo, Pa, Mo, Pa, Bo, Bo, Mo, Bo, Pa, Ma, Mo, Tb, Pa.

• Une fois classées :
Ma, Ma, Ma, Ma, Pa, Pa, Pa, Pa, Mo, Mo, Mo, Mo, Mo, Mo, Bo, Bo, Bo, Bo, Tb, Tb.

• On a donc 5 modalités ($k=5$).

• *La distribution s'écrit :*

$$\{(Ma;4),(Pa;4),(Mo;6),(Bo;4),(Tb;2)\}$$

30

- *Le Tableau statistique est comme suit :*

| $x_i$ | $n_i$ | $f_i$ |
|---|---|---|
| *Ma* | 4 | 0,2 |
| *Pa* | 4 | 0,2 |
| *Mo* | 6 | 0,3 |
| *Bo* | 4 | 0,2 |
| *Tb* | 2 | 0,1 |
| **Total** | **n=20** | **1** |

## 2) Cas du caractère quantitatif :

### a) Variable Statistique Discrète (v.s.d)

–Soit $X$ le caractère qui désigne le nombre d'enfants par ménage pour les *20 femmes* :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 4 | 0 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 4 |
| 2 | 3 | 3 | 5 | 2 | 1 | 3 | 3 | 4 | 5 |

–La distribution est alors :

$$\{(0,1),(1,3),(2,5),(3,6),(4,3),(5,2)\}$$

- Le tableau est alors le suivant :

| $x_i$ | $n_i$ | $f_i$% |
|---|---|---|
| 0 | 1 | 05 |
| 1 | 3 | 15 |
| 2 | 5 | 25 |
| 3 | 6 | 30 |
| 4 | 3 | 15 |
| 5 | 2 | 10 |
| Total | 20 | 100 |

**Question :** *Combien de femmes ont au moins ou au plus 3 enfants ?*

### i) Effectifs et fréquences cumulés :

*) *Effectifs et fréquences cumulés croissants*:

– Soit $N_i$ le $i^{ème}$ effectif cumulé croissant associé à $x_i$

$$N_i = n_1 + n_2 + \cdots + n_i = \sum_{j=1}^{i} n_j$$

$N_i$ est le nombre d'individus présentant au plus la modalité $x_i$.

$$N_4 = n_1 + n_2 + n_3 + n_4 = \sum_{j=1}^{4} n_j = 15$$

On dit que 15 femmes ont au plus $x_4 = 3$ enfants.

En divisant l'égalité ci-dessus par $n=20$, on obtient la **fréquence cumulée croissante:**

$$F_4 = f_1 + f_2 + f_3 + f_4 = \sum_{j=1}^{4} f_j = 75\%$$

On dit que 75% des femmes ont au plus $x_4 = 3$ enfants.

En général, on a :

$$F_i = f_1 + f_2 + \cdots + f_i = \sum_{j=1}^{i} f_j$$

*Effectifs et fréquences cumulés décroissants*:

En sommant cette fois à partir du $i^{ème}$ effectif jusqu'au dernier, on obtient le $i^{ème}$ effectif cumulé décroissant, par exemple :

$$N_{3\searrow} = n_3 + n_4 + n_5 + n_6 = \sum_{j=3}^{6} n_j = 16$$

- Le tableau complet est comme suit :

| $x_i$ | $n_i$ | $N_i$ | $N_{i\searrow}$ | $F_i$ | $F_{i\searrow}$ |
|---|---|---|---|---|---|
| $x_1$ | $n_1$ | $n_1$ | $n$ | $\boxed{f_1}$ | $1$ |
| $x_2$ | $n_2$ | $n_1+n_2$ | $n_2+n_3+\dots+n_k$ | $f_1+f_2$ | $f_2+f_3+\dots+f_k$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $x_i$ | $n_i$ | $n_1+n_2+\dots+n_i$ | $n_i+n_{i+1}+\dots+n_k$ | $f_1+f_2+\dots+f_i$ | $f_i+f_{i+1}+\dots+f_k$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $x_k$ | $n_k$ | $n$ | $n_k$ | $1$ | $f_k$ |
| Total | $n$ | — | — | — | — |

On dit que 16 femmes ont au moins $x_3=2$ enfants.

En divisant l'égalité ci-dessus par $n=20$, on obtient la **fréquence cumulée décroissante:**

$$F_{3\searrow} = f_3+f_4+f_5+f_6=\sum_{j=3}^{6} f_j = 80\%$$

On dit que 80% des ménages ont <u>au moins</u> $x_3=2$ enfants.

- Pour notre exemple, on a :

| $x_i$ | $n_i$ | $f_i\%$ | $N_i$ | $N_{i\searrow}$ | $F_i\%$ | $F_{i\searrow}\%$ |
|---|---|---|---|---|---|---|
| 0 | 1 | 5 | 1 | 20 | 5 | 100 |
| 1 | 3 | 15 | 4 | 19 | 20 | 95 |
| 2 | 5 | 25 | 9 | 16 | 45 | 80 |
| 3 | 6 | 30 | 15 | 11 | 75 | 55 |
| 4 | 3 | 15 | 18 | 5 | 90 | 25 |
| 5 | 2 | 10 | 20 | 2 | 100 | 10 |
| Total | 20 | 100 | | | Au plus | Au moins |

- ***Interprétation*** :

✓Il y a 19 ménages (soit 95%) qui ont **<u>au moins</u>** 1 enfant.

✓Il y a 9 ménages (soit 45%) qui ont **<u>au plus</u>** 2 enfants.

- ***Remarque*** :

- $$F_{i\searrow}= f_i+f_{i+1}+\dots+f_k=\underbrace{\sum_{j=1}^{k} f_j}_{=1}-\sum_{j=1}^{i-1} f_j$$

$$=1-F_{i-1}$$

**<u>b) Variable statistique continue</u>** <u>(v.s.c)</u>

- Dans le cas d'une v.s.c., les modalités sont regroupées en classes. Soit *k* le nombre de ces classes :
- $\underbrace{[e_0, e_1[}_{1^{ère}}$ ; $\underbrace{[e_1, e_2[}_{2^{ème}}$ ; ... ; $\underbrace{[e_{i-1}, e_i[}_{i^{ème}}$ ; ... ; $\underbrace{[e_{k-1}, e_k[}_{k^{ème}}$
- Pour la $i^{ème}$ classe, on note :
- $\boxed{a_i=e_i-e_{i-1}}$ l'amplitude de cette classe.
- $\boxed{x_i=\frac{e_{i-1}+e_i}{2}}$ le centre de cette classe.

| N° classe | Les classes | $n_i$ | $f_i$ |
|---|---|---|---|
| 1 | $[e_0, e_1[$ | $n_1$ | $f_1$ |
| 2 | $[e_1, e_2[$ | $n_2$ | $f_2$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| i | $[e_{i-1}, e_i[$ | $n_i$ | $f_i$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| k | $[e_{k-1}, e_k[$ | $n_k$ | $f_k$ |
| Total | — | $n$ | 1 |

***Exemple*** : on relève le revenu des 20 femmes(€)

| Revenu X | *Effectif* $n_i$ |
|---|---|
| 9000 | 1 |
| 10000 | 2 |
| 15000 | 1 |
| 18000 | 1 |
| 30000 | 1 |
| 40000 | 3 |
| 45000 | 1 |
| 50000 | 1 |
| 55000 | 1 |
| 60000 | 3 |
| 80000 | 2 |
| 85000 | 1 |
| 90000 | 1 |
| 120000 | 1 |
| TOTAL | 20 |

Il est plus commode de regrouper les revenus en classe, par exemple, on choisi 4 classes de même amplitude

| Classes Revenus | effectifs |
|---|---|
| [0 ; 35[ | ● |
| [35 ; 70[ | ● |
| [70 ; 105[ | ● |
| [105 ; 140[ | ? |
| TOTAL | ● |

Complétons notre tableau :

| Classes | $n_i$ | $f_i\%$ | $F_i\%$ | $F_{i\searrow}\%$ |
|---|---|---|---|---|
| [0 ; 35[ | 6 | 30 | 30 | 100 |
| [35 ; 70[ | 9 | 45 | 75 | 70 |
| [70 ; 105[ | 4 | 20 | 95 | 25 |
| [105 ; 140[ | 1 | 5 | 100 | 5 |
| Total | 20 | 100 | Au plus | Au moins |

- ***Interprétation*** :

✓Il y a $F_2=75\%$ des femmes touchent **au plus** $e_2= 70\ m€$. **([$e_1$,$e_2$[)**

✓Il y a $F_{3\searrow}=25\%$ des femmes touchent **au moins** $e_2= 70\ m€$. **([$e_2$,$e_3$[)**

**III) REPRESENTATIONS GRAPHIQUES**

–*Le graphique est une traduction visuelle de l'information qu'elle soit qualitative ou quantitative.*

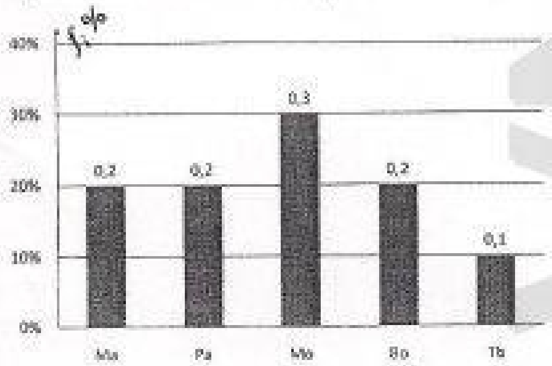
**1) Cas du caractère qualitatif :**

**–A) Le graphique en tuyaux d'orgue**

est formé de rectangles de même base constante et dont les hauteurs sont proportionnelles aux effectifs (ou fréquences) des modalités associées.

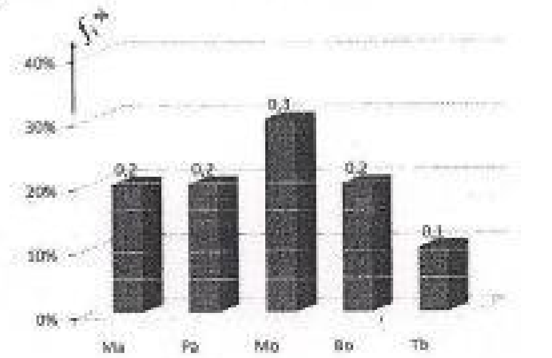
•**Exemple :** On reprend les 20 femmes

**Répartition des 20 femmes selon l'opinion**




•**Exemple :** On reprend les 20 femmes

**Répartition des 20 femmes selon l'opinion**




09/10/2011

### B) Le diagramme circulaire

C'est une représentation en disque. Chaque secteur est proportionnel à l'effectif (ou fréquence) de la modalité associée. En terme d'angle, à $x_i$ on associe l'angle au centre $\alpha_i$, du secteur $i$, vérifiant:

$$\boxed{\alpha_i = c\, f_i \quad ; \quad i=1,\dots,K}$$

*Or*

$$\sum_{i=1}^{K} \alpha_i = c \sum_{i=1}^{K} f_i \Rightarrow 360° = c \times 1 \Rightarrow \boxed{c = 360°}$$

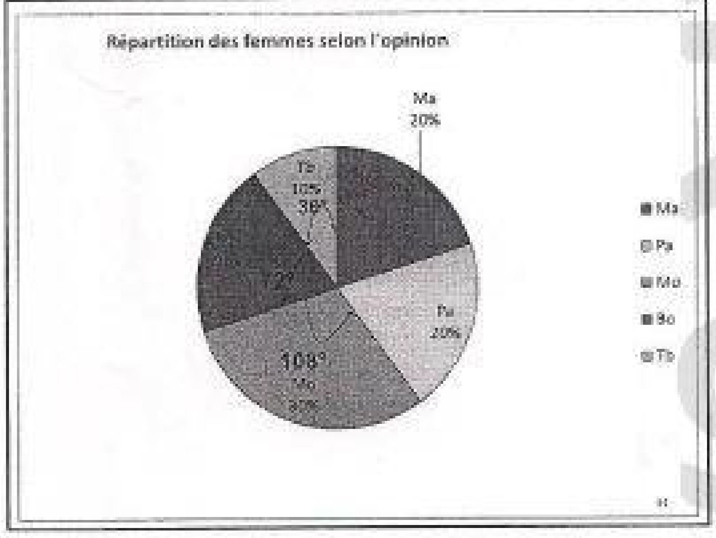
L'égalité devient:

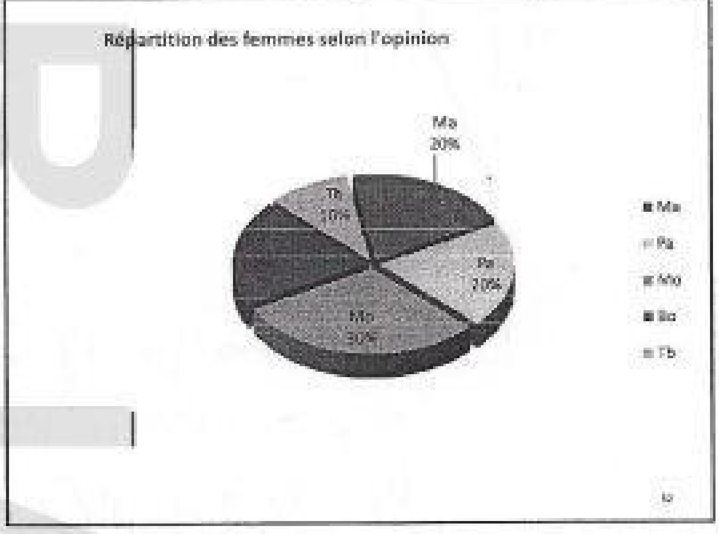
$$\alpha_i = 360 \times f_i \quad ; \quad i=1,\dots,k$$

*Ou*

$$\alpha_i = 3{,}6 \times f_i\,\% \quad ; \quad i=1,\dots,k$$

| $x_i$ | $f_i\,\%$ | $\alpha_i\,°$ |
|---|---|---|
| Ma | 20 | 72 |
| Pa | 20 | 72 |
| Mo | 30 | 108 |
| Bo | 20 | 72 |
| Tb | 10 | 36 |
| Total | 100 | 360 |



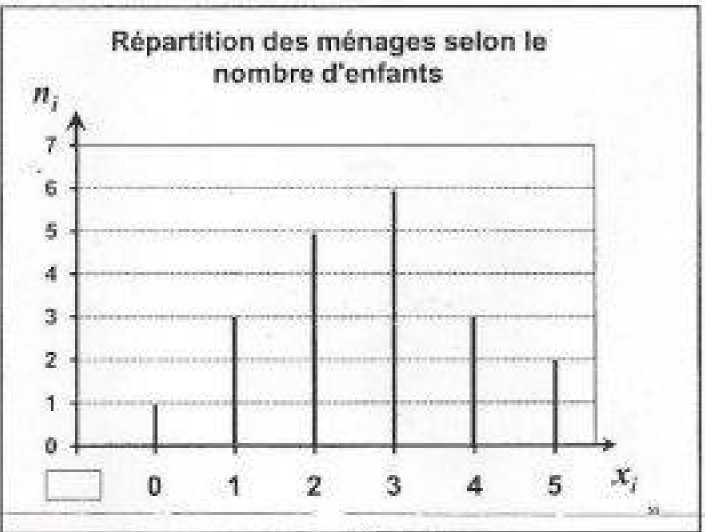

### 2) Cas du caractère quantitatif :

**A) Les V.S.D.** On utilise deux types de graphiques selon que l'on considère les effectifs (ou fréquences) simple ou les effectifs (ou fréquences) cumulés:

***i) Diagramme en bâtons***

À chaque modalité $x_i$ on associe un segment de longueur proportionnelle à l'effectif (ou fréquence).

**Exemple :** On reprend l'exemple des 20 femmes




## –ii) Courbe Cumulative

–**Définition :** *On appelle fonction de répartition* ***F(x)****, la fonction qui à chaque valeur* $x$ *de* $R$ *associe* *<u>la proportion</u> d'individus pour lesquels la valeur de la variable* $X$ *est <u>inférieure ou égale</u> à* $x$.

•***<u>Notation</u>*** **:**

$$\boxed{F(x) = P(X \leq x)}$$

–***Remarques*** :

–Si $x = x_i$ alors $F(x_i) = f_1 + \cdots + f_i = F_i$ .

–Si $x_i \leq x < x_{i+1}$ alors $F(x) = F_i + 0 = F_i$ .

–**Conclusion :**

$F(x) = F_i$ pour tout $x$ tel que $x_i \leq x < x_{i+1}$

*La représentation graphique de* $F(x)$ *est appelée Courbe cumulative, c'est une courbe « en escalier » dont les paliers sont horizontaux, puisque* $F(x)$ *est constante sur chaque intervalle* $[x_i , x_{i+1}[$.

–***Propriété de*** $F$:

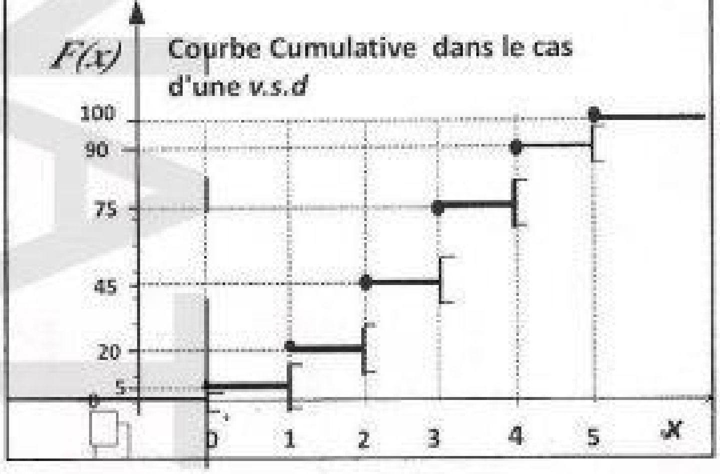
- La fonction $F$ est définie sur $R$ et à valeurs dans $[0 , 1]$.
- $F(x) = 0$ si $x < x_1$
- $F(x) = 1$ si $x \geq x_k$
- $F(-\infty) = 0$ et $F(+\infty) = 1$
- $F$ est constante sur chaque intervalle séparant deux modalités consécutives .

•***Remarque :*** On obtient la courbe cumulative des effectifs en remplaçant les $F_i$ par les $N_i$.

- Pour notre exemple, on a :

| $x_i$ | $n_i$ | $f_i$% | $F_i$% |
|---|---|---|---|
| 0 | 1 | 5 | 5 |
| 1 | 3 | 15 | 20 |
| 2 | 5 | 25 | 45 |
| 3 | 6 | 30 | 75 |
| 4 | 3 | 15 | 90 |
| 5 | 2 | 10 | 100 |
| Total | 20 | 100 | Au plus |

**Exemple :** On reprend les 20 femmes



–B) **Les V.S.C.** *On a souvent recourt à trois types de graphiques :*

### –i) *Histogramme*

–À chaque classe, on associe un rectangle dont la base est égale à l'amplitude de la classe et dont la hauteur est de telle sorte que sa surface ($S_i = b_i \times h_i$) soit proportionnelle à la fréquence de la classe. La juxtaposition de tous ces rectangles forment un histogramme.

La procédure à suivre :

– Toutes les classes ont même amplitude ($a_i = cte = a$ ; $i=1,\ldots,k$) alors $h_i = f_i$ (ou $n_i$).

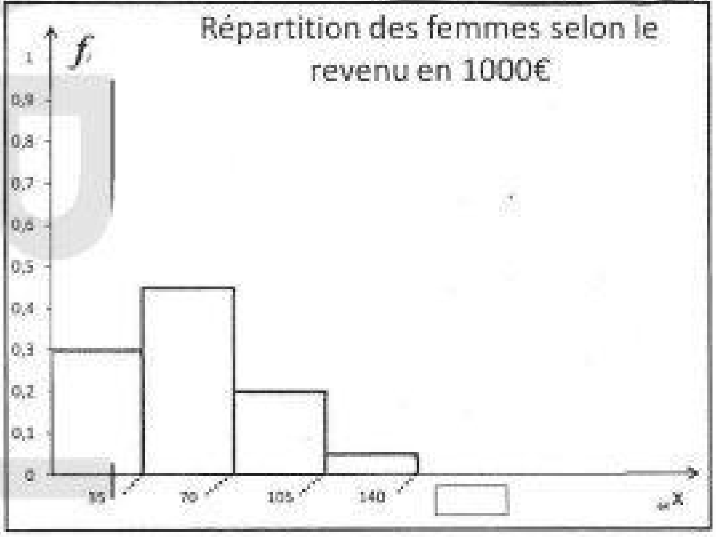
– Au moins une classe a une amplitude différente des autres : dans ce cas on choisi une amplitude de référence $a_r$ (par exemple la plus petite ou la plus répandu) Par suite, on corrige la fréquence des classes différentes en la divisant par l'amplitude associée et en multipliant par $a_r$ :

$$\boxed{h_i = f_i' = \frac{f_i}{a_i} \times a_r}$$

$$\boxed{h_i = a_r \times \frac{f_i}{a_i}}$$

**Exemple :** On reprend les 20 femmes
Toutes les classes ont même amplitude donc pas besoin de corriger les fréquences

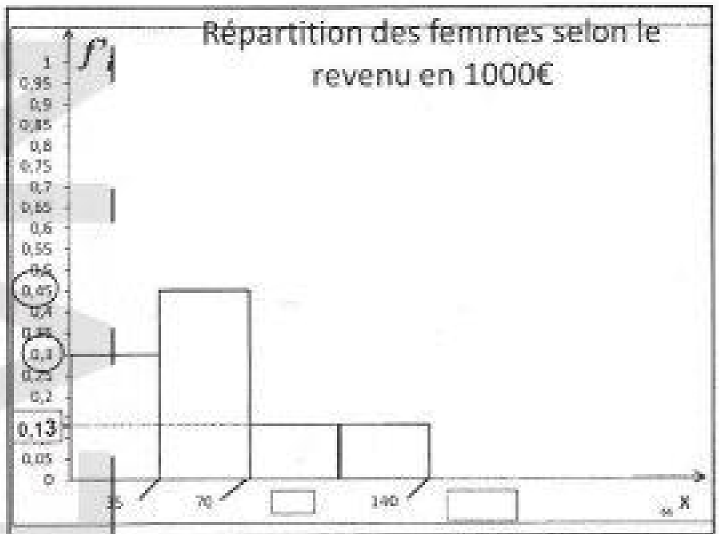
| *Classes* | $a_i$ | $f_i$ |
|---|---|---|
| [0 ; 35[ | 35 | 0,30 |
| [35 ; 70[ | 35 | 0,45 |
| [70 ; 105[ | 35 | 0,20 |
| [105 ; 140[ | 35 | 0,05 |
| Total | $a_r$=35 | 1 |




**Exemple : Si** On regroupe les 2 dernières classes

| *Classes* | $f_i$ | $a_i$ | $l_i$ | $f_i'$ |
|---|---|---|---|---|
| [0 ; 35[ | 0,30 | 35 | 1 | 0,30 |
| [35 ; 70[ | 0,45 | 35 | 1 | 0,45 |
| [70 ; 140[ | 0,25 | 70 | 2 | 0,13 |
| Total | 1 | $a_r$=35 | --- | -- |

Rem : $f_i'$ s'appelle *densité de fréquence*
(sans pourcentage)

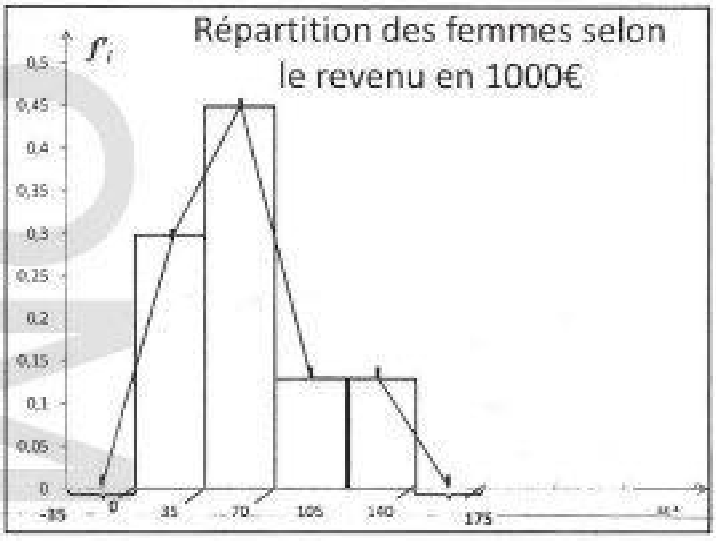


### –ii) Polygone de fréquences

– On subdivise l'histogramme en sous rectangles de même base égale à l'amplitude de référence $a_r$. $a_r$ étant choisie comme la **plus petite** des amplitudes et **vérifiant** :

$$a_i = ka_r \;;\quad k = 1,2,3,4,...$$

–Après avoir ajouter aux extrémités de l'histogramme deux rectangles fictifs de hauteur nulle et de base $a_r$ ,on joint, par des segments de droites, les milieux des sommets des sous rectangles ainsi obtenus.



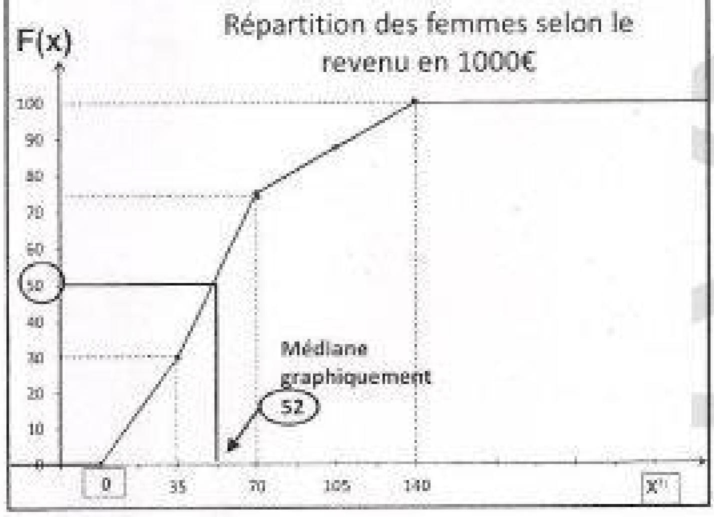

### –ii) Courbe Cumulative

–On construit la courbe de la fréquence cumulée en joignant les points $(e_i, F_i)$, où $e_i$ est la borne supérieure de la $i^{ème}$ classe $[e_{i-1}, e_i[$ et $F_i$ est la fréquence cumulée de cette même classe. On note

$$F_i = P(X \le e_i)$$

**Exemple :** On reprend les 20 femmes

| Classes | $f_i$% | $F_i$% |
|---|---|---|
| [0 ; 35[ | 30 | 30 |
| [35 ; 70[ | 45 | 75 |
| [70 ; 140[ | 25 | 100 |
| Total | 1 | Au plus |




# *Chapitre* II

- *Les caractéristiques de Tendance centrale*

- *Ici, il s'agit de faire une synthèse de l'information, contenue dans la série brute, par le chiffre; et ce en calculant des paramètres dits de tendance centrale, qui caractérisent l'ordre de grandeur des observations.*
- *Dans ce chapitre, on analysera trois de ces paramètres qui sont :*
  *les moyennes, le mode et la médiane.*

## I) LES MOYENNES

- **(1) *La moyenne arithmétique***

- **(a) *Définition* :** *La moyenne arithmétique, notée $\bar{x}$, d'une variable statistique X de distribution $\{(x_i, n_i)_{1\le i\le k}\}$ est la quantité :*

$$\boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} f_i x_i}$$

- Où, $n$ est la taille de la population, et les $x_i$ sont les modalités dans le cas d'une v.s.d. et les centres des classes dans le cas d'une v.s.c.

- **Exemple 1 :** On reprend l'exemple des 20 femmes selon le nb d'enfants

| $x_i$ | $n_i$ | $n_i x_i$ | $f_i$ | $f_i x_i$ |
|---|---|---|---|---|
| 0 | 1 | 0 | 0,05 | 0 |
| 1 | 3 | 03 | 0,15 | 0,15 |
| 2 | 5 | 10 | 0,25 | 0,50 |
| 3 | 6 | 18 | 0,30 | 0,90 |
| 4 | 3 | 12 | 0,15 | 0,60 |
| 5 | 2 | 10 | 0,10 | 0,50 |
| Total | 20 | 53 | 1 | 2,65 |

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{6} n_i x_i = \frac{53}{20} = 2{,}65$$

$$ou\ \ \bar{x} = \sum_{i=1}^{6} f_i x_i = 2{,}65$$

**Exemple 2 :** Pour les revenus des femmes

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{3} n_i x_i = \sum_{i=1}^{3} f_i x_i = 55125€$$

| Classes | $f_i$ | $x_i$ | $f_i x_i$ |
|---|---|---|---|
| [0 ; 35[ | 0,30 | 17,5 | 5,25 |
| [35 ; 70[ | 0,45 | 52,5 | 23,625 |
| [70 ; 140[ | 0,25 | 105 | 26,25 |
| Total | 1 | | 55,125 |

### (b) *Changement d'origine et d'échelle*

- ***Propriété :*** Soit $X$ une variable statistique de moyenne arithmétique $\overline{x}$. Si $Y$ est une variable statistique telle que $Y=aX+b$, où $a$ et $b$ sont des réels quelconques, alors la moyenne arithmétique de $Y$ est :

$$\boxed{\overline{y} = a\overline{x} + b}$$

Démonstration :

$$\overline{y} = \sum_{i=1}^{k} f_i y_i = \sum_{i=1}^{k} f_i (ax_i + b)$$

$$= a\underbrace{\sum_{i=1}^{k} f_i x_i}_{\overline{x}} + b\underbrace{\sum_{i=1}^{k} f_i}_{1} = a\overline{x} + b$$

### • (C) *Propriétés algébriques de la moyenne arithmétique*

- *i)* la moyenne des écarts à la moyenne arithmétique est nulle :

$$\sum_{i=1}^{k} f_i (x_i - \overline{x}) = 0$$

- ***ii)*** La moyenne des carrés des écarts à une constante a est minimale pour $a = \overline{x}$

$$\min_{a\in ◆} \sum_{i=1}^{k} f_i (x_i - a)^2 = \sum_{i=1}^{k} f_i (x_i - \overline{x})^2$$

**Démonstration :**

$$\sum_{i=1}^{k} f_i (x_i - a)^2 = \sum_{i=1}^{k} f_i \left[(x_i - \overline{x}) + (\overline{x} - a)\right]^2$$

$$= \sum_{i=1}^{k} f_i (x_i - \overline{x})^2 + \sum_{i=1}^{k} f_i (\overline{x} - a)^2$$

$$+ 2\sum_{i=1}^{k} f_i (x_i - \overline{x})(\overline{x} - a)$$

$$= \sum_{i=1}^{k} f_i (x_i - \overline{x})^2 + (\overline{x} - a)^2 \underbrace{\sum_{i=1}^{k} f_i}_{=1}$$

$$+ 2(\overline{x} - a)\underbrace{\sum_{i=1}^{k} f_i (x_i - \overline{x})}_{=0}$$

$$= \sum_{i=1}^{k} f_i (x_i - \overline{x})^2 + (\overline{x} - a)^2$$

### (d) *Propriété de l'agrégation*

- *Soit une population* P *de taille $n$, composée de m sous populations* $P_1, P_2, \ldots, P_m$ ; *de tailles respectives* $n_1, n_2, \ldots n_m$ *et de moyennes respectives* $\overline{x}_1, \overline{x}_2, \ldots \overline{x}_m$. *Alors la moyenne arithmétique* $\overline{x}$ *de la population* P *est donnée par :*

$$\boxed{\overline{x} = \frac{1}{n}\sum_{i=1}^{m} n_i \overline{x}_i}$$

- **Exemple :**

Le salaire moyen des cadres dans l'entreprise E est de 4000 DH.

- Le salaire moyen des cadres masculins est de 4200 DH.
- Le salaire moyen des cadres féminins est de 3000 DH.

1. Quelle est la répartition hommes - femmes des cadres ?