جامعة السلطان مولاي سليمان
Université Sultan Moulay Slimane

الكلية المتعددة التخصصات - بني ملال
FACULTE POLYDISCIPLINAIRE - BENI MELLAL

Statistique descriptive

Pr. FALLOUL My El Mehdi

Semestre 1

Année universitaire: 2016/2017

Objectifs du cours

- **Familiariser l'étudiant avec les principaux outils de la statistique descriptive et lui fournir les outils et techniques nécessaires pour résoudre les problèmes statistiques requérant la connaissance des mathématiques.**

Plan du cours

1. Introduction générale
2. Les tableaux statistiques
3. Les représentations graphiques
4. La réduction des données
5. Réduction des données (séries bidimensionnelles)
6. Les indices statistiques
7. Notion d'ajustement et de corrélation
8. Notion des séries chronologiques

Chapitre 1

Introduction générale

1.1 Définition

- La statistique (est une science) désigne l'ensemble des **méthodes scientifiques** qui permettent **d'analyser quantitativement un ensemble d'informations cohérent**.

- La statistique au sens large comprend deux branches :

✓ La **statistique descriptive,** qui est un ensemble de méthodes permettant de **décrire** et d'**analyser quantitativement des unités statistiques** d'une population**.**

✓ La **statistique mathématique (inférentielle)** dont l'objet est de **formuler des lois** à partir de l'observation d'échantillons, c'est-à-dire de tirages limités effectués au sein d'une population**. Elle intervient dans les enquêtes et les sondages.** Elle s'appuie sur la statistique descriptive, mais aussi sur le calcul des probabilités.

1.2 Vocabulaire

- **Population statistique:** est l'ensemble des éléments sur lesquels porte l'étude. Les éléments de la population sont appelés **individus statistiques (ou unités statistiques**). Un **échantillon** de taille n est un sous-ensemble formé de n individus de la population (n <=N).

- **Enquête:** c'est l'ensemble des opérations qui ont pour but de collecter de façon organisée des informations relatives à une population.

- **Recensement:** Lors s'une enquête, si toutes les unités statistiques de la population sont observées individuellement, l'enquête est dite complète ou exhaustive. On parle aussi de recensement (exemple recensement de la population d'un pays).

- **Sondage:** c'est l'étude d'une partie de la population, c'est ce qu'on appelle enquête partielle ou par échantillonnage(ou **sondage**).

1.2 Vocabulaire (suite)

- **Variable statistique ou caractère**: est la chose que l'on étudie et qui est commune à tous les individus de la population de référence. L'ensemble des résultats s'appelle **série statistique**.

- **Caractère qualitatif (non mesurable)**: c'est lorsque les modalités d'un caractère ne peuvent pas s'exprimer par des nombres. Exe(le caractère couleur a pour modalité: vert, jaune, bleu).

- On parle de **caractère qualitatif ordinal** quand les modalités peuvent être ordonnées (hiérarchisées). Exemples : niveau d'études, classe sociale, grade, etc. Dans le cas contraire, on parle de **caractère qualitatif nominal**. (Ex: couleur des yeux, état civil, sexe, lieu de résidence, etc.)

1.2 Vocabulaire (suite 1)

- Une variable quantitative est dite **discrète** si elle ne prend que des valeurs isolées. (**exe**: Un âge, une note arrondie etc). Une variable quantitative est **dite continue** si elle peut prendre toutes les valeurs comprises entre 2 nombres(exe: distance entre la maison et la faculté),

- **Modalités:** chaque caractère (variable statistique) possède deux ou plusieurs modalités. Ce sont les différentes situations ou les unités statistiques peuvent se trouver à l'égard du caractère considéré. (exe1: le caractère **« nationalité »** peut avoir comme **modalités: marocaine, algérienne, tunisienne**. Exe2: le caractère **« nombre d'enfants par famille »** peut avoir comme **modalités; 0,1,2**).

- **Nomenclature:** c'est la liste de toutes les modalités dans un caractère qualitatif, exemple (sexe: féminin, masculin)

1.2 Vocabulaire (suite 1)

Population

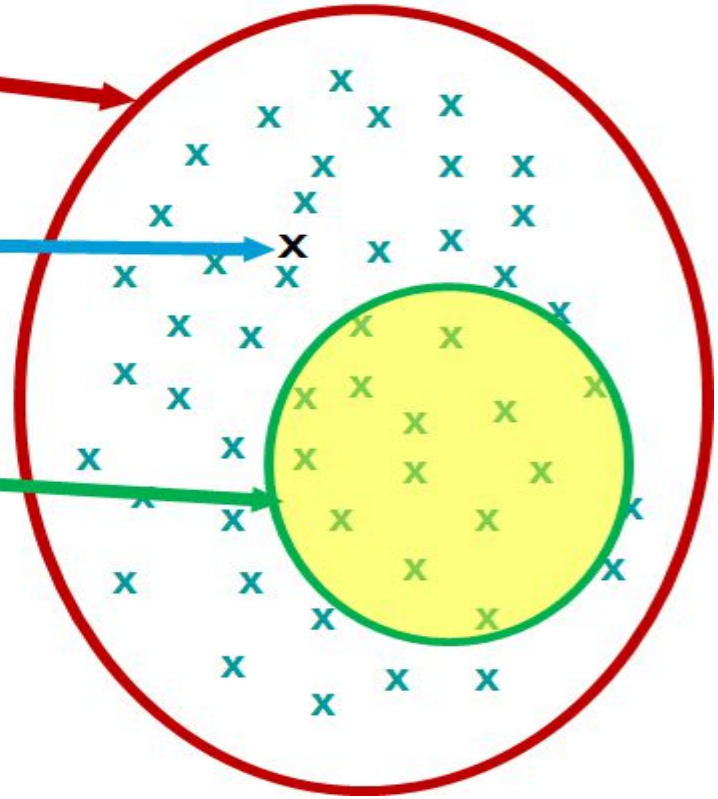
Ensemble de référence

Individu

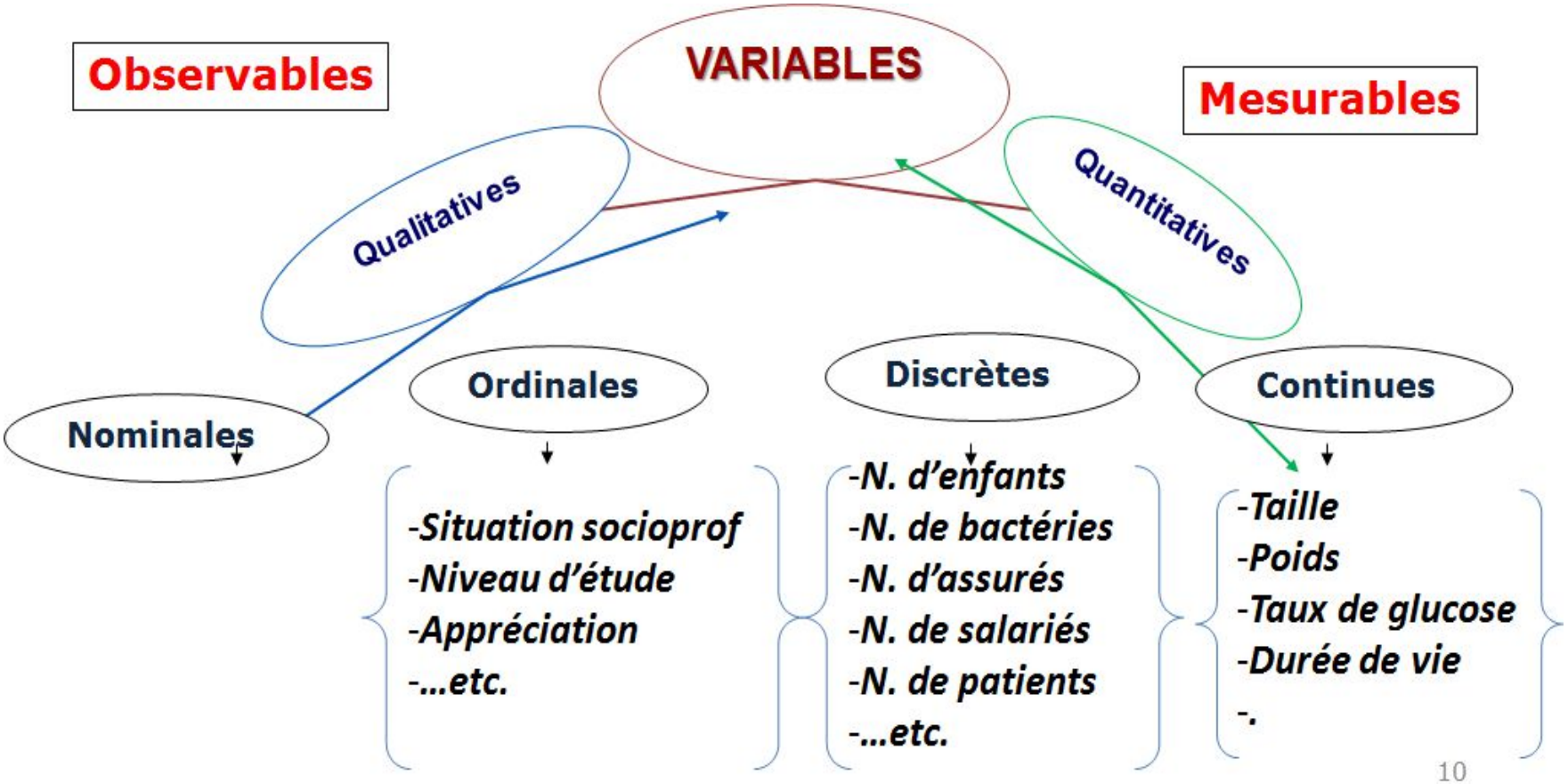
Elément de la population

Echantillon

Sous-ensemble de la population.



1.2 Vocabulaire (suite 2)



1.3 Indices de sommation

- ***Soit***

1+2+3+4+...+n on peut écrire $\sum_{i=1}^{n} i = \sum_{1 \le i \le n} i$

$$\sum_{1 \le i \le n} a = \sum_{i=1}^{n} a = na \quad ; \quad \sum_{0 \le i \le n} a = \sum_{i=0}^{n} a = (n+1)a$$

- **Règles de calcul**

$$\sum_{i=1}^{n} (a.u_i) = a \sum_{i=1}^{n} u_i$$

$$\sum_{i=1}^{n} (u_i + v_i) = \sum_{i=1}^{n} u_i + \sum_{i=1}^{n} v_i$$

$$(\sum_{i=1}^{n} u_i).(\sum_{j=1}^{p} v_j) = \sum_{\substack{1 \le i \le n \\ 1 \le j \le p}} (u_i.v_j)$$

1.4 indice du produit

- ***Soit***

$$\underbrace{a \times a \times \ldots \times a}_{n \ \ facteurs}$$, on peut écrire $$\prod_{1 \le i \le n} a = \prod_{i=1}^{n} a = a^n$$

$$1 \times 2 \times 3 \times \ldots \times n$$, on peut écrire $$\prod_{i=1}^{n} i = \prod_{1 \le i \le n} i = n!$$

- **Règles de calcul**

$$\prod_{i=1}^{n} a = a^n \ ; \ \prod_{i=0}^{n} a = a^{n+1}$$

$$\prod_{i=1}^{n} (u_i . v_i) = (\prod_{i=1}^{n} u_i).(\prod_{i=1}^{n} v_i)$$

1.3 Exemples

- **Exemple 1:** On a procédé au recensement des 50 salariés de la société STM en relevant les salaires horaires perçus:

✓ **Unité statistique:** un salarié de la société STM
✓ **Population:** l'ensemble des 50 salariés de la société STM
✓ **Caractère:** le salaire horaire
✓ **Type de caractère:** caractère quantitatif ou variable statistique

- **Exemple 2:** une enquête sur la nationalité des touristes visitant le Maroc a concerné un échantillon de 500 touristes.

✓ **Unité statistique:** un touriste
✓ **Population:** l'ensemble des touriste visitant le Maroc
✓ **Caractère:** nationalité
✓ **Type de caractère:** qualitatif

Chapitre 2: Les tableaux statistiques

2.1 Données brutes

- on appelle données brutes des données qu'on rassemble sans se soucier de la notion d'ordre.

Exemple : On a procédé au recensement des 50 salariés de la société STM en relevant les salaires horaires perçus:

34	36	45	62	43	63	26	55	57	61
87	78	77	75	74	25	15	18	20	44
96	94	103	110	88	116	125	47	85	74
14	19	17	87	92	88	75	48	95	74
45	48	98	75	45	74	85	75	47	26

2.2 Série statistique

- Une série statistique est une simple énumération des observations

$$x_1, x_2, x_3, \ldots\ldots, x_i, \ldots\ldots, x_n$$

- Ces observations étant rangées par ordre croissant:

$$x_1 \leq x_2 \leq x_3 \leq \ldots\ldots \leq x_i \leq \ldots\ldots \leq x_n$$

- **n** est le nombre total des observations appelé aussi effectif. Une même observation peut se répéter plusieurs fois. La différence entre la valeur la plus grande et la valeur la plus petite est appelée étendue.

$$Etendue = X\max - X\min$$

2.3 Distribution des fréquences
2.3.1 Cas d'un caractère qualitatif

- Lorsque les observations sont nombreuses, il est nécessaire de les condenser sous forme d'un tableau statistique appelé distribution des fréquences sous la forme suivante:

Modalités	**Fréquences absolues *ni***	**Fréquences relatives *fi***
M1 M2 . . Mi . . Mk	n1 n2 . . ni . . nk	f1 f2 . . fi . . fk
Total	**n**	**1**

- **Règles de calcul:**

$$n_1 + n_2 + \ldots\ldots + n_k = \sum_{i}^{k} n_i = n$$

$f_i = \frac{n_i}{n}$ est la proportion de la modalité Mi. Elle est dite la fréquence relative , elle est souvent exprimé en pourcentage.

$$f_i = \frac{n_i}{n} * 100\%$$

La somme des fréquences relatives est toujours égale à 1

$$\sum_{i=1}^{k} f_i = f_1 + f_2 + \ldots. + f_i + \ldots + f_k = \frac{n_1 + n_2 + \ldots + n_i + \ldots n_k}{n} = \frac{n}{n} = 1$$

<u>Exemple 1</u>: une enquête sur la nationalité des touristes visitant le Maroc a concerné un échantillon de 500 touristes.

Les résultats sont condensés dans **la distribution des fréquences suivantes**:

Nationalité	**Nombre de touristes (fréquence absolue)**	**Pourcentage des touristes (fréquence relative)**
Française	85	17%
Allemande	106	21,2%
Italienne	62	12,4%
Hollandaise	44	8,8%
Belge	40	8%
Américaine	70	14,0%
Autres nationalités	93	18,6%
Total	**500**	**100%**

2.3 Distribution des fréquences
2.3.2 Cas d'une variable statistique discrète

- On définit les fréquences absolues **(ni)** et les fréquences relatives fi. En plus, on peut ajouter les effectifs cumulés croissant **(ECC)** et décroissants **(ECD),** fréquences cumulés croissants **(FCC)** et décroissant **(FCD).**

valeurs de la variable Xi	**Effectif ni**	**Fréquences relatives fi**	**ECC**	**ECD**	**FCC**	**FCD**
V1	n1	f1	n1	n	f1	1
V2	n2	f2	n1+n2	n-n1	f1+f2	1-f1
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Vi	ni	fi	n1+n2..+ni	n-n1..-ni-1	f1+f2..+fi	1-f1-f2..-fi-1
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Vk	nk	fk	n	nk	1	fk
Total	**n**	**1**				

- Une enquête chez 1000 commerçants porte sur le nombre d'employés

nombre d'employés	Nombre de commerçants ni	Fréquences relatives fi	ECC	ECD	FCC	FCD
0	50	5%	50	1000	5%	100%
1	100	10%	150	950	150%	95%
2	200	20%	350	850	625%	85%
3	150	15%	500	650	35%	65%
4	120	12%	620	500	%	50%
5	160	16%	780	380	78%	38%
6	130	13%	910	220	91%	22%
7	90	9%	1000	90	100%	9%
Total	**1000**	**100%**				

- Le nombre de commerçants employant au plus 5 employés (au maximum 5 employés) est 780 soit 78% des commerçants.
- Le nombre de commerçants employant au moins 3 employés (au minimum 3 employés) est 650 soit 65% des commerçants.

2.3 Distribution des fréquences
2.3.3 Cas d'une variable statistique continu

- Méthode pour construire une distribution de fréquence à partir des données brutes :

1) On calcule l'étendue de la série;
2) On partage l'étendue en classes de même amplitude (lorsque c'est impossible, on considère des classes d'amplitudes variables), en pratique le nombre de classes est entre 5 et 20 (selon les données), de préférence (les points centraux doivent coïncider avec les données observées);
3) Pour chaque classe on détermine le nombre d'observations (comptage des données appartenant à chaque classe;

Exemple:

On a mesuré le poids en kilogramme de 80 personnes.

➢ Les données brutes sont comme suit :

68	84	75	82	68	90	62	88	76	93
73	79	88	73	60	93	71	59	85	75
61	65	75	87	74	62	95	78	63	72
66	78	82	75	94	77	69	74	68	60
96	78	89	61	75	95	60	79	83	71
79	62	67	97	78	85	76	65	71	75
65	80	73	57	88	78	62	76	53	74
86	67	73	81	72	63	76	75	85	77

- La plus grande valeur est : 97
- La plus petite valeur est : 53
- L'étendue est : 97-53= 44
- En fixant le nombre de classes à 10, l'amplitude constante des classes est 44/10= 4,4 soit équivaut à une amplitude de 4

Poids	Point central	Fréquences absolues ni	Fréquences absolues cumulés		Fréquences relatives fi	Fréquences relatives cumulées	
			C	D		C	D
50-54	52	1	1	80	1.25%	1.25%	100%
55-59	57	2	3	79	2.5%	3.75%	98.75%
60-64	62	11	14	77	13.75%	17.5%	96.25%
65-69	67	10	24	66	12.5%	30%	82.5%
70-74	72	12	36	56	15%	45%	70%
75-79	77	21	57	44	26.25%	71.25%	55%
80-84	82	6	63	23	7.5%	78.75%	28.75%
85-89	87	9	72	17	11.25%	90%	21.25%
90-94	92	4	76	8	5%	95%	10%
95-99	97	4	80	4	5%	100%	5%
Total		**80**			**100%**		

- Le nombre de personnes pesant au moins 70 kg (au moins 70 kg ou plus de 69 kg) est 56 soit 70% des personnes pesées;
- Le nombre de personnes pesant au plus 84kg (au maximum 84 ou moins de 85 kg) est 63 soit 78,75% des personnes pesées.

2.4 Distribution des fréquences à deux variables
2.4.1 Série statistique double

- La statistique descriptive à 2 dimensions a pour **but de caractériser les relations** qui existent entre **deux séries d'observations considérées simultanément** qui peuvent être de nature (qualitative, quantitative, continue ou discontinue).

- Une série statistique double est une simple énumération des observations de deux variables statistiques X et Y.

$$x_1, x_2, x_3, \ldots\ldots, x_i, \ldots\ldots, x_n$$
$$y_1, y_2, y_3, \ldots\ldots, y_i, \ldots\ldots, y_n$$

- Les observations correspondent à un couple de valeurs (xi,yi)
- **n** est le nombre total d'observations, appelé aussi effectif

Exemple

- On a relevé les notes de mathématiques et de statistiques obtenus par 12 étudiants dans un examen final.

✓ **Unité statistique**: un étudiant;

✓ **Population**: 12 étudiants;

✓ **Premier caractère**: note de mathématique, c'est une variable statistique continue;

✓ **Deuxième caractère**: note de statistique, c'est une variable statistique continue.

➢ La série statistique double est

Numéro étudiant	1	2	3	4	5	6	7	8	9	10	11	12
Note Math	11	14	9	12	10	6	15	12	10	10	8	13
Note statistiques	10	15	11	11	9	8	14	13	11	12	10	12

2.4.2 Tableaux de contingence

x y	y1	y2	y3	…	yj	…	yp	ni.
x1	n11	n12	n13	…	n1j	…	n1p	n1.
x2	n21	n22	n23	…	n2j	…	n2p	n2.
x3	n31	n32	n33	…	n3j	…	n3p	n3.
…	…	…	…	…	…	…	…	...
Xi	ni1	ni2	ni3	…	nij	…	nip	ni.
…	…	…	…	…	…	…	…	…
xk	nk1	nik	nik	…	nij	…	nip	nk.
n.j	n.1	n.2	n.3	…	n.j	…	n.p	n

- nij est le nombre d'individus qui représente à la fois la modalité xi et la modalité xj, c'est la fréquence absolue conjointe:

$$n = \sum_{i=1}^{k} \sum_{j=1}^{p} n_{ij}$$

- Fij est la fréquence relative conjointe, c'est la proportion des individus qui représente à la fois la modalité xi et

2.4.2 Tableaux de contingence

- ***fij*** est la proportion d'individus qui représente à la fois la modalité xi et la modalité xj, c'est la fréquence relative conjointe:

$$f_{ij} = \frac{n_{ij}}{n}$$

- ***ni.*** est le nombre d'individus qui possèdent la modalité X quelque soit la modalité de Y:

$$n_{i.} = n_{i1} + n_{i2} + \ldots + n_{ij} + \ldots + n_{ip} = \sum_{j=1}^{p} nij$$

- ***n.j*** est le nombre d'individus qui possèdent la modalité Y quelque soit la modalité X.

$$n_{.j} = n_{1j} + n_{2j} + \ldots + n_{ij} + \ldots n_{kj} = \sum_{i=1}^{k} n_{ij}$$

$$n = \sum_{i=1}^{k} \sum_{j=1}^{p} n_{ij} = \sum_{i=1}^{k} n_{i.} = \sum_{j=1}^{p} n_{.j}$$

Exemple

La répartition de 300 salariés d'une entreprise selon l'âge et la situation familiale est représentée dans le tableau de contingence suivant:

Age	**20 - 30**		**30-40**		**40-50**		**50-60**		**Total**	
Situation familiale	***nij***	***fij %***	***nij***	***fij%***	***nij***	***fij%***	***nij***	***fij%***	***nij***	***fij%***
Célibataire	**38**	**12,7**	31	10,3	10	3,3	5	1,7	**84**	28
Marié	13	4,3	35	11,7	56	18,7	41	13,6	**145**	48,3
Divorcé	8	2,7	12	4	8	2,7	8	2,7	**36**	12,1
Veuf	4	1,3	6	2	13	4,3	12	4	**35**	11,6
Total	**63**	**21**	**84**	**28**	**87**	**29**	**66**	**22**	300	100

- Parmi les **300 salariés** de cette entreprise, il y a **38 salariés** célibataires **20 à 30 ans**, soit une proportion de 12,7% de l'ensemble des salariés de cette entreprise.

2.4.3 Distributions marginales

- La distribution marginale **du caractère X** est la distribution à une dimension des individus de la population qui présentent une **modalité de X quelque soit la modalité de Y**.
- La distribution marginale **du caractère Y** est la distribution à une dimension des individus de la population qui présentent une **modalité de Y quelque soit la modalité de X.**

distribution marginale de x

X	ni.	fi.
X1	n1.	f1.
X2	n2.	f2.
X3	n3.	f3.
…	…	…
Xi	ni.	fi.
…	…	…
Xk	nk.	fk.
	n	1

distribution marginale de y

X	n.j	fi.
Y1	n.1	f1.
Y2	n.2	f2.
Y3	n.3	f3.
…	…	…
Yj	n.j	fj.
…	…	…
Yp	n.p	f.p
	n	1

Exemple 1

- La distribution marginale de la situation familiale

Situation familiale	**nij**	**fij**
Célibataire	84	28%
Marié	145	48,3%
Divorcé	36	12,1%
Veuf	35	11,6%
Total	**300**	**100%**

➢ Parmi les 300 salariés de l'entreprise: 145 sont mariés, 84 sont célibataires, 36 sont divorcés, et 35 sont veufs, ce qui représentent respectivement, 48,3%, 28%, 12,1%, et 11, 6 % sont veufs.

Exemple 2

- La distribution marginale de l'âge des salariés:

Situation familiale	nij	fij
20 à 30 ans	63	21%
30 à 40 ans	84	28%
40 à 50 ans	87	29%
50 à 60 ans	66	22%
Total	**63**	**100%**

- 63 salariés de l'entreprise sont âgé entre 20 et 30 ans : 84 sont âgés entre 30 et 40 ans, 87 sont âgés entre 40 et 50 ans et 66 sont âgé entre 50 et 60 ans avec des proportions respectives de 21%, 28%, 29% et 22%.

2.4.4 Distributions conditionnelles

- La distribution marginale **du caractère X** est la distribution à une dimension des individus de la population qui présentent une **modalité de X quelque soit la modalité de Y**.

- *distribution conditionnelle de X/ Y=yi* *distribution conditionnelle de y/X=xi*

X	ni.	fi.
X1	n1.	f1.
X2	n2.	f2.
X3	n3.	f3.
…	…	…
Xi	ni.	fi.
…	…	…
Xk	nk.	fk.
	n	1

X	n.j	fi.
Y1	n.1	f1.
Y2	n.2	f2.
Y3	n.3	f3.
…	…	…
Yj	n.j	fj.
…	…	…
Yp	n.p	f.p
	n	1

Exemple 1

- La distribution conditionnelle de la situation familiale des salariés âgés entre 20 a 30 ans:

Situation familiale	**nij**	**fij**
Célibataire	38	60%
Marié	13	21%
Divorcé	8	13%
Veuf	4	6%
Total	**63**	**100%**

➢ Parmi les 63 salariés de l'entreprise âgé de 20 à 30 ans : 13 sont mariés, 8 sont divorcés, et 4 sont veufs, ce qui représentent respectivement, 60%, 21%, 13%, et 6 % des salariés âgés de 20 à 30 ans.

Exemple 2

- La distribution conditionnelle de l'âge des couples mariés

Age	**nij**	**fij**
20 à 30 ans	13	9%
30 à 40 ans	35	24%
40 à 50 ans	56	39%
50 à 60 ans	41	28%
Total	**145**	**100%**

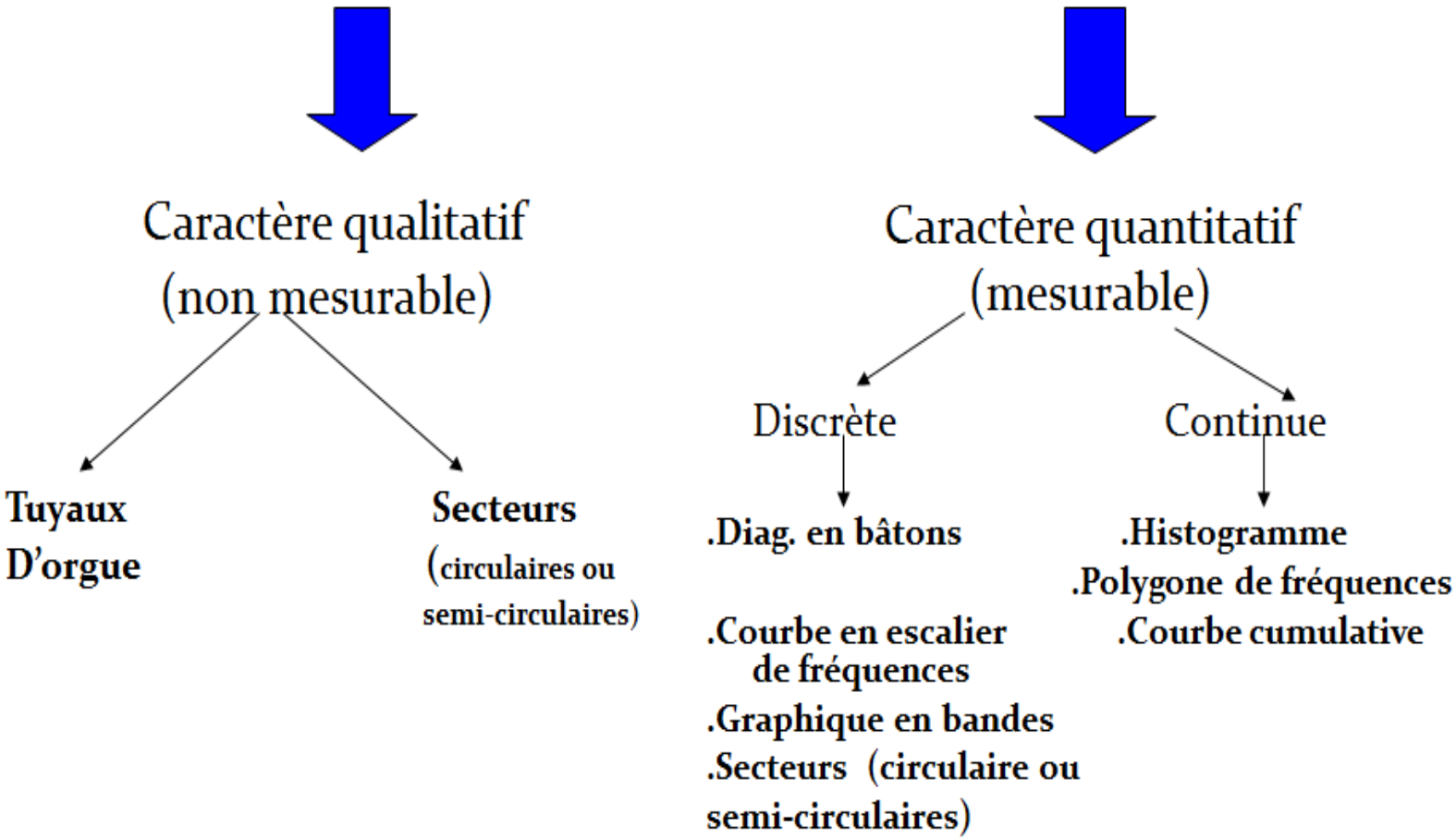
➢ 13 salariés mariés de cette entreprise sont âgé entre 20 et 30 ce qui représente une proportion de 9%, 35 salariés mariés sont âgé entre 30 et 40 ans ce qui représente une proportion de 24%.

Chapitre 3

Les représentations graphiques

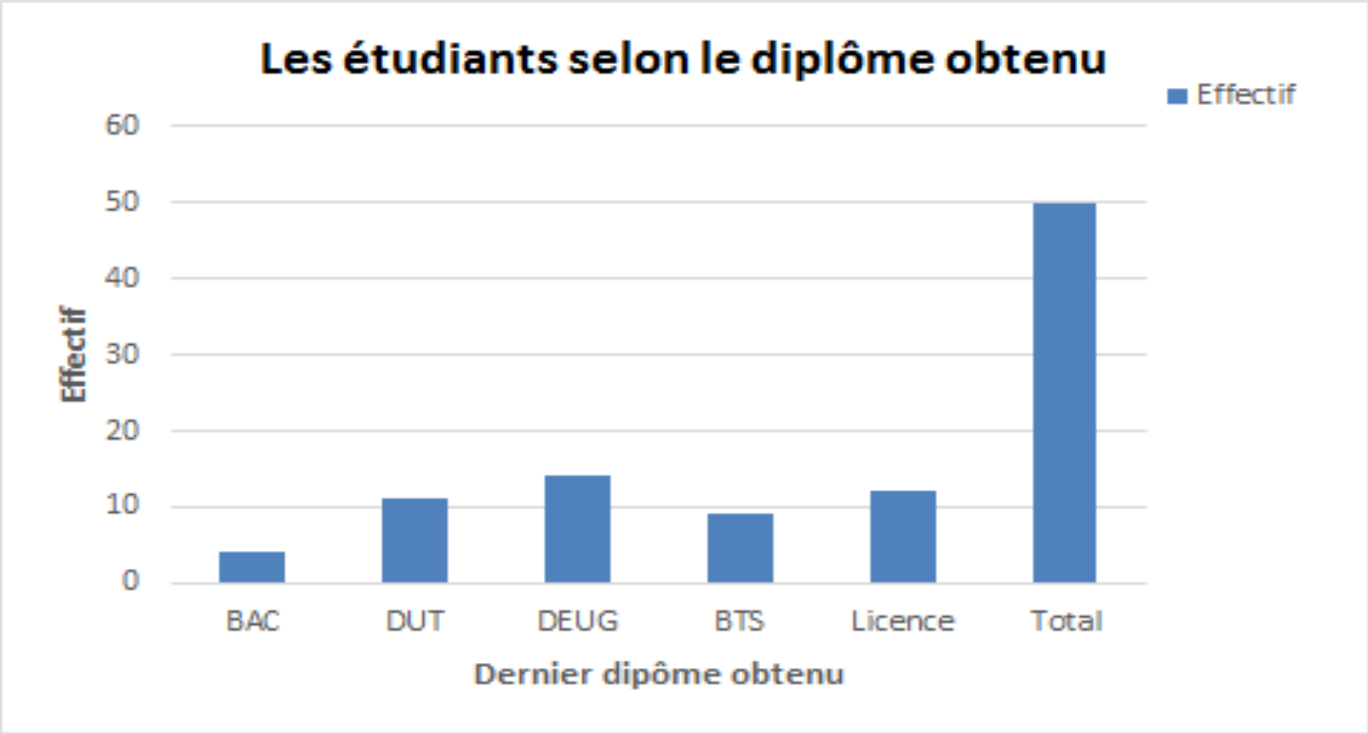
3.2 Présentation

- La représentation dépend de **la nature du caractère étudié**



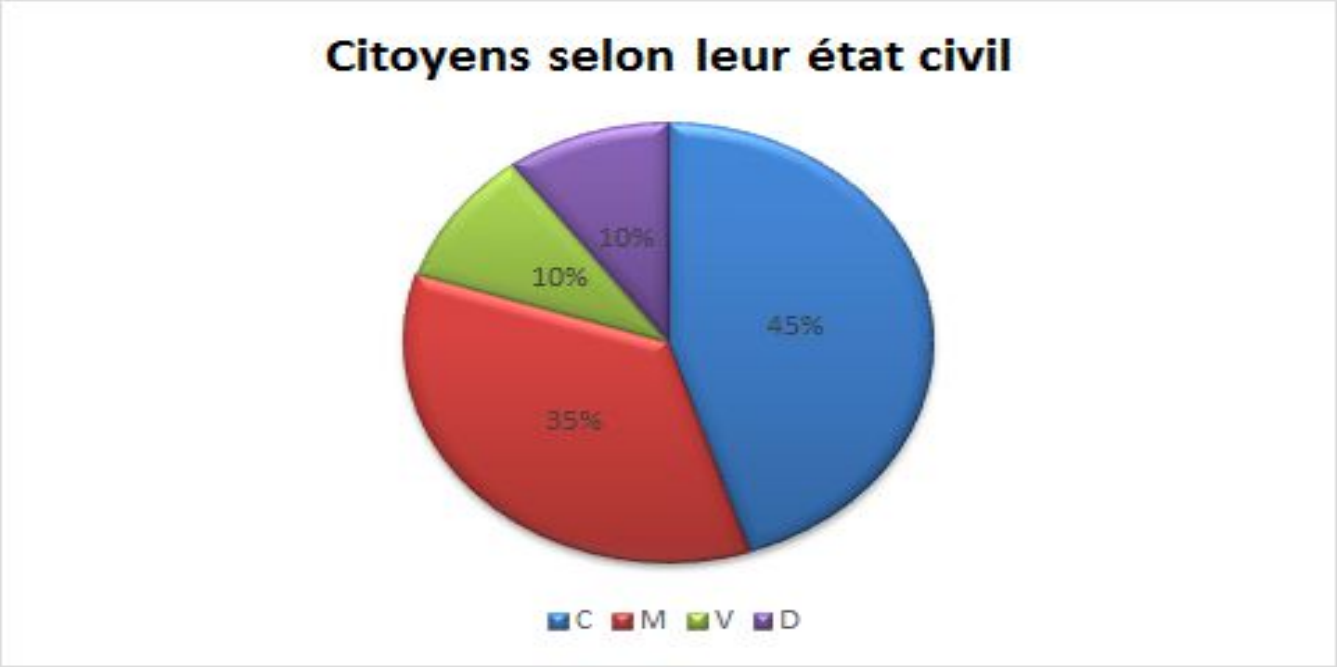
3.3 cas d'un caractère qualitatif (tuyaux d'orgue)

Dernier diplôme obtenu	Effectif
BAC	4
DUT	11
DEUG	14
BTS	9
Licence	12
Total	50



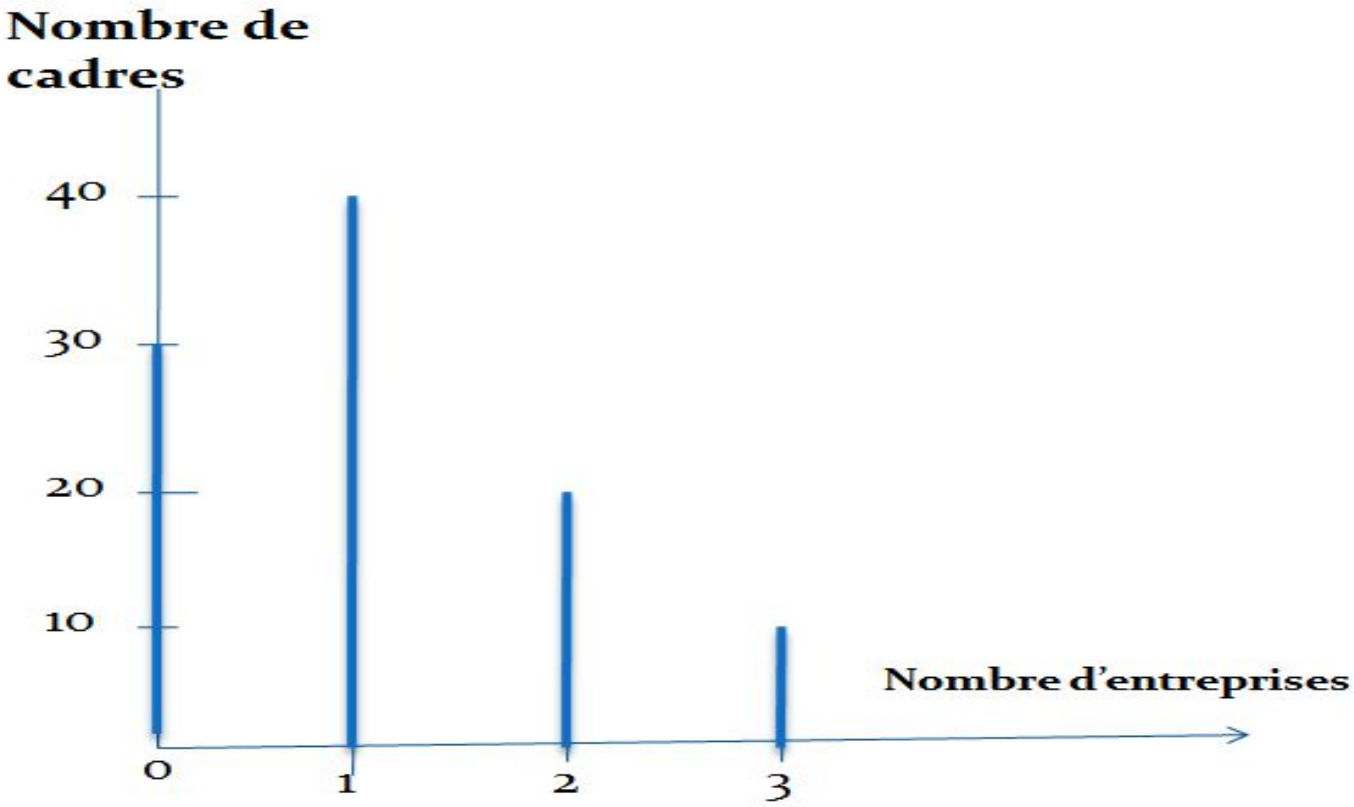
3.4 cas d'un caractère qualitatif (secteurs)

Etat civil	n_i	f_i	degré (l'angle au centre)
C	9	0,45	162°
M	7	0,35	126°
V	2	0,1	36°
D	2	0,1	36°
Total	**20**	**1**	**360**



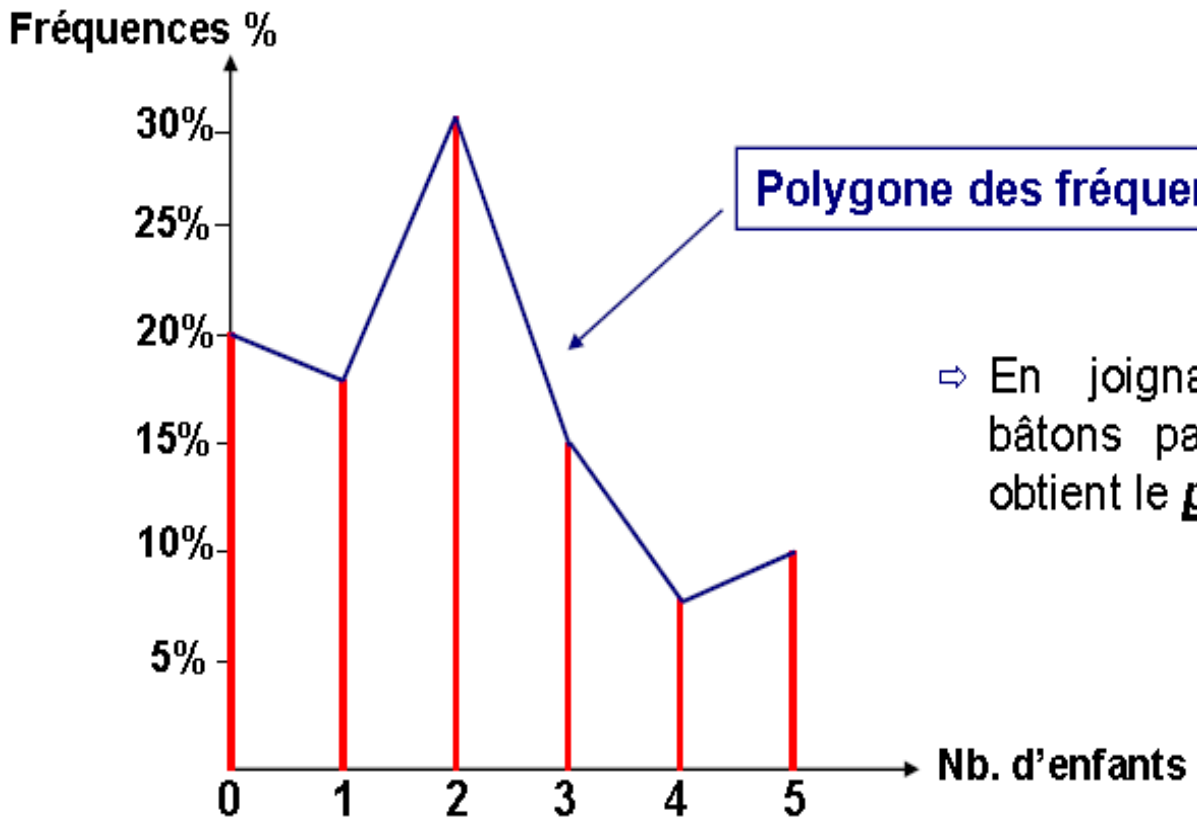
3.5 cas d'un caractère quantitatif discret (diagramme en bâtons)

Nombres de cadres	Effectif n_i	Fréquences f_i
0	30	0,3
1	40	0,4
2	20	0,2
3	10	0,1
Total	**100**	**1**



3.6 cas d'un caractère quantitatif discret
(diagramme en bâtons et polygone de fréquences)

Distribution des fréquences des salariés selon leur nombre d'enfants



⇨ En joignant les sommets des bâtons par une ligne brisée, on obtient le ***polygone de fréquences***

3.7 cas d'un caractère quantitatif discret
(Fonction de répartition)

- La fonction de répartition F(X) est représentée par la courbe cumulative des fréquences (effectifs):

Exemple

Nb d'enfants	Effectifs	Fréquences %	F(X) Moins de	F(X) Plus de
0	8	20%	20%	100%
1	7	17,5%	37,5%	80%
2	12	30%	60,5%	62,5%
3	6	15%	82,5%	32,5%
4	3	7,5%	90%	17,5%
5	4	10%	100%	10%
Total	40	100	-	-

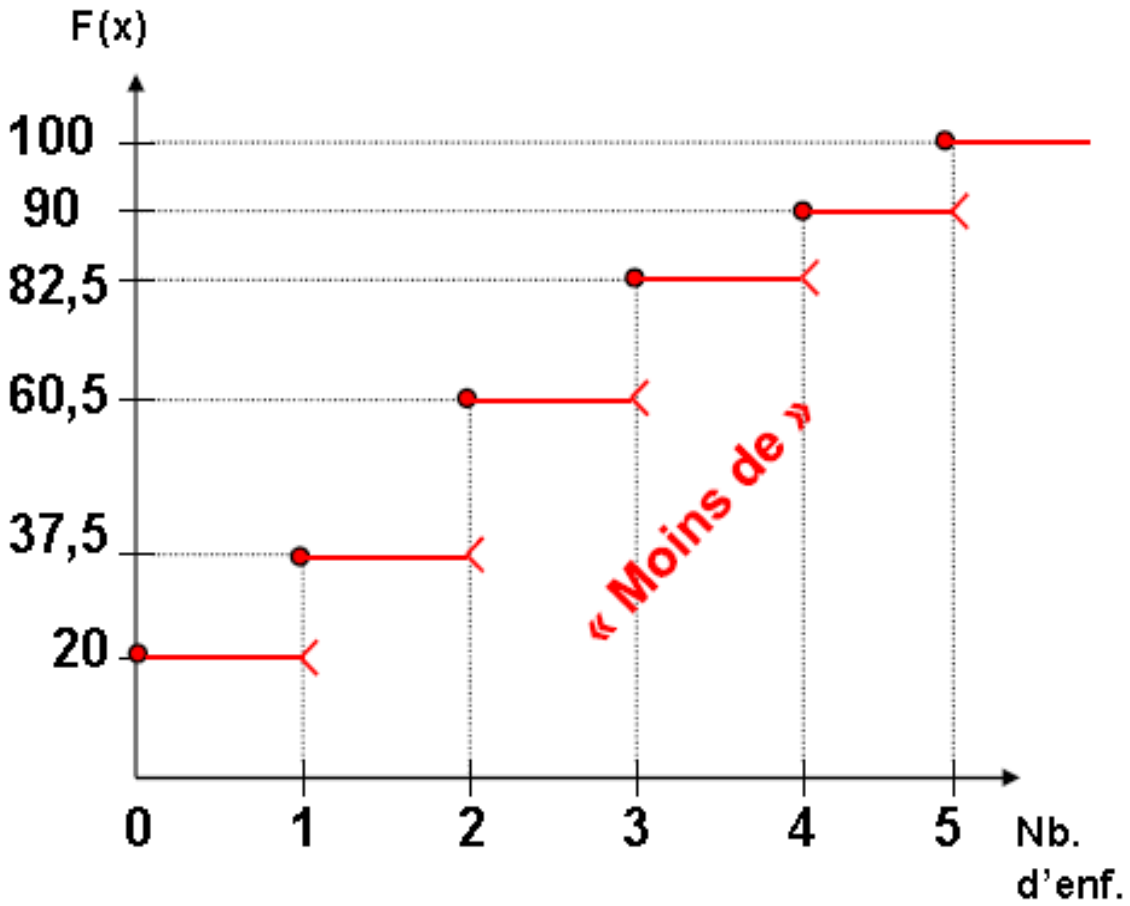
3.8 cas d'un caractère quantitatif discret
(Courbe cumulative croissante)

- La fonction de répartition F(X) est définie sur {0,1,2,3,4,5}, on commence par représenter les points dont on connait les coordonnées.

$F(x) = 20\% \quad \forall\, x < 1$

$F(x) = 37{,}5\% \quad \forall\, 1 \leq x < 2$

$F(x) = 60{,}5\% \quad \forall\, 2 \leq x < 3$



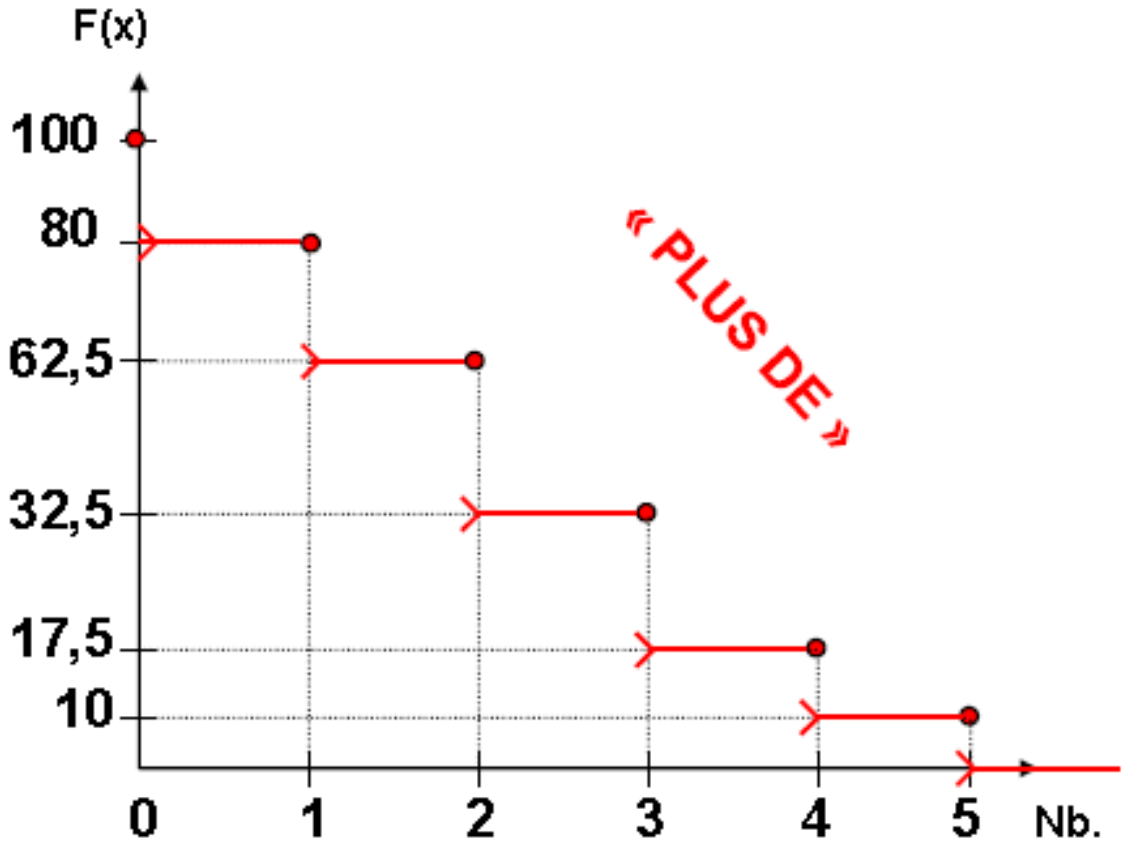
3.9 cas d'un caractère quantitatif discret
(Courbe cumulative décroissante)

- La fonction de répartition F(X) est définie sur {0,1,2,3,4,5}, on commence par représenter les points dont on connait les coordonnées.

$F(x) = 80\% \ \forall\, 0 < x \leq 1$

$F(x) = 62{,}5\% \ \forall\, 1 < x \leq 2$

$F(x) = 32{,}5\% \ \forall\, 2 < x \leq 3$

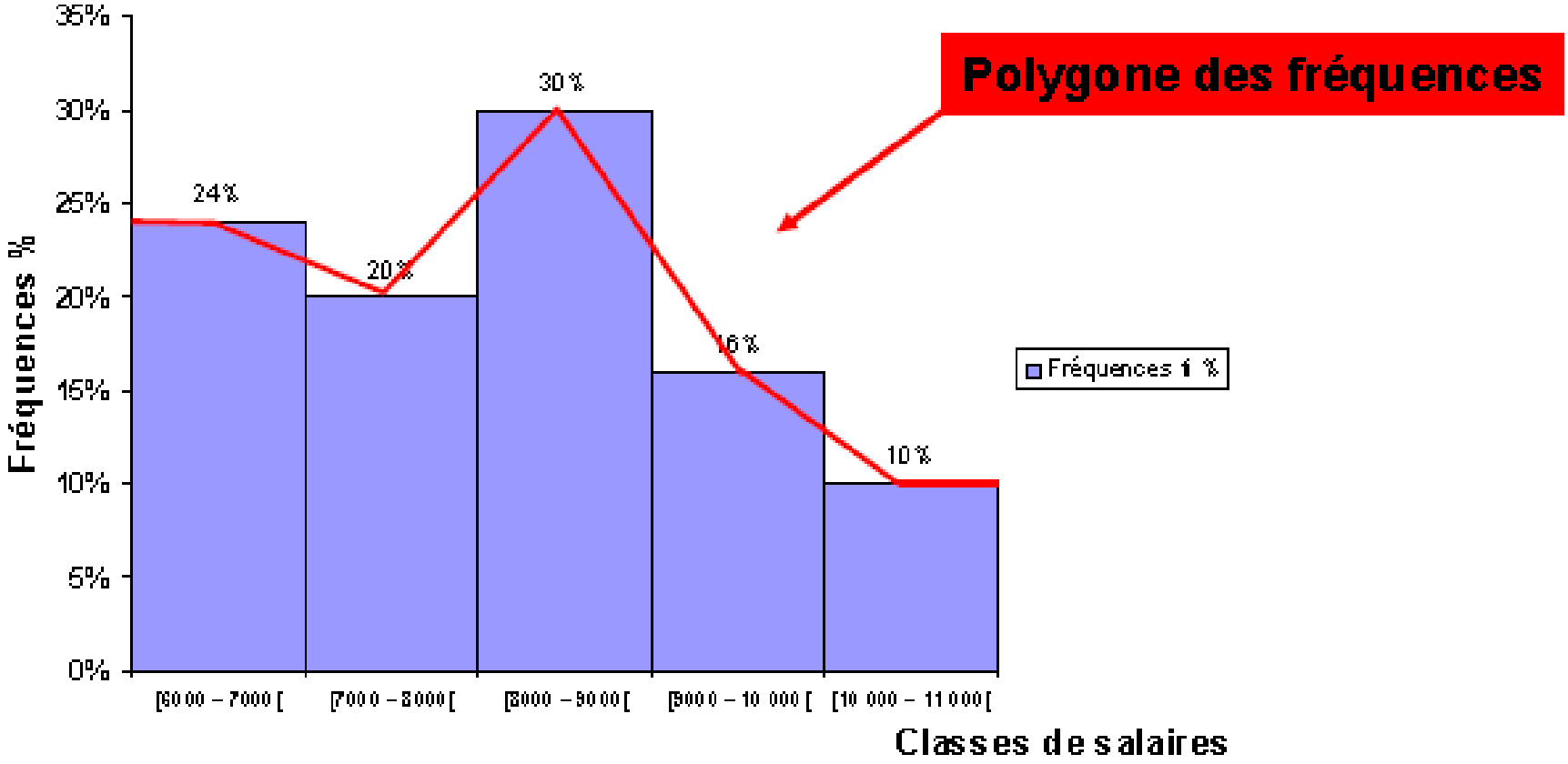


3.10 cas d'un caractère quantitatif continu (l'histogramme cas des classes d'amplitudes égales)

- Exemple : on souhaite étudier le salaires de 50 employés de la société STM

Salaires	Fréquences absolus ***ni***	Fréquences relatives ***fi***	Amplitudes ***ai***
6000-7000	12	24%	1000
7000-8000	10	20%	1000
8000-9000	15	30%	1000
9000-10000	8	16%	1000
10000-11000	5	10%	1000
Total	**50**	**100%**	

3.10 cas d'un caractère quantitatif continu
(l'histogramme cas des classes d'amplitudes égales)

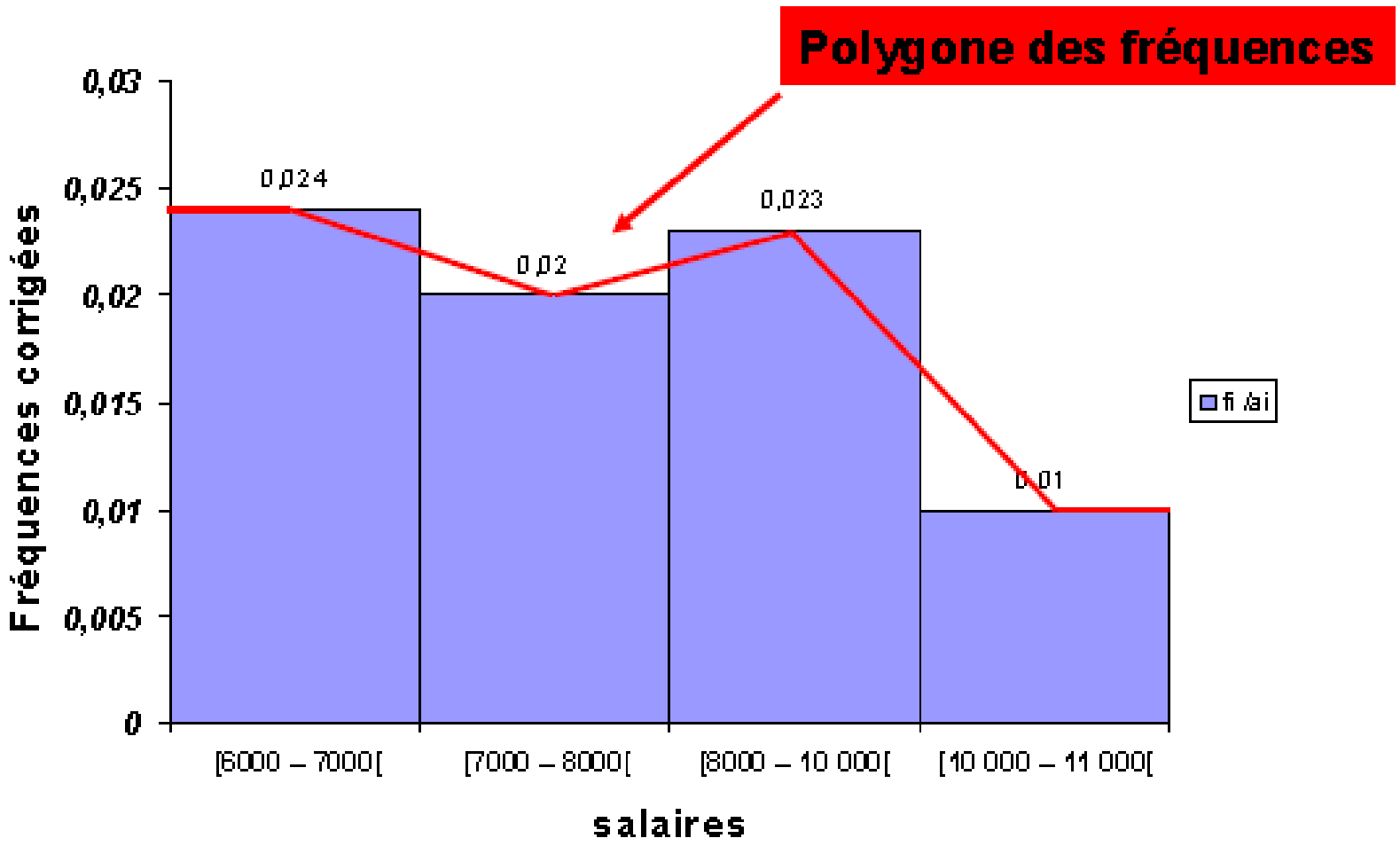


3.11 cas d'un caractère quantitatif continu (l'histogramme cas des classes d'amplitudes inégales)

- Exemple : on souhaite étudier le salaires de 50 employés de la société ABC

Salaires	Fréquences absolus ***ni***	Fréquences relatives ***fi***	Amplitudes ***ai***	***fi/ai***
6000-7000	12	24%	1000	0,024
7000-8000	10	20%	1000	0,02
8000-1000	23	46%	2000	0,023
10000-11000	5	10%	1000	0,01
Total	**50**	**100%**	-	-

3.11 cas d'un caractère quantitatif continu
(l'histogramme cas des classes d'amplitudes inégales)



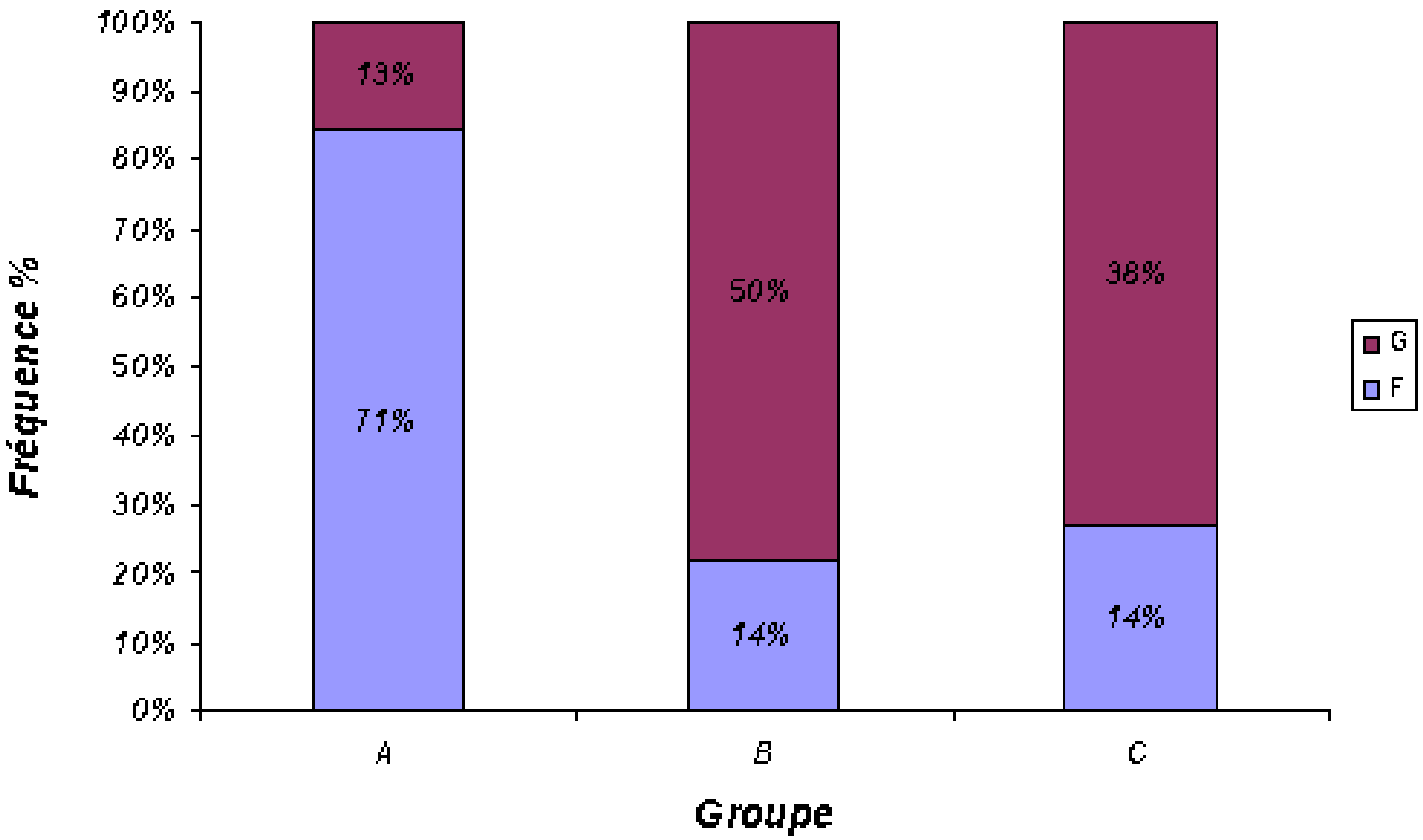
3.12 Représentation de caractère à deux dimensions

- Exemple : La variable sexe conditionnement à la variable groupe

Groupe Sexe	A	B	C	Total
F	83%	20%	25%	47%
G	17%	80%	75%	53%
Total	100%	100%	100%	100%

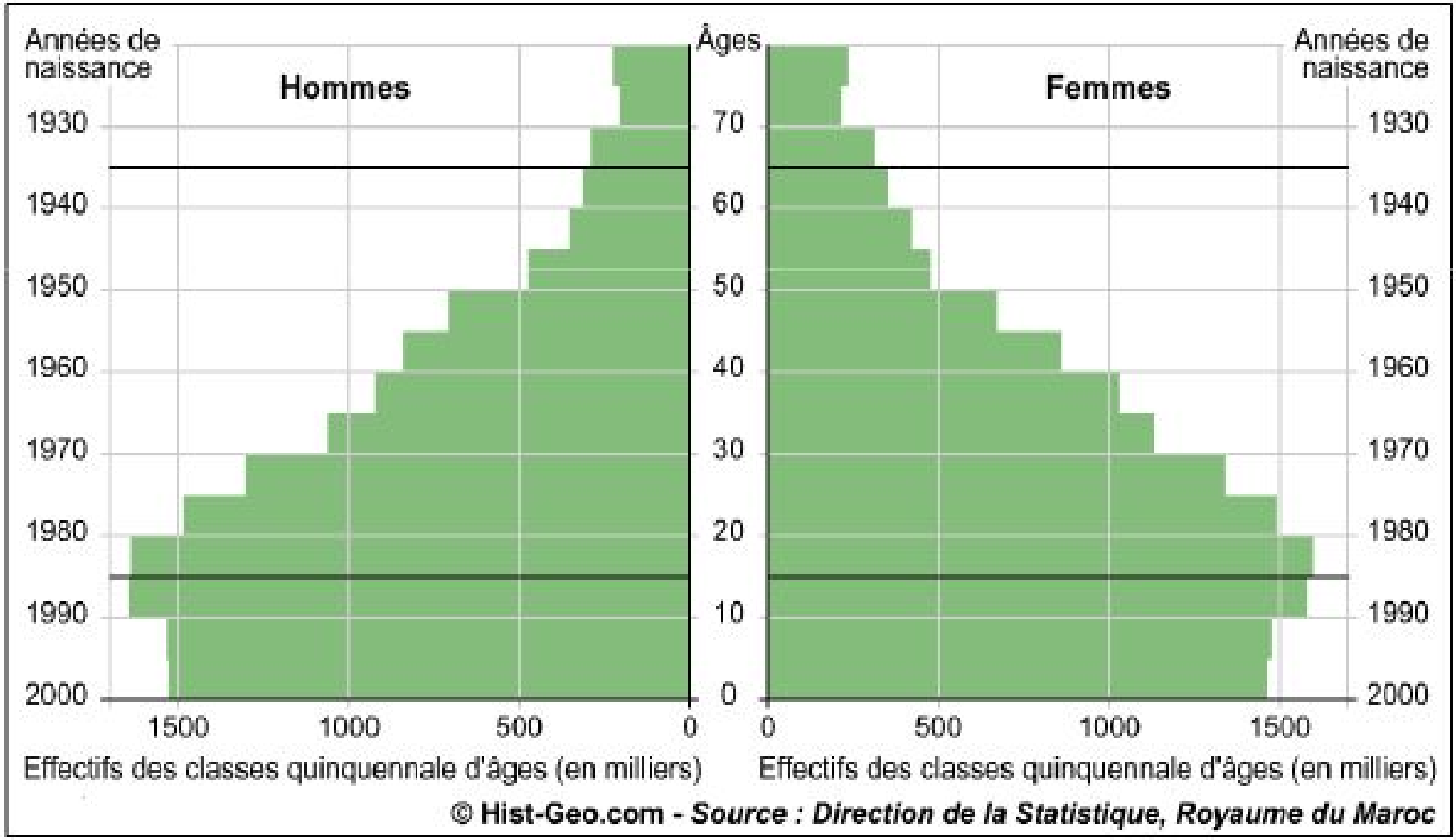
3.12 Représentation de caractère à deux dimensions

- Diagramme des fréquences de la variable sexe conditionnellement à la variable âge



3.12 Représentation de caractère à deux dimensions

- Pyramid des âges de la population Marocain au 1 er Juillet 2000



3.13 Représentation logarithmique

- Jusque là : on a vu des représentations arithmétiques uniquement. Ces représentations ont **deux inconvénients** :
- **1/ On ne peut pas juxtaposer des grandeurs (ou distributions) à échelles très différentes :**

Exemple : Projeter indice des prix à l'importation dans le temps avec le volume biens importés

- **2/ On ne peut pas comparer les variations**

Année	indice (1990 base 100)	variation indice des prix	importations	variation Importations
1990	100		75000	
1991	125	25	65000	-10000
1992	135	10	58000	-7000
1993	150	15	48000	-10000
1994	178	28	32000	-16000
1995	200	22	24000	-8000
1996	120	-80	48000	24000
1997	101	-19	67000	19000
1998	76	-25	85000	18000
1999	54	-22	102000	17000

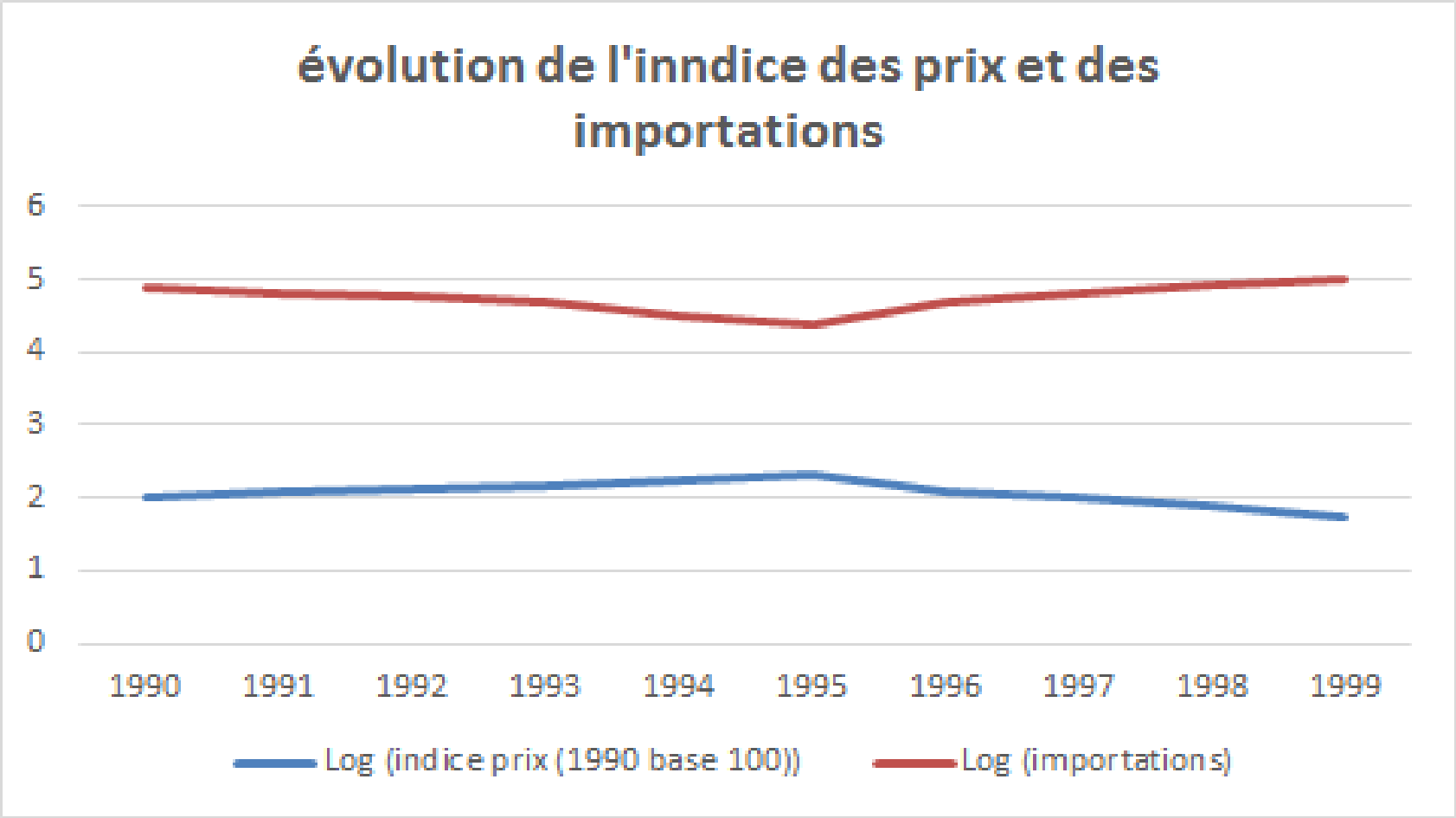
3.13 Représentation logarithmique

Année	Log (indice prix (1990 base 100))	Log (importations)	Dlog(indice Prix)	Dlog(Importations)
1990	2	4,87	0,09	-0,062
1991	2,09	4,81	0,03	-0,04
1992	2,13	4,76	0,045	-0,08
1993	2,17	4,68	0,07	-0,17
1994	2,25	4,50	0,05	-0,12
1995	2,30	4,38	-0,22	0,30
1996	2,07	4,68	-0,07	0,14
1997	2,01	4,82	-0,12	0,10
1998	1,88	4,92	-0,14	0,07
1999	1,73	5,01	-1,73	-5,01

- Les variations des logs sont comparables à des taux de croissance des variables (pour des petites variations) :
- **Démonstration :** Soit x une variable quelconque

$$d(\log x) = f'(x) = \frac{dx}{x}$$ (taux de croissance)

3.13 Représentation logarithmique



Chapitre 4
La réduction des données

4.1 tendance centrale univarié
4.1.1 Le mode (variable discrète)

- **Définition :** Le mode est la valeur de la variable associée au plus grand nombre d'effectif (ou encore à la plus grande fréquence).

Salaire	Effectif	fréquence	Mode
1000	25	0,25	
1800	45	0,45	1800 (mode)
2200	30	0,3	

4.1 tendance centrale univarié
4.1.2 Le mode (variable continu)

- **Classe à amplitude égale**

Salaires	Effectif	fréquence	Classe modale	'Mode' estimé
[1000-2000[	25	0,25		
[2000-3000[	45	0,45	[2000-3000[	2500
[3000-4000[	30	0,3		

- **Classe à amplitude égale: (cc = petite amplitude /ai)*fi**

Salaires	Effectif ni	Fréquence fi	Amplitude ai	Fréquence Ajustée fi*cc	Classe modale	'Mode' estimé
[1000-2000[	25	0,25	1000	0.125		
[2000-4500[	55	0,55	2500	0.11		
[4500-5000[	20	0,2	500	0.2	[4500-5000[	4750

4.1 tendance centrale univarié
4.1.3 La médiane (variable discrète)

- **Définition :** La médiane est la valeur d'une série ***ordonnée*** partageant celle-ci en **deux sous ensembles à taille égale**.

✓ Chaque valeur observée est unique dans la série:

- Série à nombre d'observations impairs : [3, 5, 7, 9, 10 ,11, 12]. ***Mé=9***
- Série à nombre d'observations pairs : [3, 5, 7, 9, 10, 11,12, 13].

➢ ***Intervalle Médian=[9,10],*** *on peut l'estimer* ***à 9.5 (médiane)***

✓ Chaque valeur observée plusieurs fois dans la série:

- **Exemple simple**: [3, 5, 7, 9, 9, 9, 9, 9, 10 ,11,12].
- La médiane ici est 9 : car la valeur qui coupe l'échantillon en 2 échantillons égales est 9.
- **Remarque:** ceci étant dit, plus de 50% des observations ont une valeur inférieure ou égale à neuf (8/11=72% des valeurs '<' ou '=' 9) !

4.1 tendance centrale univarié
4.1.3 La médiane (variable discrète)

- **Exemple plus général :** grille de salaire (3 rémunérations fixes) pour 101 personnes

Salaire	Effectif	Fréquence	F cumulée
1000	25	0,25	0.25
1800 (médiane)	45	0,45	0.70
2200	30	0,3	1

- 70% sont rémunérés d'une valeur inférieure ou égale à 1800.
- Mais, en ordonnant l'échantillon (plus petite à la plus grande valeur) : de la 26ème personne observée à la 70ème personne observée : salaire donné 1800. Donc, la 51ème personne observée (ou l'observation qui coupe l'échantillon en deux) est 1800.
- Mé=1800
- (mais on ne peut pas dire que 50% gagnent au plus (ou au moins) 1800)

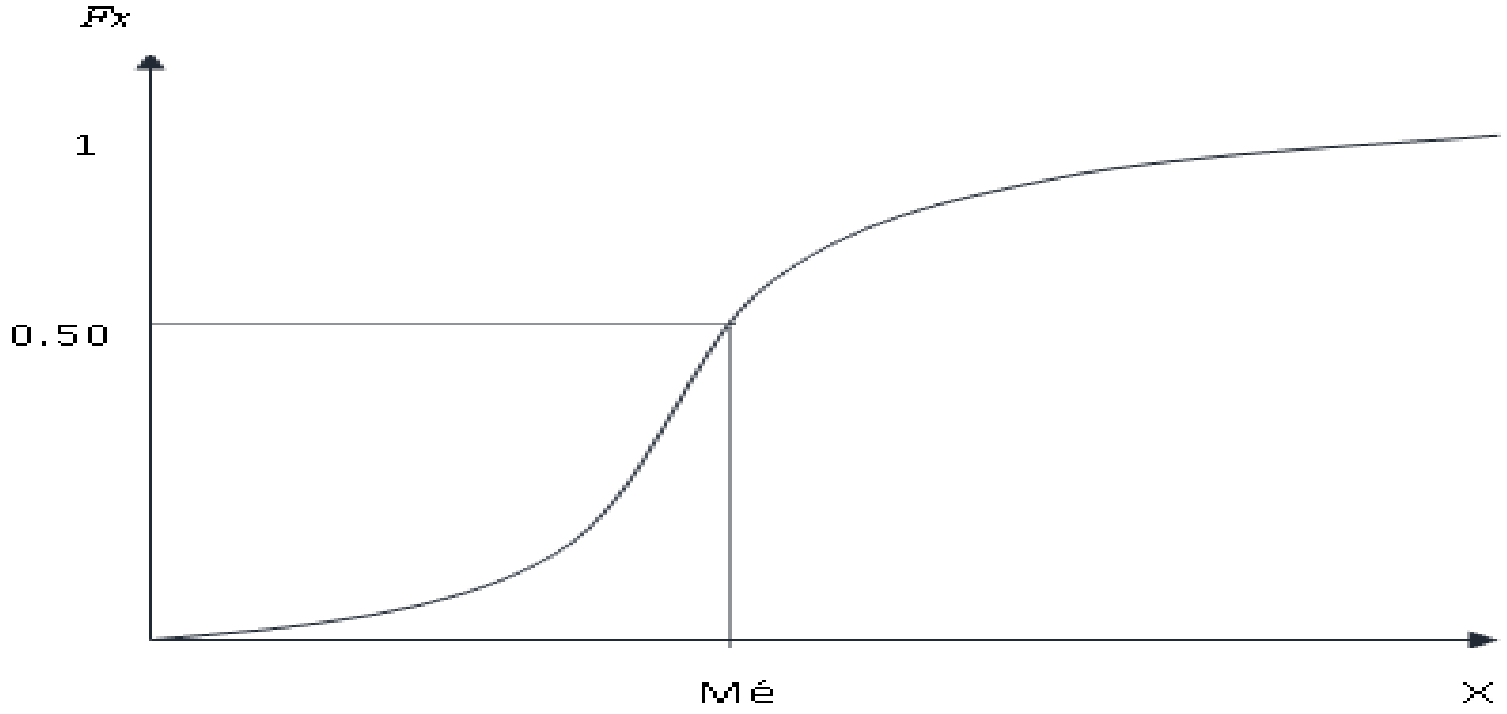
4.1 tendance centrale univarié
4.1.4 La médiane (variable continue)

- **Méthode de détermination de la médiane (graphique):**

- On trace la fonction de répartition *F*

- On localise la valeur sur la fonction *F* pour laquelle 50% de la population est associée

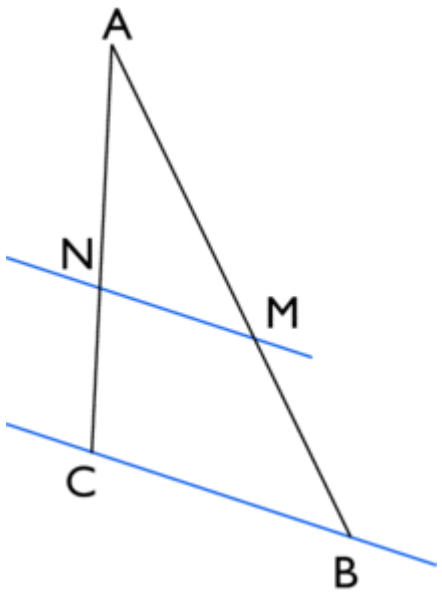
- Cette valeur est la médiane

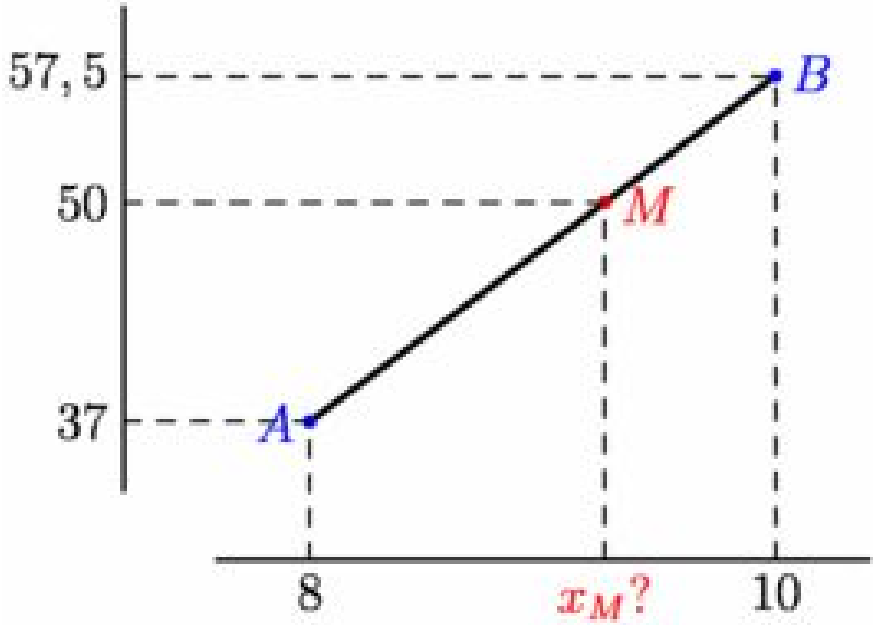


Rappel (interpolation linéaire)

- **Rappel de l'interpolation linéaire (Théorème de Thales)**

✓ ABC est un triangle. M se trouve sur le segment [AB] et N sur le segment [AC]. D'après le théorème de Thalès, **si les droites (BC) et (MN) sont parallèles,** alors on a l'égalité: $\frac{AM}{AB} = \frac{AN}{AC} = \frac{MN}{BC}$





4.1 tendance centrale univarié
4.1.4 La médiane (variable continue)

- **Méthode de détermination de la médiane (interpolation linéaire)**
- **Exemple :** Distribution des Salaires observés en continu

Salaires	Effectif	Fréquence (fi)	F cumulée (Répartition)
[1000-2000[	20	0,20	0.20
[2000-4000[	50	0,50	0.70
[4000-6000[	30	0,30	1

- **La classe** ou se trouve la médiane est 2000-4000

$$0.2 \rightarrow 2000$$
$$0.5 \rightarrow Me$$
$$0.7 \rightarrow 7000$$

$$Mé = 2000 + \left(\frac{(4000 - 2000)}{0.7 - 0.2}\right)[0.5 - 0.2] = 3200$$

4.2 Les quartiles « ntiles »

- **Définition :** Valeur de la variable qui partage la série en plusieurs (ou encore ‘n’) sous-groupes égaux.

Quantile (‘n’tile)	Nb de sous ensembles obtenues (n)	Nombre de quantiles (n-1)	Chaque écart inter-quantile représente :	Fonction de répartition $F('n'tile_h)$
Médiane	2	1 (la médiane)	50% (écart Mé-valeur extrêmes)	$F(Mé) = 50\%$
Quartile	4	3 (Q1, Q2, Q3)	25%	$F(Q1) = 25\%$ $F(Q2) = 50\%$ $F(Q3) = 75\%$
Décile	10	9 (D1,…,D9)	10%	$F(D1) = 10\%$ … $F(D9) = 90\%$
Centile	100	99 (C1,…, C99)	1%	$F(C1) = 1\%$ … $F(C99) = 99\%$

4.2 Les quartiles « ntiles »

- **Exemple :** Soit une population de 80 salariés classés d'après le niveau de leur salaire journalier :

Classes (en DH)	n_i	$n_i c$
[90, 100[	5	5
[100, 110[	9	14
[110, 120[	16	30
[120, 130[	25	55
[130, 140[	13	68
[140, 150[	7	75
[150, 160[	3	78
[160, 170[	2	80
Total	80	

4.2 Les quartiles « ntiles »

- **Exemple :** Soit une population de 80 salariés classés d'après le niveau de leur salaire journalier :

✓ ***Les quartiles*** *:*

- n/4 = 20 se trouve dans la classe [110, 120[
- $\mathbf{Q_1}$ = 110 + (120-110).[(20-14)/(30-14)] = 113,75
- $\mathbf{M_e} = Q_2 = 120 + (130 - 120).[(40-30)/(55-30)] = 124$
- $\mathbf{Q_3}$ = 130 + (140-130).[60-55)/(68-55)] = 133,85

➢ **L'intervalle interquartile** est **(133,85-113,75),** il contient 80% des observations.

✓ ***Les déciles*** *:*

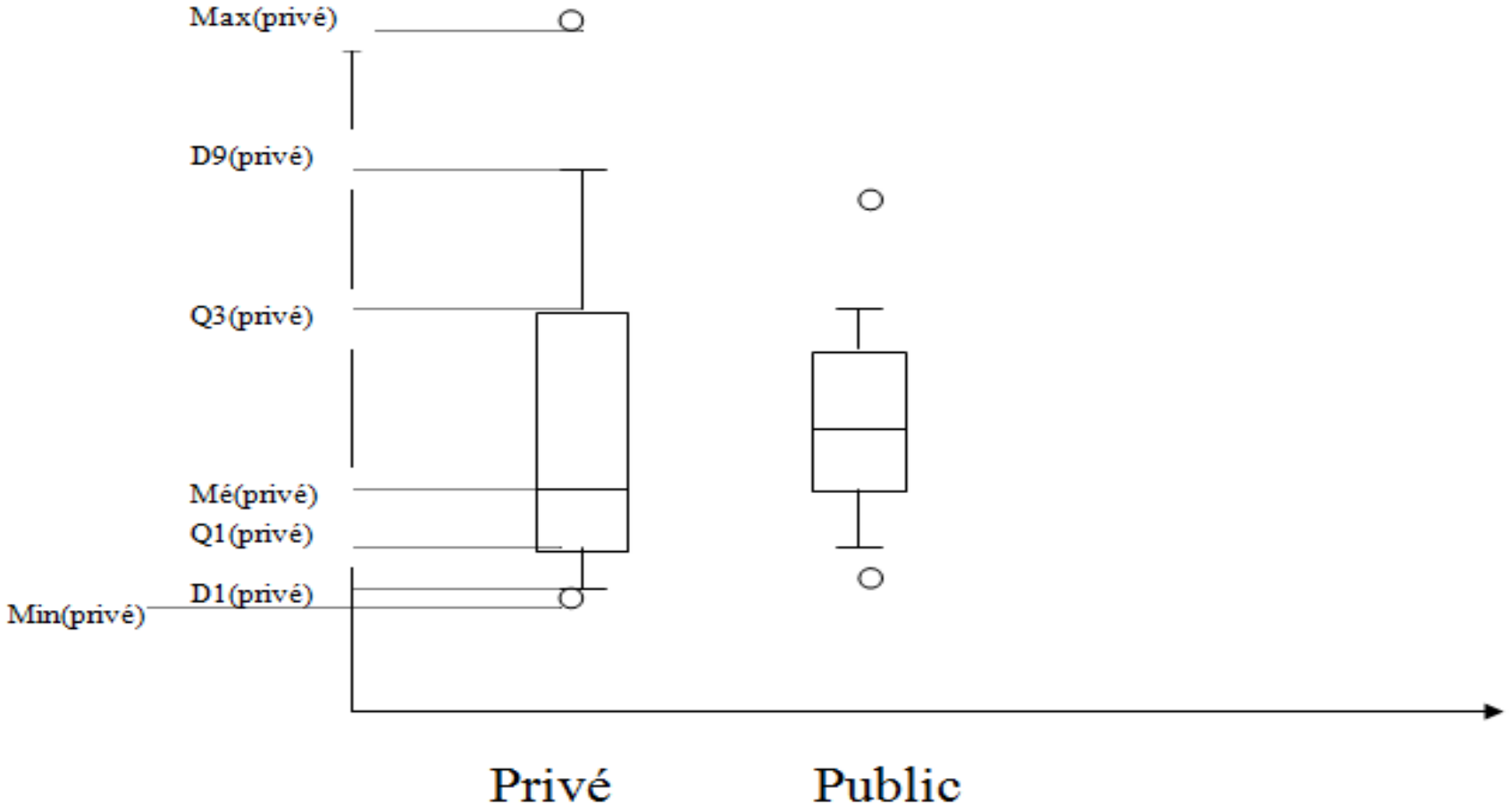
- $F(D_1) = 0,1$ $nc(D_1) = n/10$
- $F(D_2) = 0,2$ $nc(D_2) = 2n/10$
- ……. L'intervalle D9 – D1 s'appelle **intervalle interdécile** ; il contient 80% des observations.

✓ ***Les centiles*** *:*

- $F(C_1) = 0,01$ $nc(C_1) = n/100$ …..

4.3 Graphique « Box Plot »

- **Définition :** C'est un graphique qui permet de représenter, la plupart des statistiques étudiées jusque-là : **médiane, quartiles, déciles**. Aussi, il rend compte du niveau d'asymétrie, de la dispersion et des valeurs extrêmes de la distribution. **Exemple** : Etude sur deux distributions de salaires dans le secteur privé et le secteur public en France.



4.4 La moyenne
4.4.1 La moyenne arithmétique

✓ **La moyenne arithmétique simple :**

Définition : c'est le rapport entre somme des valeurs observées et le nombre d'observations.

$$\bar{x} = \frac{1}{n}\sum_i x_i$$

✓ **La moyenne arithmétique pondérée :**

Définition : somme des valeurs observées, pondérées par leur poids (ou fréquence d'apparition)

$$\bar{x} = \sum_i w_i x_i = \sum_i \frac{n_i}{n} x_i = \sum_i f_i . x_i$$

La moyenne arithmétique pondérée (variable continue):

$$\bar{x} = \frac{1}{n}\sum_i n_i x_i^c \quad \textit{avec} \quad x_i^c = \frac{x_{i+1} + x_i}{2}$$

4.4 La moyenne
4.4.2 généralisation de la moyenne

Paramètre 'r'	Moyenne	Formule
R=1	Moyenne Arithmétique (pondérée)	$\overline{x} = \left[\frac{1}{n} \sum_i n_i x_i \right]^{1/r}$
R=2	Moyenne Quadratique	$\overline{q} = \left[\frac{1}{n} \sum_i n_i x_i^2 \right]^{1/2}$
R=-1	Moyenne Harmonique	$\overline{h} = \left[\frac{1}{n} \sum_i n_i x_i^{-1} \right]^{-1}$, facilité de calcul : $\frac{1}{h} = \left[\frac{1}{n} \sum_i \frac{n_i}{x_i} \right]$
R=e, e>0 et 'e' tendant vers 0	Moyenne Géométrique	$G = \left[\frac{1}{n} \sum_i n_i x_i^e \right]^{1/e}$ avec $e \longrightarrow 0$,e>0 2 expressions Alternatives: $G = \left(\prod_i x_i^{n_i} \right)^{1/n}$ ou $G = \exp\left[\frac{1}{n} \sum_i n_i \log x_i \right]$

4.4 La moyenne
4.4.3 généralisation de la moyenne

✓ ***Propriétés:***

$$h < G < \bar{x} < Q \qquad \bar{x}.h = G^2$$

- ***G*** : **Moyenne géométrique** utilisée pour calculer des moyennes de taux de croissance.

Exemple : Un prêt à 3% à la date 1, 5% date 2 et 7% date 3. Calculer taux d'intérêt annuel Moyen:

$$1 + i = \left[(1.03)(1.05)(1.07)\right]^{1/3}$$

- ***H: Moyenne Harmonique:*** Pour faire la moyenne de vitesses lorsque la distance sur laquelle chaque vitesse pratiquée est la même.

Exemple: Une voiture parcourt un trajet à 100 km/h de moyenne et le retour à 40 km/h. La vitesse moyenne de l'aller-retour :

$$VM(H) = \frac{2}{\frac{1}{100} + \frac{1}{40}} = 57.14$$

4.4 La moyenne
4.4.3 généralisation de la moyenne

- ***q: Moyenne quadratique:*** *Prenons un rapide exemple :*

- *considérons les nombre suivants {-2, 5, -8, 9, -4 } Nous pouvons en calculer la moyenne arithmétique* ***avec l'inconvénient de voir se neutraliser les valeurs positives et négatives et d'aboutir à un résultat nul sans que cela ne nous apprenne quoi que ce soit****. En effet,.*

- *Le calcul de la moyenne quadratique pour la même série donne .*

$$MQ(Y) = \sqrt{\frac{\sum_{i=1}^{N} y_i^2}{N}} = 6.16$$

4.5 La dispersion

4.5.1 L'étendue , l'intervalle interquartile, l'écart absolu moyen

- ✓ **L'étendue*:*** est la différence entre la plus grande et la plus petite valeur des observations de la série statistique : (idée sur la dispersion). (idée souvent fausse du fait des valeurs extrêmes).

$$R = X_{\max} - X_{\min}$$

- ✓ **Les intervalles interquantiles:** (Q3-Q1;D9-D1) (bon indicateur de dispersion cependant reste vague par rapport à l'écart type).

- ✓ **L'écart absolu moyen:** l'écart absolu moyen est la moyenne arithmétique des écarts par rapport à la tendance centrale, exprimée en valeur absolue.

$$E_{\bar{x}} = \frac{1}{n} \sum n_i \left\| x_i - \bar{x} \right\|$$

- *Cet indicateur ne se prête pas facilement au calcul algébrique.*

4.5 La dispersion
4.5.2 L'écart absolu moyen (exemple)

- **Exemple:** Soit sept températures relevées dans les salles de production d'une unité de filature à la même heure:

Degrés (°c)	12	16	11	14	13	17	15	= 14
	2	2	3	0	1	3	1	$\sum = 12$

- **L'écart absolu moyen** = 1/7 x 12 = 1,71
- Les températures s'écartent en moyenne de 1,71° par rapport à leur moyenne arithmétique.

4.5 La dispersion
4.5.3 La variance et l'écart type

- ✓ **La variance:** est la moyenne des écarts (par rapport à la moyenne) au carrés

$$V(x) = \frac{1}{n}\sum_i (x_i - \bar{x})^2 \qquad V(x) = \frac{1}{n}\sum_i n_i (x_i - \bar{x})^2$$ ***(variance pondérée)***

- ✓ **L'écart-type:** est la Racine carrée de la variance ou encore, la moyenne ***quadratique*** des écarts (à la moyenne)

$$\sigma = \left(\frac{1}{n}\sum_i n_i (x_i - \bar{x})^2\right)^{1/2} = V(x)^{1/2} = \sqrt[2]{V(x)}$$

- ✓ **Exemple:** Soit sept températures relevées dans les salles de production d'une unité de filature à la même heure:

Degrés (°c)	12	16	11	14	13	17	15	= 14
	2	2	3	0	1	3	1	$\sum = 12$

4.5 La dispersion
Exemple 1

Xi degrés Tempe	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
12	-2	4
16	2	4
11	-3	9
14	0	0
13	-1	1
17	3	9
15	1	1

$$\bar{x} = 14$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{28}{7} = 4$$

4.5 La dispersion
Exemple 2

Classes	n_i	x_i	$n_i x_i$	$(x_i - \bar{x})^2$	$n_i (x_i - \bar{x})^2$
[3 ; 4[	26	3,5	91	2,496	64,896
[4 ; 5[	33	4,5	148,5	0,336	11,088
[5 ; 6[	64	5,5	352	0,176	11,264
[6 ; 7[	7	6,5	45,5	2,016	14,112
[7 ; 8[	10	7,5	75	5,856	58,56
Total	**140**		**712**		**159,92**

$$V(x) = 159{,}92/140 = 1{,}14$$

$$\sqrt{V(x)} = \sqrt{1.14} = 1.067$$

4.5 La dispersion
4.5.4 Formule développée de la variance

✓ **La formule développée de la variance (théorème de Koenig)**

$$V(x) = \frac{1}{n}\sum_i n_i (x_i - \bar{x})^2$$

$$V(x) = \frac{1}{n}\sum_i n_i (x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

$$V(x) = \frac{1}{n}(\sum_i n_i x_i^2 - 2\bar{x}\sum_i n_i x_i + \bar{x}^2 \sum_i n_i)$$

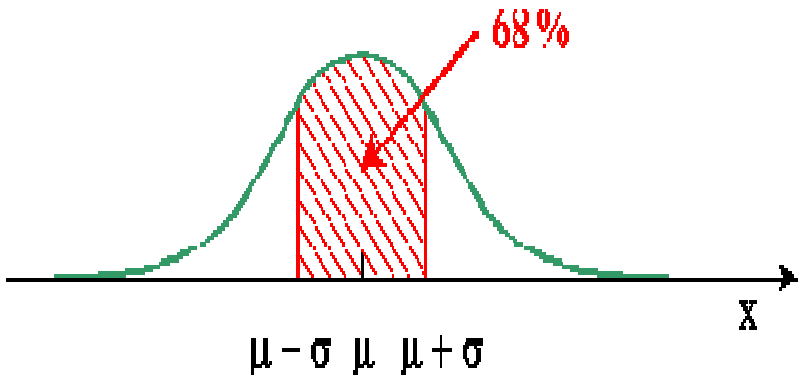
$$V(x) = \frac{1}{n}\sum_i n_i x_i^2 - 2\bar{x}\bar{x} + \bar{x}^2 = \frac{1}{n}\sum_i n_i x_i^2 - \bar{x}^2 \text{ (c.q.f.d)}$$

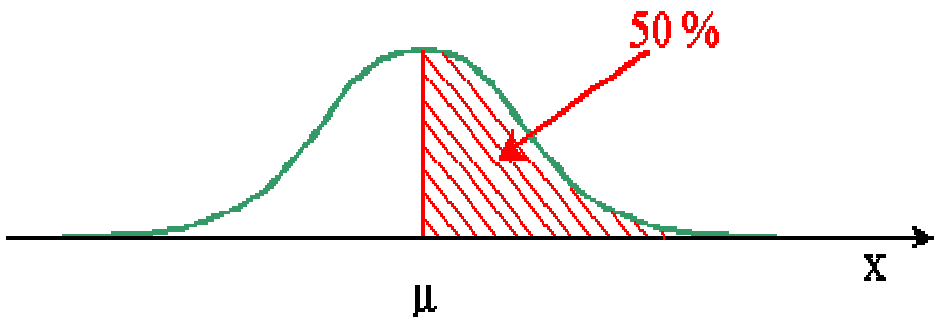
Donc on obtient

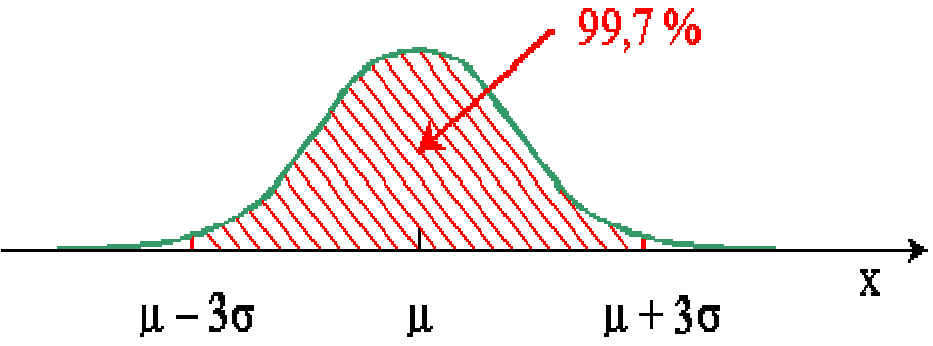
$$\boldsymbol{\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} n_i x_i^2 - \bar{x}^2} \qquad \sigma^2 = \sum_{i=1}^{k} f_i x_i^2 - \bar{x}^2$$

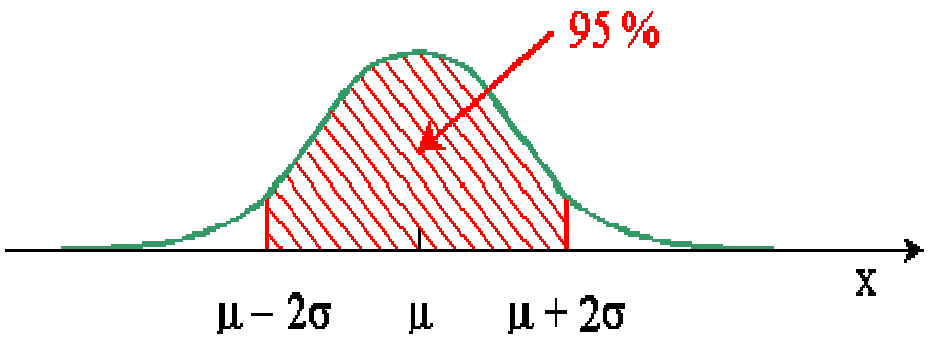
4.5 La dispersion
4.5.5 série distribué normalement

- **Autre propriété** : Quand utilise loi normale s'applique), on peut utiliser l'écart-type parallèlement à la moyenne pour calculer des intervalles de données. Si μ= **moyenne=médiane**,σ= écart-type et **x** = une valeur incluse dans l'ensemble de données, alors









4.5 La dispersion
4.5.6 Formule de décomposition de la variance

✓ La formule de décomposition de la variance

Soit une population P composée de plusieurs sous populations ($P1,..Pp$).
Soit ;

- $\overline{x}$ moyenne 'globale' et $V(x)$ variance 'globale' de la population ;
- $\overline{x_p} = \frac{1}{n_p} \sum_{i \in p} n_{i,p} x_{i,p}$ la moyenne 'locale' au sein de chaque sous-population p
- $et\ V(x_p) = \frac{1}{n_p} \left[\sum_{i \in p} n_{i,p} (x_{i,p} - \overline{x_p})^2 \right]$ la variance 'locale' au sein de chaque sous-population p.

Proposition : La variance 'globale' est la somme de :

- la moyenne des variances 'locales' (variance intra-population)
- et de la variance des moyennes 'locales' (variance interpopulation)

$$V(x) = \underbrace{\left[\frac{n_1}{n} V(x_1) + \ldots + \frac{n_P}{n} V(x_P) \right]}_{\overline{V(x_i)}\ (\text{var. Intra-population})} + \underbrace{\left[\frac{n_1}{n} (\overline{x_1} - \overline{x})^2 + \ldots + \frac{n_P}{n} (\overline{x_P} - \overline{x})^2 \right]}_{V(\overline{x})(Variance\ \text{inter-population})}$$

$$V(x) = \frac{1}{n} \sum_{p=1 \ldots P} n_{i,p} V(x_p) + \frac{1}{n} \sum_{p} n_{i,p} (\overline{x_p} - \overline{x})^2$$

4.5 La dispersion
4.5.7 Coefficient de variation

- **Définition :** Le coefficient de variation est une mesure de la dispersion relative (écart type par rapport à la moyenne) d'une série. Pour le besoin de comparer des dispersions de différentes séries.

$$CV = \frac{\sigma}{\bar{x}}$$

- **Exemple:** Demande d'importation sur une période de 30 ans

France (euros)	Allemagne (Euros)	Canada (Dollars can.)
Moyenne= 5	Moyenne=5	Moyenne=6
Ecart type=2.5	Ecart type=1	Ecart Type=4
CV=2.5/5=0.5	**CV=1/5=0.2**	**CV=0.66**

- La demande d'importation est plus dispersée au Canada et en France
  - Le **coefficient de variation** peut être utilisé pour caractériser l'homogénéité ou l'hétérogénéité d'un échantillon:
  - Si CV < 10% alors la population peut être considérée comme **homogène** ;
  - Si le CV **est entre 10% - 20%** de la population peut être considérée comme **relativement homogène**;
  - Si le CV **est entre 20% - 30%** de la population peut être considérée comme **relativement hétérogène** ;
  - Si CV > 30% de la population peut être considérée comme **hétérogène.**

4.5 La dispersion
4.5.8 Les moments

- ✓ **Définition** : **m_r**, le **moment d'ordre r par rapport à a**, est défini comme la moyenne des écarts par rapport à **a** de x, élevés à la puissance **r**.

$$m_r = \sum_{j=1}^{k} (x_j - a)^r . f_j$$

En pratique, seuls certains moments sont utilisés et donc calculés, on distingue :

- ✓ **<u>Les moments non centrés (par rapport à 0)</u> :**

$$m_r = \sum_{j=1}^{k} x_j{}^r . f_j \quad \textbf{avec} \quad m_0 = 1 \; ; \; m_1 = \overline{x} \; ; \; m_2 = \sigma^2 + \overline{x^2}$$

- ✓ **<u>Les moments centrés (par rapport à la moyenne)</u> :**

$$\mu_r = \sum_{j=1}^{k} (x_j - \overline{x})^r . f_j \quad \textbf{avec} \quad \mu_0 = 1 \; ; \; \mu_1 = 0 \; ; \; \mu_2 = m_2 - m^2{}_1 = \sigma^2$$

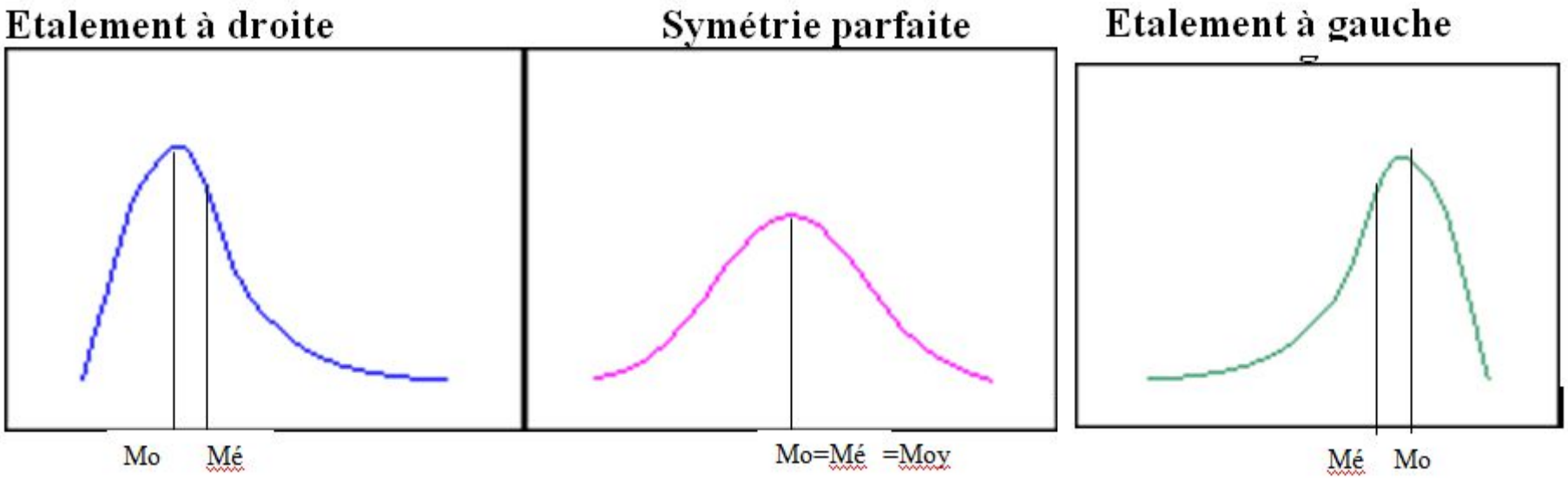
$$\mu_3 = m_3 - 3m_2 . m_1 + 2(m_1)^3$$

4.6 Les caractéristiques de forme

4.6.1 La forme d'une distribution (comparaison des tendances centrales et coefficient d'asymétrie

✓ **Comparaison de tendances centrales:**

- Si **Mo=Mé**=Moyenne alors la série est symétrique
- Si **Mo>Mé** alors la distribution est étalée vers la gauche
- Si **Mo<Mé** alors la distribution est étalée vers la droite

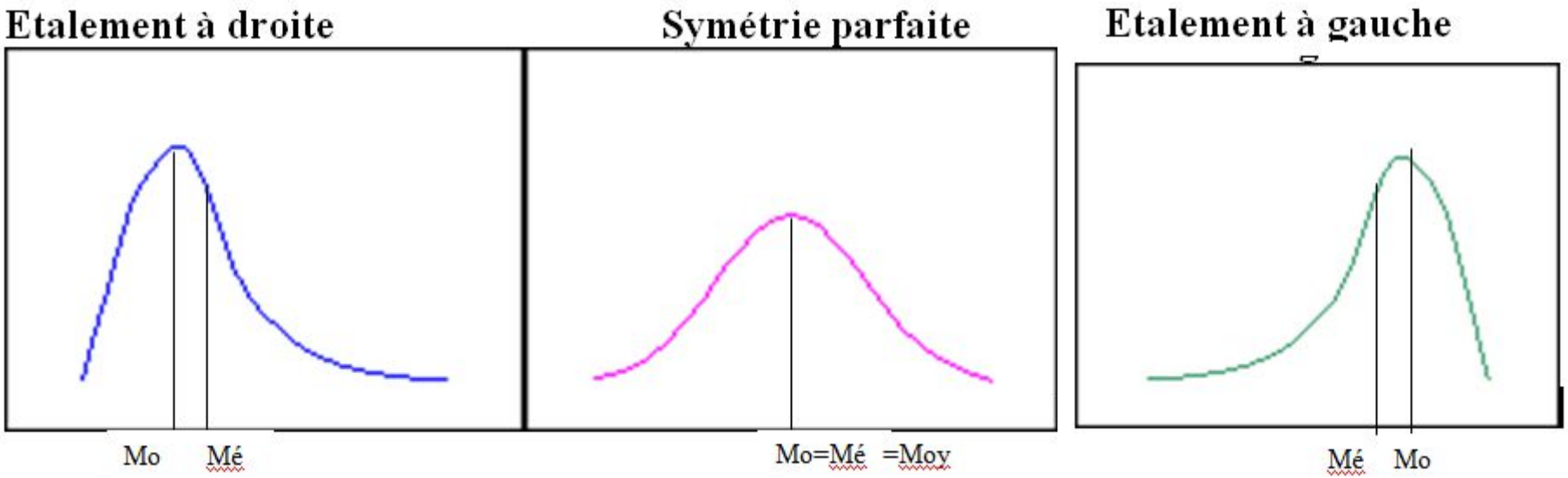


4.6 Les caractéristiques de forme

4.6.1 La forme d'une distribution (comparaison des tendances centrales et coefficient d'asymétrie et d'aplatissement

✓ **Comparaison de tendances centrales:**

- Si **Mo=Mé**=Moyenne alors la série est symétrique
- Si **Mo>Mé** alors la distribution est étalée vers la gauche
- Si **Mo<Mé** alors la distribution est étalée vers la droite



4.6 Les caractéristiques de forme
4.6.1 La forme d'une distribution (comparaison des tendances centrales et coefficient d'asymétrie et d'aplatissement

✓ **Calcul du coefficient d'asymétrie:**

1) Coefficient de Yule:

$$s=\frac{(Q3-Mé)-(Mé-Q1)}{(Q3-Mé)+(Mé-Q1)}$$

- Si **s=0**, alors il y a **symétrie**
- Si **s>0**, alors la courbe des fréquences **étalée à gauche**
- Si **s<0**, alors série est **étalée à droite**

2) **Coefficient de Pearson:**

$$p=\frac{\bar{x}-Mo}{\sigma}$$

- Si **p=0**, la série est **symétrique**
- Si **p>0** la série est **étalée à droite**
- **Si p<0** la série est **étalée à gauche**

4.6 Les caractéristiques de forme

4.6.1 La forme d'une distribution (comparaison des tendances centrales et coefficient d'asymétrie et d'aplatissement

✓ **Calcul du coefficient d'aplatissement:**

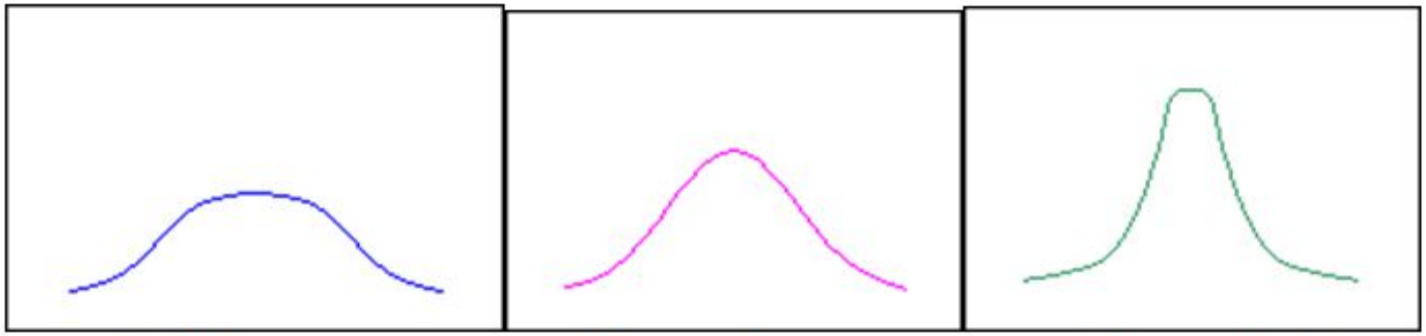
1) Coefficient de Pearson:

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4}$$

Si $\beta_2 = 3$ alors la distribution est normale

Si $\beta_2 > 3$ alors distribution est leptocurtique

Si $\beta_2 < 3$ alors distribution est platicurtique



Aplatie (platicurtique) | Normale | Concentrée (leptocurtique)

4.6 Les caractéristiques de concentration
4.6.2 Comparaison des médianes et indice de Gini

- ✓ **Définition :** La concentration est la comparaison deux séries de fréquences cumulées. (c'est un moyen de mesurer l'inégalité d'une distribution, ex l'inégalité des salaires etc…).

- ✓ ***Comparaison des médianes (médiane et médiale)***

Exemple : On étudie les salaires des salariés dans une entreprise

Salaires	xi	Effectif ni	fi	fi fcc	nixi	F(nixi) %	F(nixi) Fcc %
0-10	5	16	0,200	0,200	80	4,00	4,00
10-20	15	30	0,375	0,575	450	21,00	25,00
20-40	30	18	0,2250	0,800	540	25,00	50,00
40-70	55	10	0,125	0,925	550	25,00	75,00
70-100	85	6	0,0750	1,000	510	25,00	100,00
Total		**80**	**1,000**		**2130**		

4.6 Les caractéristiques de concentration
4.6.2 Comparaison des médianes et indice de Gini

$$C = \frac{Mé - Ml}{Intervalle\ Variation}$$

- Si **C** s'approche de **0** il y a faible concentration
- *Si **C** s'approche de **1** il y a forte concentration*

$$\frac{Ml - 20}{0.5 - 0.25} = \frac{40 - 20}{0.5 - 0.25} \rightarrow Ml = 40$$

*Et puisque **Me = 18** ce qui donne* $C = \frac{18 - 40}{90(étendue)} = 24.44\%$

4.6 Les caractéristiques de concentration
4.6.2 Comparaison des médianes et indice de Gini

✓ ***Calcul de l'indice de concentration (indice de gini):***

1) On trace la courbe de Lorentz

2) On mesure la concentration grâce à la courbe de Lorentz

$$G = \frac{Aire\ \text{concentration}}{\text{Aire du triangle}}$$ *or l'aire du triangle* ***OAB =1/2***

D'où $$G = 2 * Aire\ \text{concentration}$$

$$S = \frac{(b_1)h_1}{2} + \sum_{J=2}^{n} \frac{(b_j + B_j)h_j}{2}$$

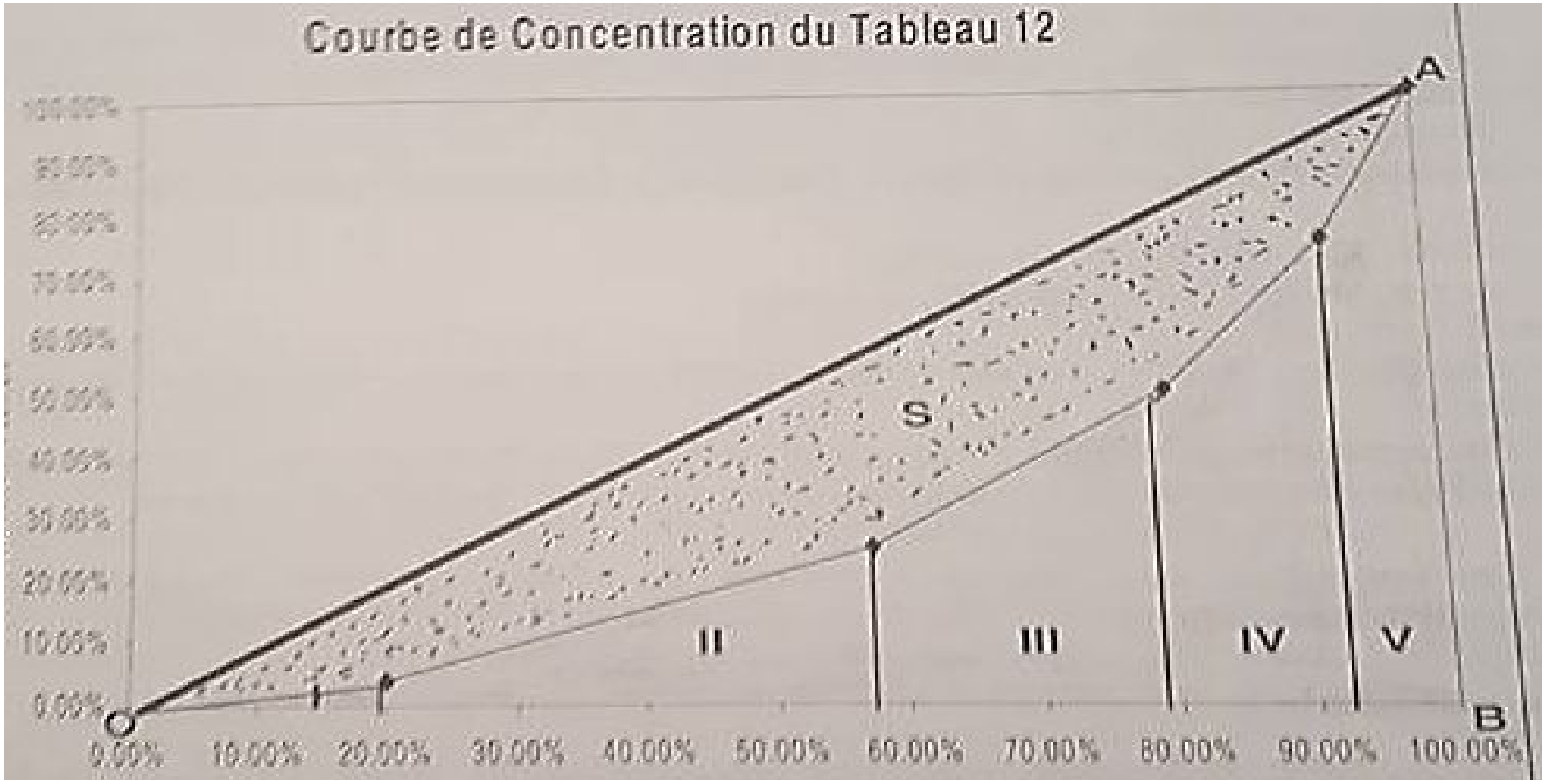
3) On calcul la surface totale (triangle + trapèzes)

4) $$concentration = (Aire\ Triangle) - (Surface\ totale)$$
$$= 1/2\ -\ Surface\ totaleAire$$

5) *Enfin*

$$G = 2 * Aire\ \text{concentration}$$

4.6 Les caractéristiques de concentration
4.6.2 Comparaison des médianes et indice de Gini



$$S = \frac{[(4 * 20) + (29 * 37.5) + (75 * 22.5) + (126 * 12.5) + (176 * 7.5)]}{2} = 2875$$

$$Concentration = 5000 - 2875 = 2125$$

$$G = \frac{2125}{5000} = 42.5\%$$

Chapitre 5
La réduction des données
(séries bivariées)

y(j) / x(i)	y_1	y_2	y_3	y_4	(7) $n_{i.}$	(8) $n_{i.}x_i$	(9) $n_{i.}x_i^2$	(10) Σ $n_{ij}y_j$	(11) Σ $n_{ij}y_j^2$	(12) Σ $n_{ij}x_iy_j$
x_1	n_{11}	n_{12}	n_{13}	n_{14}	$n_{1.}$	$n_{1.}x_1$	$n_{1.}x_1^2$	$\Sigma n_{1j}y_j$	$\Sigma n_{1j}y_j^2$	Σ $n_{1j}x_1y_j$
x_2	n_{21}	n_{22}	n_{23}	n_{24}	$n_{2.}$	$n_{2.}x_2$	$n_{2.}x_2^2$	$\Sigma n_{2j}y_j$	$\Sigma n_{2j}y_j^2$	Σ $n_{2j}x_2y_j$
x_3	n_{31}	n_{32}	n_{33}	n_{34}	$n_{3.}$	$n_{3.}x_3$	$n_{3.}x_3^2$	$\Sigma n_{3j}y_j$	$\Sigma n_{3j}y_j^2$	Σ $n_{3j}x_3\,y_j$
x_4	n_{41}	n_{42}	n_{43}	n_{44}	$n_{4.}$	$n_{4.}x_4$	$n_{4.}x_4^2$	$\Sigma n_{4j}y_j$	$\Sigma n_{4j}y_j^2$	Σ $n_{4j}x_4y_j$
x_5	n_{51}	n_{52}	n_{53}	n_{54}	$n_{5.}$	$n_{5.}x_5$	$n_{5.}x_5^2$	$\Sigma n_{5j}y_j$	$\Sigma n_{5j}y_j^2$	Σ $n_{5j}x_5y_j$
(1) $n_{.j}$	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{.4}$	$n_{..}$	$\Sigma =$	$\Sigma =$			$\Sigma\Sigma =$
(2) $n_{.j}y_j$	$n_{.1}y_1$	$n_{.2}y_2$	$n_{.3}y_3$	$n_{.4}y_4$	$\Sigma n_{.j}y_j$					
(3) $n_{.j}y_j^2$	$n_{.1}y_1^2$	$n_{.2}y_2^2$	$n_{.3}y_3^2$	$n_{.4}y_4^2$	Σ $n_{.j}y_j^2$					
(4) Σ $n_{ij}x_i$	$\Sigma n_{i1}x_i$	$\Sigma n_{i2}x_i$	$\Sigma n_{i3}x_i$	$\Sigma n_{i4}x_i$						
(5) Σ	$\Sigma n_{i1}x_i^2$	$\Sigma n_{i2}x_i^2$	$\Sigma n_{i3}x_i^2$	$\Sigma n_{i4}x_i^2$						

Exemple 1

	2	4	6	$n_{i.}$	$n_{i.}x_i$	$n_{i.}x_i^2$	$\Sigma n_{ij}yj$	$\Sigma n_{ij}y_j^2$	$\Sigma n_{ij}x_iy_j$
2	0 0	1 8	1 12	2	4	8	$\Sigma n_{1j}y_j$ = 10	$\Sigma n_{1j}y_j^2$ = 52	$\Sigma n_{1j}x_1y_j$ = 20
4	2 16	3 48	0 0	5	20	80	$\Sigma n_{2j}y_j$ = 16	$\Sigma n_{2j}y_j^2$ = 56	$\Sigma n_{2j}x_2y_j$ = 64
6	1 12	1 24	1 36	3	18	108	$\Sigma n_{3j}y_j$ = 12	$\Sigma n_{3j}y_j^2$ = 56	$\Sigma n_{3j}x_3y_j$ = 72
$n_{.j}$	3	5	2	Σ= **10**	Σ= **42**	Σ= **196**			$\Sigma\Sigma$ = **156**
$n_{.j}y_j$	6	20	12	Σ = **38**					
$n_{.j}y_j^2$	12	80	72	Σ= **164**					
$\Sigma n_{ij}x_i$	$\Sigma n_{i1}x_i$ = 14	$\Sigma n_{i2}x_i$ = 20	$\Sigma n_{i3}x_i$ = 8						
$\Sigma\ n_{ij}x_i^2$	$\Sigma n_{i1}x_i^2$ = 68	$\Sigma n_{i2}x_i^2$ = 88	$\Sigma n_{i3}x_i^2$ = 40						
$\Sigma n_{ij}x_iy_j$	Σ $n_{i1}x_iy_1$ = 28	$\Sigma n_{i2}x_iy_2$ = 80	$\Sigma n_{i3}x_iy_3$ = 48	$\Sigma\ \Sigma$= **156**					

5.1 Paramètres d'une série bivariée
5.1.1 Paramètres marginales de X et Y

$$\bar{\bar{x}} = \frac{1}{n_{..}} \sum_{i=1}^{k} n_{i.} x_i = \frac{42}{10} = 4{,}2$$

$$\bar{\bar{y}} = \frac{1}{n_{..}} \sum_{j=1}^{l} n_{.j} y_j = \frac{38}{10} = 3{,}8$$

$$V(x) = \frac{1}{n_{..}} \sum_{i=1}^{k} n_{i.} x_i^2 - \bar{\bar{x}}^2 = \frac{196}{10} - (4{,}2)^2 = 1{,}96$$

Ou

$$V(x) = \frac{1}{n_{..}} \sum_{i=1}^{k} n_{i.} (x_i - \bar{\bar{x}})^2$$

$$V(y) = \frac{1}{n_{..}} \sum_{j=1}^{l} n_{.j} y_j^2 - \bar{\bar{y}}^2 = \frac{164}{10} - (3{,}8)^2 = 1{,}96$$

Ou

$$V(y) = \frac{1}{n_{..}} \sum_{j=1}^{p} n_{.j} (y_j - \bar{\bar{Y}})^2$$

5.1 Paramètres d'une série bivariée
5.1.2 Paramètres conditionnels de X et Y

$$\bar{x}_1 = \frac{1}{n_{.1}} \sum_{i=1}^{3} n_{i1} x_i = \frac{14}{3} = 4{,}67 \qquad \bar{y}_1 = \frac{1}{n_{1.}} \sum_{j=1}^{3} n_{1j} y_j = \frac{10}{2} = 5$$

$$\bar{x}_2 = \frac{1}{n_{.2}} \sum_{i=1}^{3} n_{i2} x_i = \frac{20}{5} = 4 \qquad \bar{y}_2 = \frac{1}{n_{2.}} \sum_{j=1}^{3} n_{2j} y_j = \frac{16}{5} = 3{,}2$$

$$\bar{x}_3 = \frac{1}{n_{.3}} \sum_{i=1}^{3} n_{i3} x_i = \frac{8}{2} = 4 \qquad \bar{y}_3 = \frac{1}{n_{3.}} \sum_{j=1}^{3} n_{3j} y_j = \frac{12}{3} = 4$$

$$V_1(x) = \frac{1}{n_{.1}} \sum_{i=1}^{3} n_{i1} x_i^2 - \bar{x}_1{}^2 = \frac{68}{3} - (4{,}66)^2 = 0{,}96$$

5.1 Paramètres d'une série bivariée
5.1.2 Paramètres conditionnels de X et Y

$$V_2(x) = \frac{1}{n_{.2}} \sum_{i=1}^{3} n_{i2} x_i^2 - \bar{x}_2{}^2 = \frac{88}{5} - (4)^2 = 1{,}6$$

$$V_3(x) = \frac{1}{n_{.3}} \sum_{i=1}^{3} n_{i3} x_i^2 - \bar{x}_3{}^2 = \frac{40}{2} - (4)^2 = 4$$

$$V_1(y) = \frac{1}{n_{1.}} \sum_{j=1}^{3} n_{1j} y_j^2 - \bar{y}_1{}^2 = \frac{52}{2} - (5)^2 = 1$$

$$V_2(y) = \frac{1}{n_{2.}} \sum_{j=1}^{3} n_{2j} y_j^2 - \bar{y}_2{}^2 = \frac{56}{5} - (3{,}2)^2 = 0{,}96$$

$$V_3(y) = \frac{1}{n_{3.}} \sum_{j=1}^{3} n_{3j} y_j^2 - \bar{y}_3{}^2 = \frac{56}{3} - (4)^2 = 2{,}67$$

$$Cov(x,y) = \frac{1}{n_{..}} \sum_{i=1}^{3} \sum_{j=1}^{3} n_{ij} x_i y_j - \bar{\bar{x}}\bar{\bar{y}} = \left(\frac{1}{10}.156\right) - (4{,}2)(3{,}8)$$

$$= -0{,}36$$

5.1 Paramètres d'une série bivariée
5.1.3 Covariance par changement de variable

Soit

$$x'_i = \frac{x_i - x_0}{\alpha}, \quad y'_j = \frac{y_j - y_0}{\beta}$$

$$x_i = \alpha x'_i + x_0, \quad y_j = \beta y'_j + y_0$$

$$\bar{\bar{x}} = \alpha \bar{\bar{x}}' + x_0 \; ; \; \bar{\bar{y}} = \beta \bar{\bar{y}}' + y_0$$

Par soustraction

$$x_i - \bar{\bar{x}} = \alpha\left(x'_i - \bar{\bar{x}}'\right) \quad ; \; y_j - \bar{\bar{y}} = \beta\left(y'_j - \bar{\bar{y}}'\right)$$

Par conséquent

$$\begin{aligned} Cov(x,y) &= \frac{1}{n_{..}} \sum_i \sum_j n_{ij}(x_i - \bar{\bar{x}})(y_j - \bar{\bar{y}}) \\ &= \alpha\beta \frac{1}{n_{..}} \sum_i \sum_j n_{ij}(x'_i - \bar{x}')\beta(y'_j - \bar{\bar{y}}') \\ &= \alpha\beta Cov(x', y') \end{aligned}$$

Chapitre 6
Les indices statistiques

6.1 Les indices statistiques
6.1.1 Les indices simples

❑ Propriétés des indices élémentaires:

1) La circularité:

$$I_{t/0}(G) = \frac{I_{t/t'}(G)}{I_{t'/0}(G)} \qquad I_{t/t'}(G) = \frac{I_{t/0}(G)}{I_{t'/0}(G)}$$

- La circularité est une propriété fondamentale qui permet de comparer non seulement les dates 0 et t d'une part, 0 et t' d'autre part mais également t et t':

2) La réversibilité:

$$I_{0/t}(G) = \frac{1}{I_{t/0}(G)}$$

- La réversibilité est importante surtout lorsqu'on se réfère à un autre critère autre que le temps. Exemple comparer la même grandeur dans deux espaces différents.

3) L'enchainement:

$$I_{t/0}(G) = I_{t/t-1}(G) * I_{t-1/t-2}(G) * \ldots . * I_{1/0}(G)$$

- On obtient ainsi l'indice de la date t par rapport à la date 0 en faisant le produit des indices intermédiaires d'une date par rapport à la date précédente.

6.1 Les indices statistiques
6.1.2 Les indices synthétiques

- **Définition** : Un indice synthétique est une grandeur composite qui résume un ensemble d'indices simples basés sur des grandeurs hétérogènes (i.e. ne pouvant pas être additionnées). **Exemple** : L'indice des prix à la production agrégé.

✓ Soit une grandeur G composé de plusieurs éléments (ex: indice des prix):

$$G^1, G^2, G^3, \dots, G^i, \dots, G^k$$

✓ Les indices élémentaires sont définis par: $$I_{t/0}(G^i) = \frac{G_t^i}{G_0^i} * 100$$

✓ Les indices synthétiques élémentaires pondérées par les coefficients:

- W_0^i : L'importance relative du constituant i dans la grandeur G à la date de base 0. $$\sum_i W_0^i = 1$$
- W_t^i : L'importance relative du constituant i dans la grandeur G à la date courante: t. $$\sum_i W_t^i = 1$$

6.1 Les indices statistiques
6.1.2 Les indices synthétiques

1) L'indice Lapeyre:

- L'indice de Laspeyres est **la moyenne arithmétique pondérée** des indices élémentaires par les coefficients **w** de la date de référence (date de base: 0):

$$\mathrm{L}_{t/0}(\mathrm{G}) = \sum_i w_0^i I_{t/0}(G^i) \quad \text{d'où} \quad \mathrm{L}_{t/0}(\mathrm{G}) = \sum_i w_0^i * \frac{G_t^i}{G_0^i}$$

2) L'indice Paashe :

- L'indice de Paasche est **la moyenne harmonique** des indices élémentaires **pondérée** par les coefficients **W** à la date courante (date : t)

$$\mathrm{P}_{t/0}(\mathrm{G}) = \frac{1}{\sum_i w_t^i * \frac{G_0^i}{G_t^i}}$$

3) L'indice de Fisher:

- L'indice de Fisher est la moyenne géométrique simple des indices de Laspeyres et de Paasche.

$$F_{t/0}(\mathrm{G}) = \sqrt{L_{t/0}(G) * P_{t/0}(G)}$$

6.1 Les indices statistiques
6.1.2 Les indices synthétiques

- **Exemple:** Le tableau suivant présente des données observées sur les importations du Maroc en matières premières pendant les trois dernières années dans le secteur du textile:

	2010		2011	
Mat Prem	Poids	Montant	Poids	Montant
Fil	10%	500	5%	600
Tissu	70%	1000	80%	1500
Fournitures	20%	3000	15%	2700

1) Indices élémentaires des matières premières de 2011 par rapport à 2010:

Fil $I_{11/10}(M^1) = \frac{M_{11}^1}{M_{10}^1} * 100 = \frac{600}{500} * 100 = 120$ (augmentation de 20%)

Tissu $I_{11/10}(M^2) = \frac{M_{11}^2}{M_{10}^2} * 100 = \frac{1500}{1000} * 100 = 150$ (augment de 50%)

Fournitures $I_{11/10}(M^3) = \frac{M_{11}^3}{M_{10}^3} * 100 = \frac{2700}{3000} * 100 = 90$

6.1 Les indices statistiques
6.1.2 Les indices synthétiques

1) Indice Paashe:

- Les coefficients sont ceux de la date courante : 2011

$$P_{11/10}(M) = \frac{1}{\sum_i w_{11}^i * \frac{M_{10}^i}{M_{11}^i}} \qquad P_{11/10}(M) = \frac{1}{0.05 * \frac{500}{600} + 0.80 * \frac{1000}{1500} + 0.15 * \frac{3000}{2700}}$$

$P_{11/10}(M) = 134.83$ (une augmentation des importation de 34,83%)

2) Indice Lapeyre:

$$F_{11/10}(M) = \sqrt{L_{11/10}(M) * P_{11/10}(M)} \qquad F_{11/10}(M) = \sqrt{135 * 134.83}$$

$F_{11/10}(M) = 134.91$ (une augmentation des importation de 34,91%)

2) Indice Fisher:

$$F_{11/10}(M) = \sqrt{L_{11/10}(M) * P_{11/10}(M)} \qquad F_{11/10}(M) = \sqrt{135 * 134.83}$$

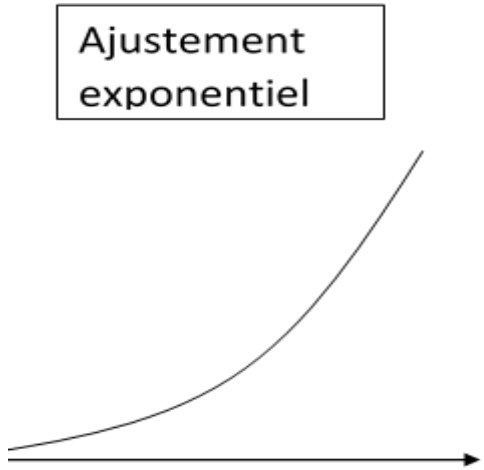
$$F_{11/10}(M) = 134.91$$

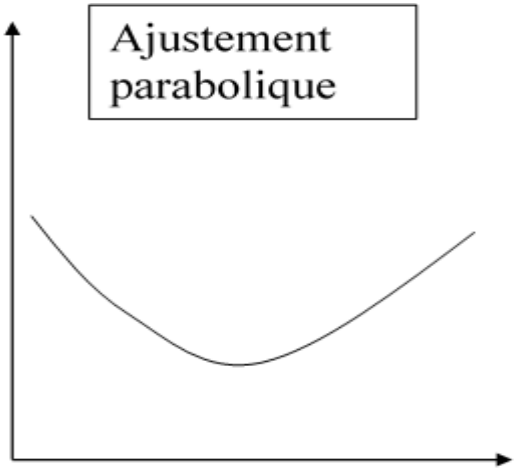
Chapitre 7

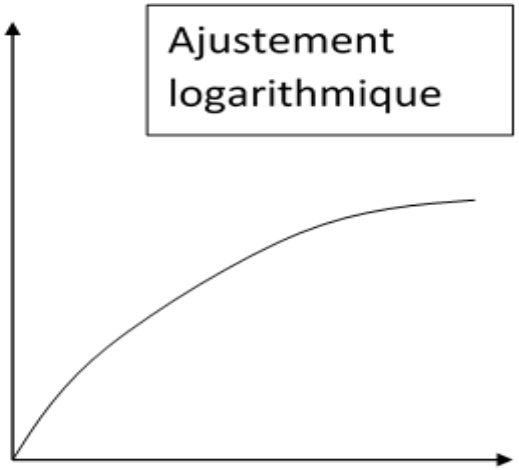
Relations entre variables : régressions et corrélations

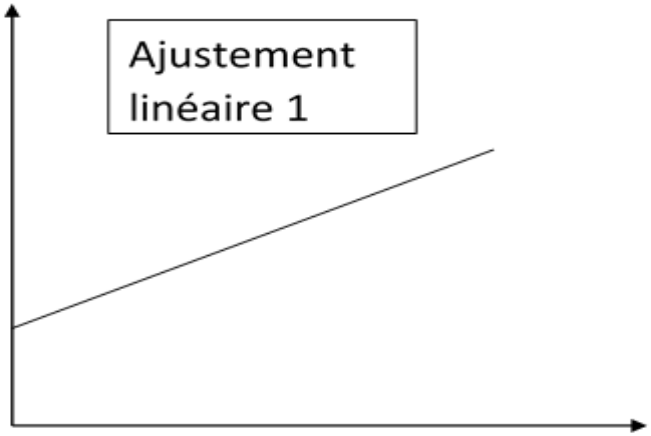
7.1 Régression et corrélation
7.1.1 Introduction

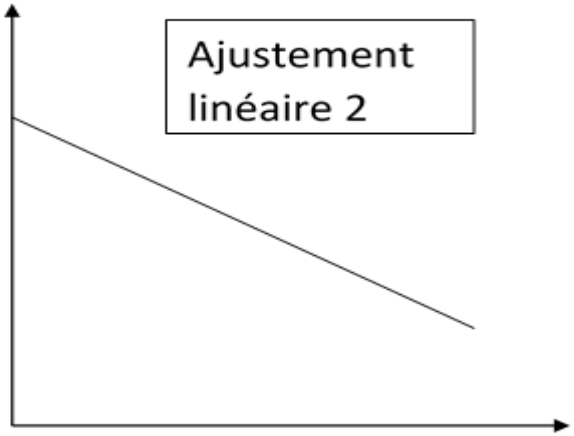
- ✓ **Définition** Les courbes de régressions (ajustements) sont un moyen graphique de synthétiser la liaison existante entre deux variables (ou le nuage de points formé par ces deux variables). (Ex quelques ajustements***)***





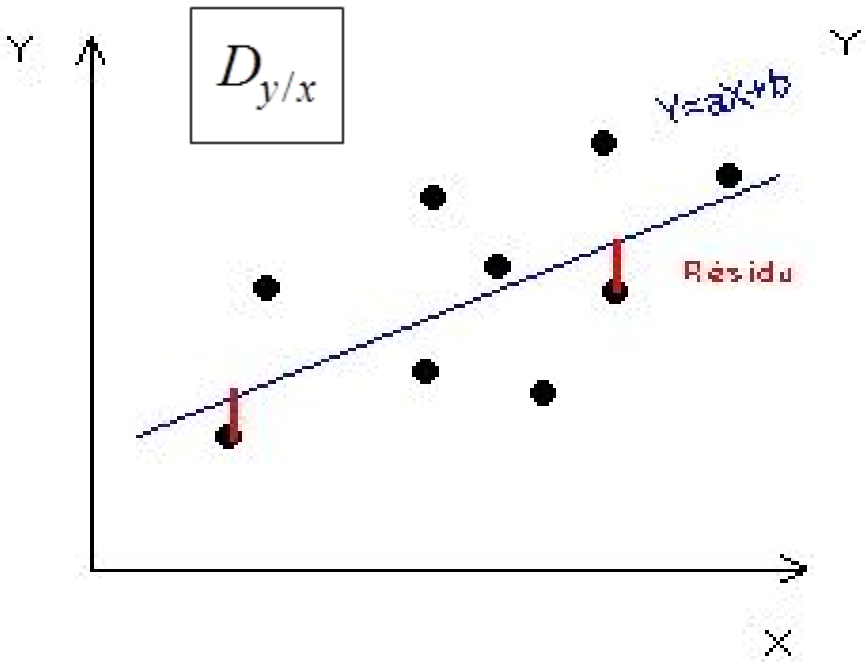


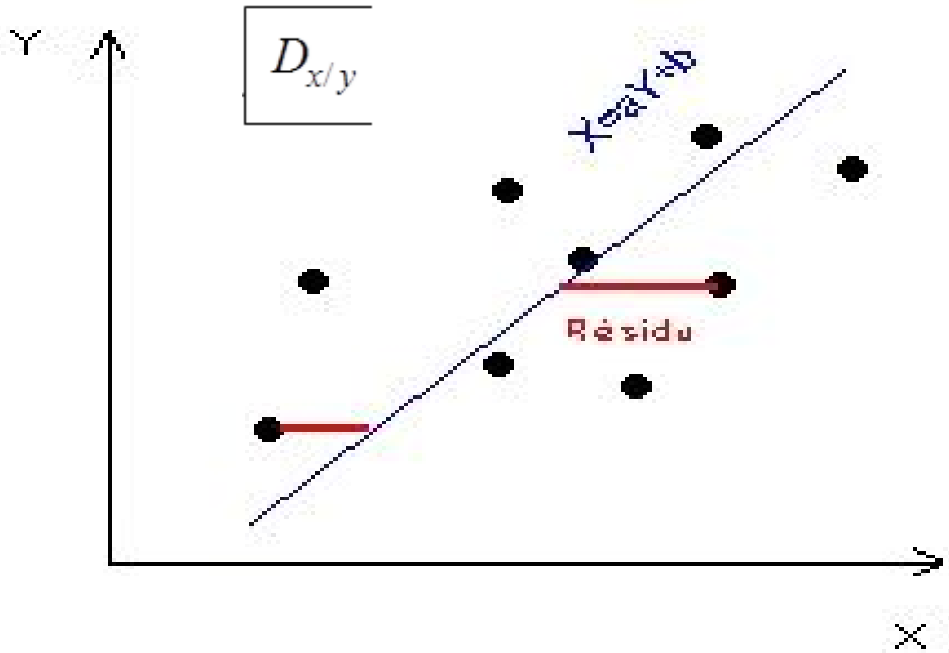




7.1 Régression et corrélation
7.1.2 La méthode des moindres carrées (MCO)

- ✓ **Définition:** la méthode linéaire des moindres carrés un nuage de points par deux droites possibles qui lient y à x, **tel que la distance entre le nuage de points et chaque droite est <u>minimale</u>**.





Objectif: choix de $\hat{y}_i = \hat{a}x_i + \hat{b}$ $(D_{y/x})$ tel que: $\sum_i \left(y_i - \hat{y}_i\right)^2$ est un mimimum	Objectif: choisir $\hat{x}_i = \hat{a}' y_i + \hat{b}'$ $(D_{x/y})$ tel que: $\sum_i \left(x_i - \hat{x}_i\right)^2$ est un minimum.

7.1 Régression et corrélation
7.1.2 La méthode des moindres carrées (MCO)

- ✓ **Démonstration pour Dy/x** (la démo pour s'obtient symétriquement):
- ✓ **Objectif:** Choisir une droite qui minimise les écarts des observations à cette droite (droite passant par le milieu du nuage de points).

$$Min\ \frac{1}{n}\sum_i \left(y_i - \hat{y}_i\right)^2 = \frac{1}{n}\sum_i \left(y_i - \hat{a}x_i - \hat{b}\right)^2$$

<u>Dériver / b:</u>

$$f'(b) = \frac{df}{db} = (-1).2\frac{1}{n}\sum_i \left(y_i - ax_i - b\right) = 0$$ (dériver par rapport à b: $u^2 \rightarrow 2uu'$)

$$\Leftrightarrow \frac{\sum_i y_i}{n} = a\frac{\sum_i x_i}{n} + b$$

$$\Leftrightarrow \quad \bar{y} = a\bar{x} + b \quad (1)$$

(résultat 'en passant': la droite doit passer par les points moyens)

Dériver / a:

$$\text{et } f'(a) = \frac{df}{da} = \frac{1}{n}\sum_i (-x_i).2\left(y_i - ax_i - b\right) = 0 \quad (2)$$

en remplaçant b par son expression (1), dans l'équation (2), on a:

$$f'(a) = \frac{df}{da} = -2.\frac{1}{n}\sum_i x_i.\left(y_i - ax_i - \bar{y} + a\bar{x}\right) = 0$$

Arranger les termes:

$$f'(a) = \frac{df}{da} = -2.\frac{1}{n}\sum_i x_i.\left((y_i - \bar{y}) - a(x_i - \bar{x})\right) = 0$$

$$\Leftrightarrow \hat{a} = \frac{\left(\sum x_i(y_i - \bar{y}) \big/ n\right)}{\left(\sum\sum x_i(x_i - \bar{x}) \big/ n\right)} = \frac{\frac{\sum x_i y_i}{n} - \overline{xy}}{\frac{\sum x_i^2}{n} - \bar{x}^2} = \frac{Cov(x,y)}{Var(x)} \quad (3)$$

7.1 Régression et corrélation
7.1.2 La méthode des moindres carrées (MCO)

Dériver / a:

Ainsi:

- **Tracé de** $D_{y/x}$ **correspond à la droite:** $\hat{y}_i = \hat{a}x_i + \hat{b}$ **avec** $\hat{a} = \frac{Cov(x,y)}{Var(x)}$ **et** $\hat{b} = \bar{y} - \hat{a}\bar{x}$

- **Par démonstration symétrique: Tracé de** $D_{x/y}$ **correspond à**

$\hat{x}_j = \hat{a}'y_j + \hat{b}'$ **(ou encore :** $y_j = \frac{1}{\hat{a}'}x_i + \frac{1}{\hat{b}'}$ **)** **avec** $\hat{a}' = \frac{Cov(x,y)}{Var(y)}$ **et** $\hat{b}' = \bar{x} - \hat{a}'\bar{y}$

✓ **Le coefficient de corrélation :** (-1< = r < = 1)

$$r = (a.a')^{1/2} = \left(\frac{Cov(x,y)^2}{Var(x).Var(y)} \right)^{1/2} = \left(\frac{Cov(x,y)}{\sigma_x.\sigma_y} \right)$$

- si *r* est proche de 0, il **n'y a pas de relation linéaire** entre X et Y
- si *r* est proche de -1, il existe **une forte relation linéaire négative** entre X et Y
- si *r* est proche de 1, il existe **une forte relation linéaire positive** entre X et Y

7.1 Régression et corrélation
7.1.2 La méthode des moindres carrées (MCO)

✓ **Le coefficient de corrélation :** (0<= r < = 1)

Le coefficient de détermination (appelé r^2) est le carré du coefficient de corrélation 'r'.

<u>Proposition:</u>

r^2 mesure le part de la variabilité totale de y expliquée par x (*ou encore* par la droite de régression).

- Plus r^2 est grand (tend vers 1) et mieux la droite $\hat{y}=\hat{a}x+\hat{b}$ résume le nuage de points (y). → évolution de x décrit bien celle de y.
- Qd. r^2 est petit (tend vers 0) → évolution de x semble être indépendante de celle de y.

Chapitre 8:

Les séries chronologiques

8.1 Les séries chronologiques
8.1 Introduction

= ✓ **Définition** : On appelle « série chronologique » toutes suite temporelle d'observations chiffrées, les observation sont effectuées à des intervalles de temps réguliers (années, mois, jours,…).

✓ Une SC comporte quartes composantes (mouvements):

- Une **composante Extra saisonnière** : il s'agit d'un mouvement à long terme qualifié par trend de la série chronologique (Ti)
- Une **composante cyclique** où l'amplitude d'un mouvement est variable pouvant dépasser l'année (Ci)
- Une **composante saisonnière** : des fluctuations périodiques peuvent apparaître à l'intérieur de l'année et qui peuvent se répéter chaque année à la même période (Si)
- - Une **composante aléatoire** ou imprévisible où l'intensité de variation est réduite. (Ai).

✓ Il existe 2 méthode d'une chronique:

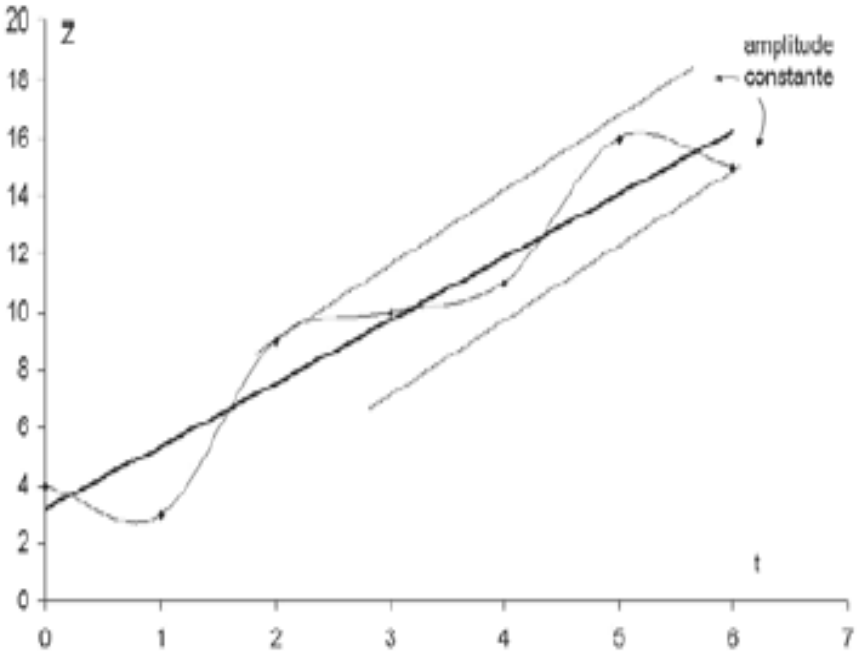
➢ **Le modèle additif :** Yi= Ti + Ci + Si + Ai

➢ **Le modèle multiplicatif :** Yi= Ti * Ci * Si *Ai (plus fréquent)

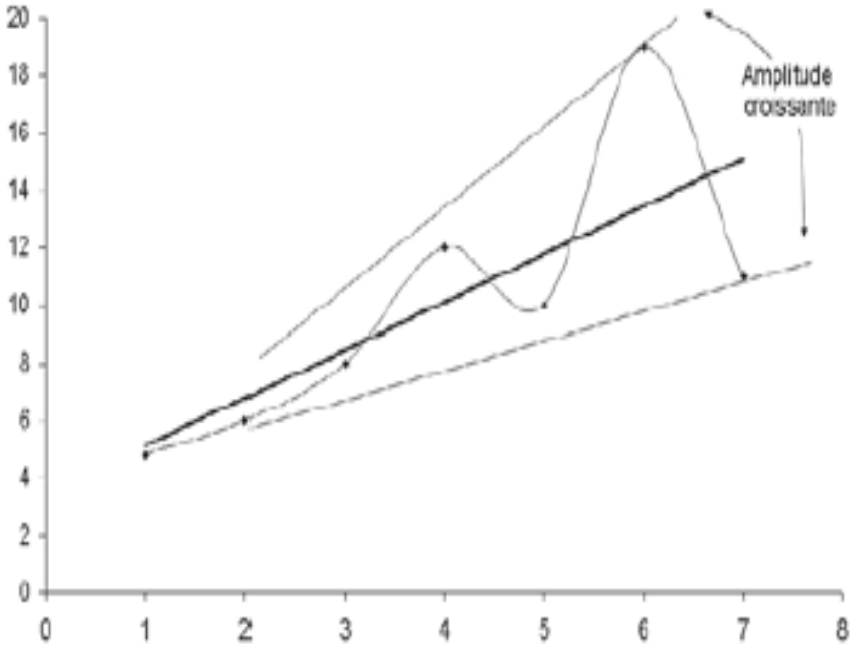
8.1 Les séries chronologiques
8.1.2 Les types de modèles de la SC

- **Les modèles additif et multiplicatif**

(a) Modèle additif



(b) Modèle multiplicatif



8.1 Les séries chronologiques
8.1.2 Périodicité

- ✓ Une SC peut être (**journalières, mensuelles, trimestrielles**…)

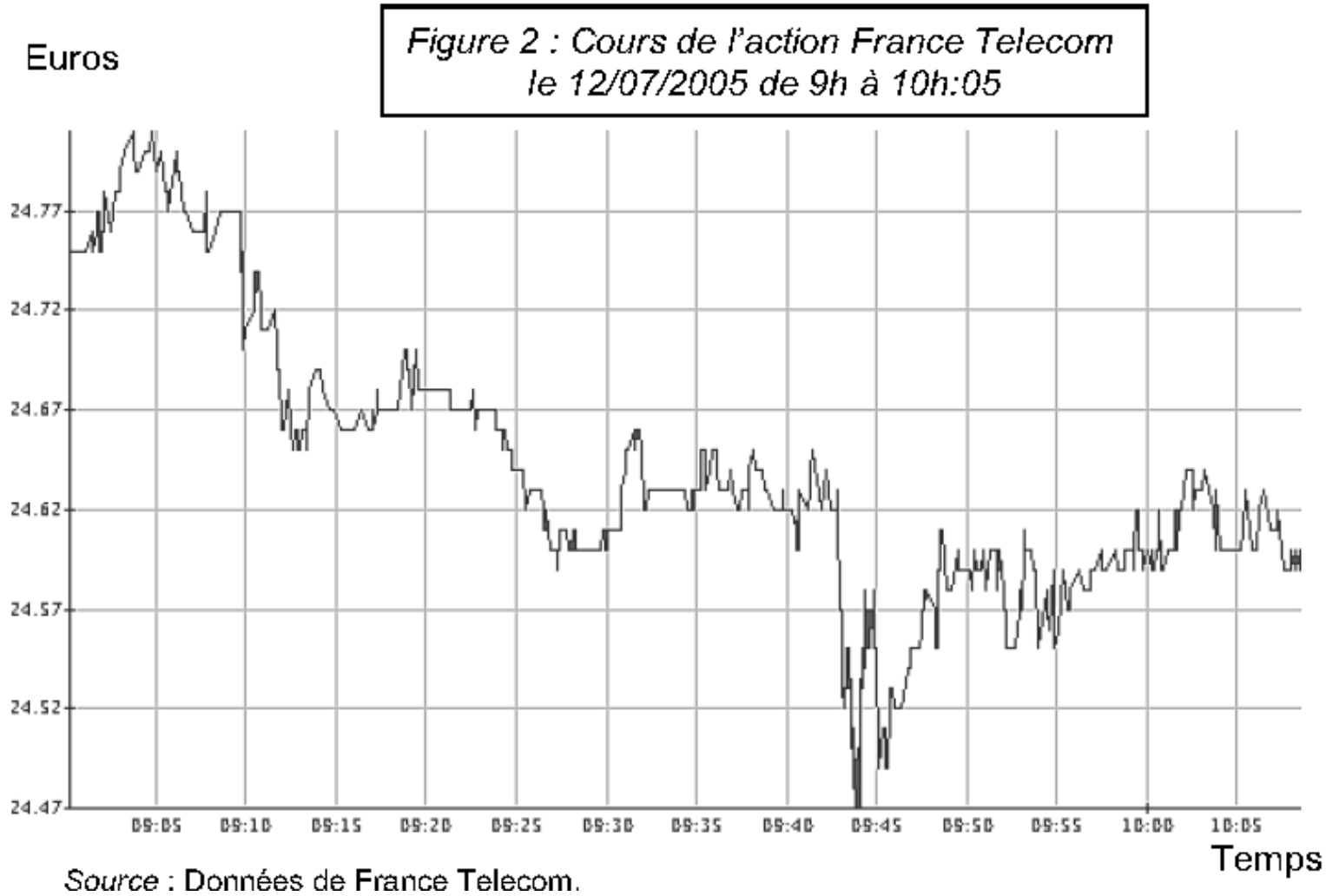


Figure 2 : Cours de l'action France Telecom le 12/07/2005 de 9h à 10h:05

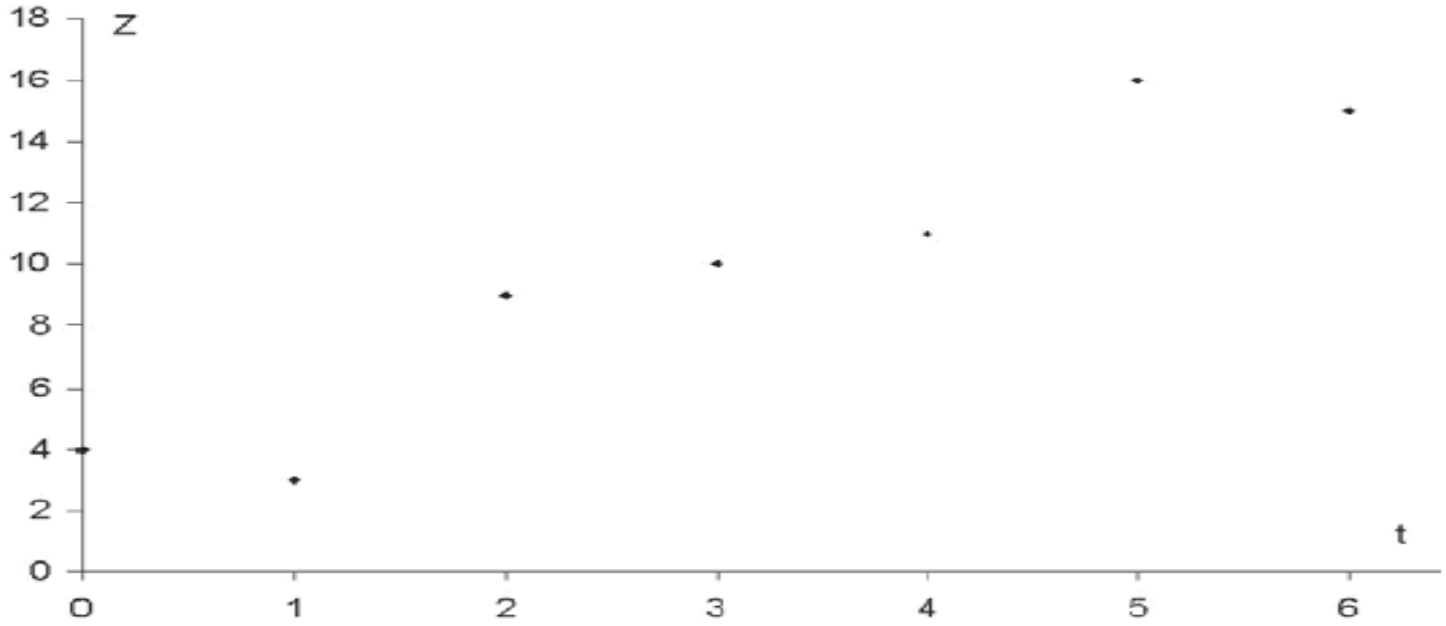
Source : Données de France Telecom.

8.1 Les séries chronologiques
8.1.3 Détermination de la tendance d'une SC

- ✓ **Détermination par régression linéaire (MC0):**

Exemple: soit la SC suivante repéré par l'indice t

t	0	1	2	3	4	5	6
z	4	3	9	10	11	16	15



8.1 Les séries chronologiques
8.1.3 Détermination de la tendance d'une SC

=

t_i	z_i	$z_i t_i$	t_i^2	$z_i t_i^2$
0	4	0	0	0
1	3	3	1	3
2	9	18	4	36
3	10	30	9	90
4	11	44	16	176
5	16	80	25	400
6	15	90	36	540
21	**68**	**265**	**91**	**1245**

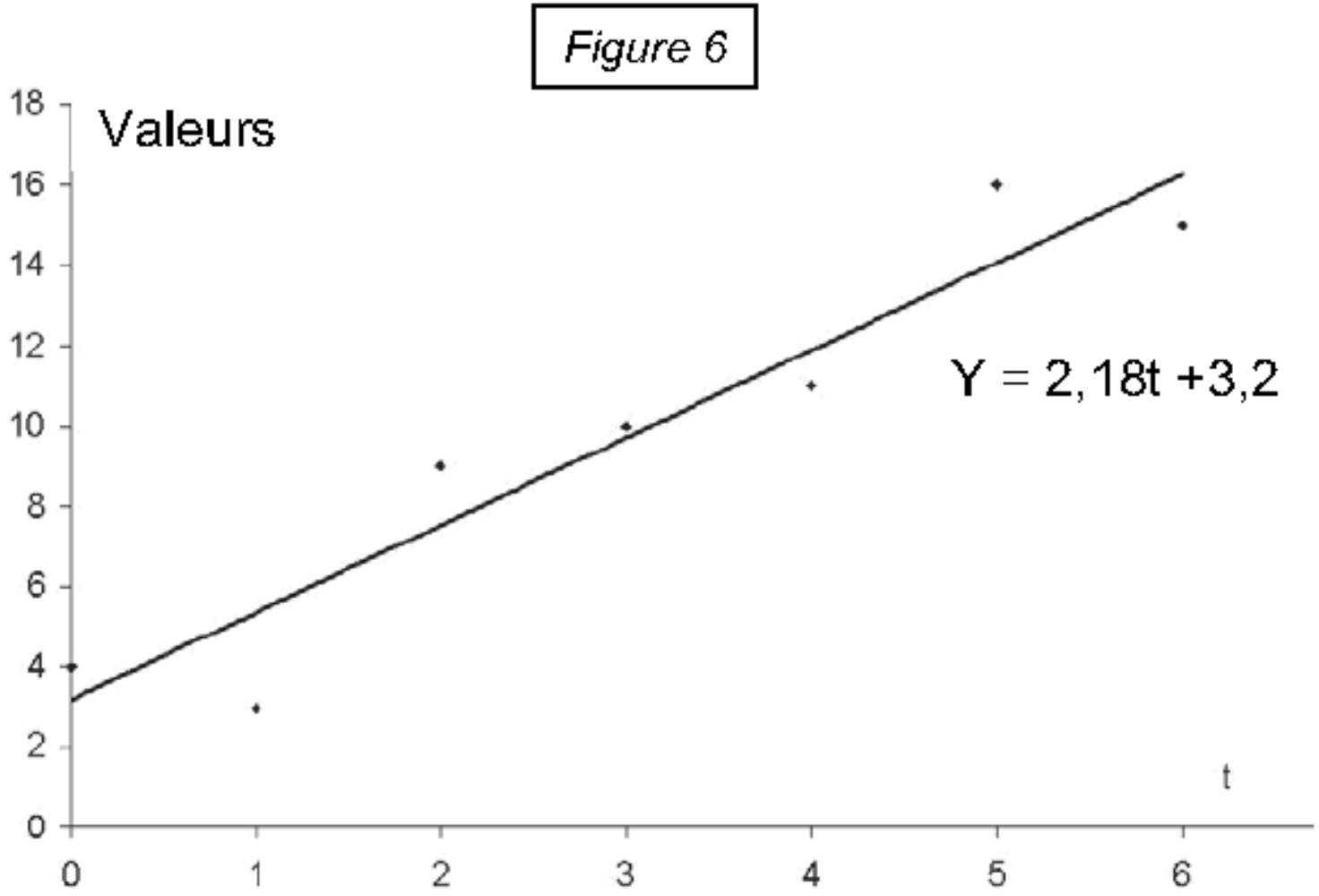
$$a = \frac{\sum_i z_i t_i \quad n\overset{=}{t}\overset{=}{z}}{\sum_i t_i^2 \quad n\left(\overset{=}{t}\right)^2} = \frac{265 \quad 7\times 3\times 9,714}{91 \quad 7\times 3^2} = \frac{61}{28} = 2,18 \qquad b = \overset{=}{z} \quad a\overset{=}{t} = 9,714 \quad 2,17\times 3 = 3,2$$

$$f_t = at + b = 2,17 + 3,2$$

8.1 Les séries chronologiques
8.1.3 Détermination de la tendance d'une SC

- ✓ **Exemple les deux moyens mobiles d'ordre 2**

=



8.1 Les séries chronologiques
8.1.3 Détermination de la tendance d'une SC

✓ **Détermination par la méthode des moyennes mobiles:**

=

Date	Cours de clôture(€)	Ordre 2	Ordre 3
13/07/2005	24,66		
12/07/2005	24,61	24,635	
11/07/2005	24,73	24,67	24,666
08/07/2005	24,53	24,63	24,623
07/07/2005	24,01	24,27	24,42
06/07/2005	24,16	24,09	24,23
05/07/2005	24	24,08	24,06
04/07/2005	24,18	24,09	24,11
01/07/2005	24,27	24,23	24,15
30/06/2005	24,16	24,22	24,20
29/06/2005	23,8	23,98	24,08
28/06/2005	22,6	23,20	23,52
27/06/2005	22,58	22,59	22,99
24/06/2005	22,66	22,62	22,61
23/06/2005	22,93	22,80	22,72
22/06/2005	22,97	22,95	22,85
21/06/2005	23,02	23,00	22,97
20/06/2005	22,85	22,94	22,95
17/06/2005	22,94	22,90	22,94
16/06/2005	22,68	22,81	22,82
15/06/2005	22,48	22,58	22,70
14/06/2005	22,6	22,54	22,59
13/06/2005	22,74	22,67	22,61

8.1 Les séries chronologiques
8.1.3 Détermination de la tendance d'une SC

- ✓ **Exemple des deux moyens mobiles d'ordre 2**

=

$$\frac{24,66+24,61}{2}=24,635 \qquad \frac{24,61+24,73}{2}=24,67$$

- ✓ **Exemple des deux moyens mobiles d'ordre 3**

$$\frac{24,66+24,61+24,73}{3}=24,666 \qquad \frac{24,61+24,73+24,53}{3}=24,623$$

8.1 Les séries chronologiques
8.1.3 Détermination de la tendance d'une SC

=

Figure 7 (a) : Moyenne mobile d'ordre 2

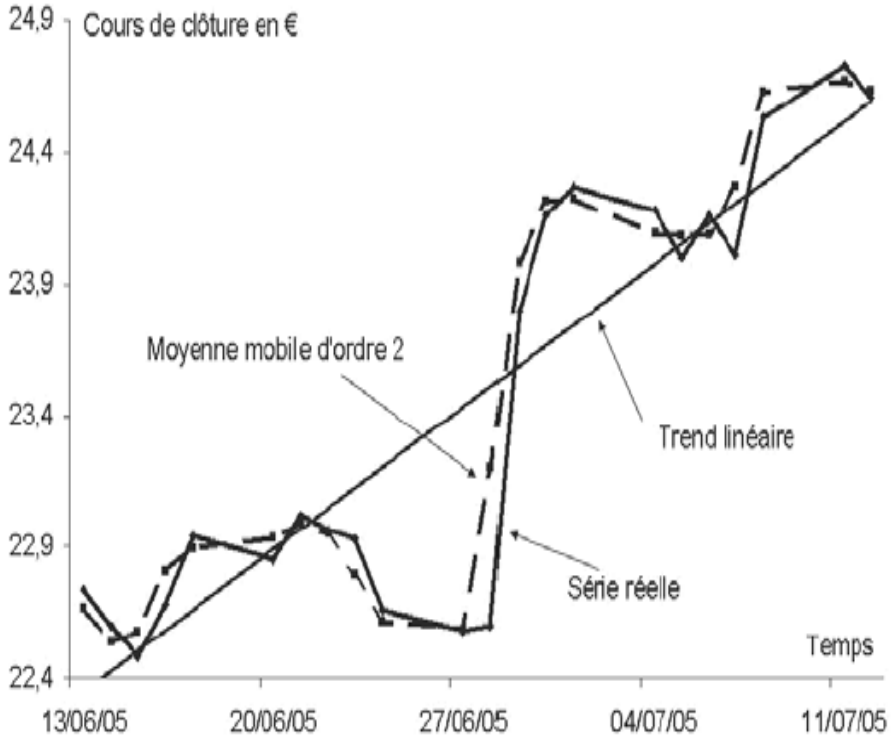
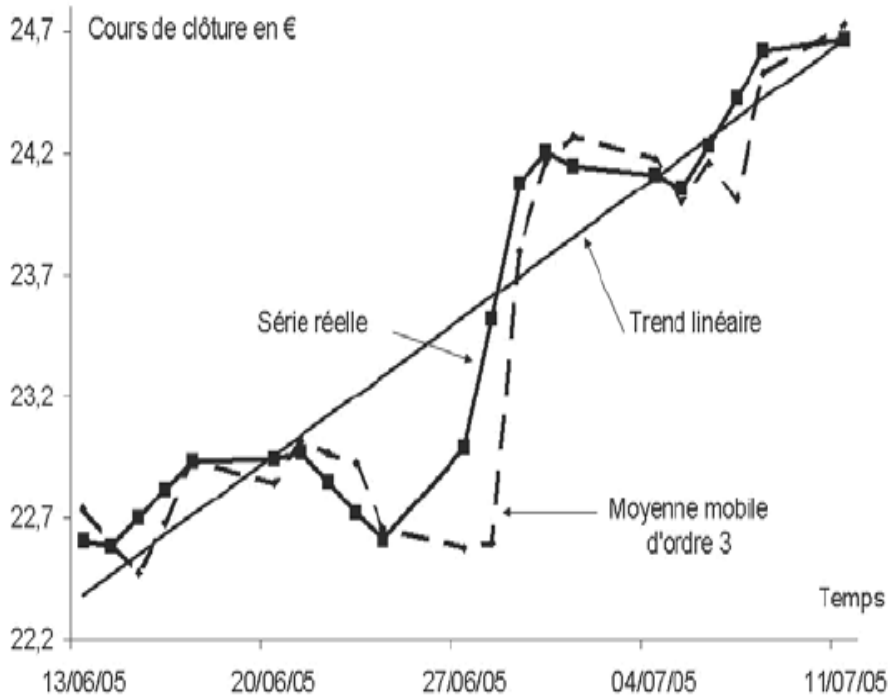


Figure 7 (b) : Moyenne mobile d'ordre 3



8.1 Les séries chronologiques
8.1.4 Détermination des coefficients saisonniers

✓ **Exemple:** on dispose des ventes trimestrielles de 4 années d'un magasin de vente de pièces mécaniques:

	Les ventes historiques (en Trimestres)			
les années	1	2	3	4
2003	3200	3250	3640	3100
2004	3150	3350	3100	3250
2005	3250	3450	3500	3250
2006	3150	3450	3350	3250

- La droite d'ajustement est de la forme Y = 3,22.X + 3265.75

	Les valeurs ajustés Y' (Trend)			
les années	1	2	3	4
2003	3268,97	3281,85	3294,74	3307,62
2004	3272,19	3285,07	3297,96	3310,84
2005	3275,41	3288,29	3301,18	3314,06
2006	3278,63	3291,51	3304,40	3317,28

- coefficients saisonniers (CS) sont les moyennes arithmétiques des rapport au trend (Y /Y') par période (trimestre, mois...)

8.1 Les séries chronologiques
8.1.4 Détermination des coefficients saisonniers

=

	Valeurs réels valeurs ajustée			
les années	1	2	3	4
2003	0,98	0,99	1,10	0,94
2004	0,96	1,02	0,94	0,98
2005	0,99	1,05	1,06	0,98
2006	0,96	1,05	1,01	0,98
Somme	3.98	4.11	4.11	3.88
Coefficients saisonniers	**0.995**	**1.0275**	**1.0275**	**0.97**

	série corrigée des variations saisonnières Y' /CS			
les années	1	2	3	4
2003	3284.42	3194,01	3206,56	3409,92
2004	3288.63	3197,15	3209,69	3413,24
2005	3291.86	3200,28	3212,83	3416,56
2006	3278,63	3203,42	3215,96	3419,88

	Les prévisions des ventes de l'année 2007			
Eléments	1	2	3	4
prévisions	3320,49	3323,71	3326,93	3330,15
Prévisions corrigés Des effets saisonniers	3303.88	3415.11	3418.42	3230.24

Bibliographie (indicative)

- Adil El Marhoum et Adil El Abassi, Statistiques descriptives, Université Cadi Ayyad, 2000
- Bernard **Grais**, Statistique descriptive, Dunod, 2000.
- Bernard **Py**, Statistique descriptive, Economica, 2009.
- J. **Hubler**, Statistiques appliquées à l'économie, Bréal, 1996.
- Alain **Piller**, Statistique descriptive, Premium Editeur, 2004.