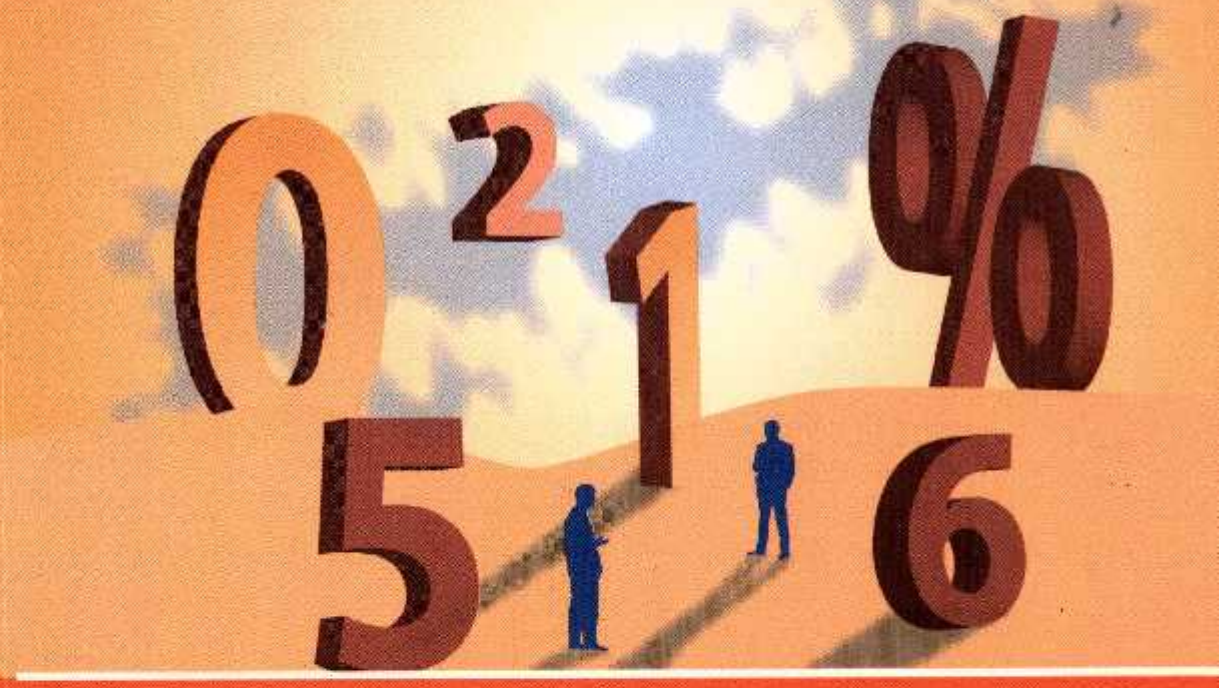
ENCYCLOPÉDIE DE MATHÉMATIQUES

# STATISTIQUES DESCRIPTIVES

## COURS ET EXERCICES




CENTRE DE RECHERCHE EN MATHEMATIQUES DE L'IGA

Adil ELMARHOUM
INGÉNIEUR D'ÉTAT

Mohamed DIOURI
DOCTEUR INGENIEUR



LES EDITIONS TOUBKAL

IGA

INSTITUT SUPERIEUR DU GENIE APPLIQUE

**Mohamed DIOURI**
Docteur Ingénieur
Président Fondateur de l'IGA

**Adil ELMARHOUM**
Docteur en statistique et informatique appliquée
Professeur Habilité Université Mohamed V Agdal

# STATISTIQUE DESCRIPTIVE

## Cours et exercices

**Collection Sciences Techniques et Managements des éditions TOUBKAL**
**Publications du Centre de Recherche en Mathématiques (CRM) de l'IGA**

# STATISTIQUE DESCRIPTIVE

## Cours et exercices


Dépôt légal N°2006/2774
I.S.S.N. 9954 – 496 – 03 – 3

Les livres de la collection Sciences Techniques et Management
sont co-édités par les éditions **TOUBKAL** et l'Institut supérieur du Génie Appliqué, **IGA.**

*A la mémoire de Myriam*
*M D*

*A mes chers enfants Zineb et Adam*
*A.E*

# LIMINAIRE

On dit souvent que l'on peut faire dire ce qu'on veut aux statistiques ! C'est bien connu, entre le verre à moitié plein et le verre à moitié vide la différence d'interprétation nous interpelle, mais avant de pouvoir interpréter un ensemble de données, il est indispensable de savoir comment représenter, dans un tableau ou par un graphique, une série statistique, comment en faire les premiers traitements et surtout comment présenter les résultats de ces calculs.

Ce sont là, les objectifs de ce livre !

Le présent livre est un livre de cours.

La méthode adoptée peut se résumer dans les deux points suivants :

- Chaque chapitre est traité d'une façon exhaustive pour englober tous les concepts et toutes les démonstrations des formules statistiques.
  Il renferme, en plus, un ensemble d'exemples d'application avec solutions et surtout les méthodes de résolution.
- A la fin de chaque chapitre, le lecteur trouvera, ensuite un ensemble d'exercices d'application qui lui permettra de s'entraîner à résoudre des problèmes classiques de statistique.

Signalons, à cet effet, que pour toutes les solutions proposées pour les exemples, nous avons utilisé l'ordinateur avec des logiciels de graphisme et de gestionnaires de tableaux et nous encourageons vivement autant les étudiants que les professeurs d'en faire de même pour tout problème de statistique.

Cette utilisation de l'ordinateur nous amène à avertir nos lecteurs que les résultats des calculs donnés dans les tableaux et ailleurs différeront de ceux qu'on pourrait obtenir grâce à une calculette pour la simple raison que la puissance de précision d'un ordinateur ne peut jamais être égalée par une calculette.

Ce livre est ainsi destiné aux étudiants qui désirent acquérir une certaine adresse à la résolution de problèmes de statistique descriptive et aux professeurs qui recherchent un ensemble d'exercices didactiques de statistique descriptive à proposer à la réflexion de leurs étudiants.

**Les auteurs**

Casablanca, octobre 2006.

# SOMMAIRE


**INTRODUCTION** **9**

**PARTIE 1- STATISTIQUE DESCRIPTIVE A UNE SEULE VARIABLE** **13**

**CH. 1. TABLEAUX ET GRAPHIQUES** **15**
1.1. Tableaux statistiques 15
1.2. Représentations graphiques 28
1.3. Exercices d'application 37

**CH. 2. CARACTERISTIQUES DE TENDANCE CENTRALE** **43**
2.1. Les moyennes 43
2.2. Le mode 61
2.3. La médiane 63
2.4. La médiale 66
2.5. Les fractiles 69
2.6. Exercices d'application 73

**CH. 3. CARACTERISTIQUES DE DISPERSION** **78**
3.1. Ecart absolu moyen 78
3.2. Variance 82
3.3. Ecart type 86
3.4. Coefficient de variation 87
3.5. Indice de concentration 93
3.6. Exercices d'application 104

**PARTIE 2 - STATISTIQUE DESCRIPTIVE A DEUX VARIABLES** **111**

**CH. 4. REGRESSION ET CORRELATION** **113**
4.1. Introduction 113
4.2. Régression simple 113
4.3. Qualité de l'ajustement 130
4.4. Calcul des prévisions 137
4.5. Régression non linéaire simple 138
4.6. Régression multiple 142
4.7. Exercices d'application 153

**CH. 5. LES SERIES CHRONOLOGIQUES** **163**
5.1. Définition 163
5.2. Représentation graphique 164
5.3. Les principaux mouvements des séries chronologiques 166
5.4. Les schémas de composition 167
5.5. Les méthodes de lissage 169
5.6. Etude du trend 178
5.7. Etude de la composante saisonnière 183
5.8. Exercices d'application 195

**CH. 6. INDICES STATISTIQUES** **205**
6.1. Les indices élémentaires 205
6.2. Les indices synthétiques 211
6.3. Les indices synthétiques pondérés 216
6.4. Les principaux indices synthétiques 217
6.5. L'indice des prix à la consommation 220
6.6. Indices boursiers 233
6.7. Exercices d'application 234

**BIBLIOGRAPHIE** **246**

# INTRODUCTION

## HISTORIQUE.

L'activité qui consiste à recueillir des données permettant de connaître la situation des États remonte à la plus haute antiquité. On cite, d'une part, l'empereur chinois Yao, organisant le recensement des productions agricoles en 2238 avant J.-C., et, d'autre part, l'institution des recensements de la population chez les Égyptiens, en 1700 avant J.-C.

Au début du XVI$^{e}$ siècle, on commença à tenir en Angleterre un registre des décès et des naissances. En France, les intendants Sully, Colbert et Vauban commandèrent de nombreux inventaires et enquêtes. En 1662, l'Anglais John Graunt constata une certaine constance dans le rapport du nombre de naissances féminines à celui des naissances masculines.

On attribue la création du terme « statistique » à un professeur allemand Göttingen, G. Achenwall (1719-1772), qui aurait en 1746 créé le mot *Statistik*, dérivé de la notion *Staatskunde*.

Mais c'est seulement au XIX$^{e}$ siècle qu'on découvrit que la théorie des probabilités pouvait constituer une aide précieuse à la méthode statistique. Ce rapprochement, déjà perçu par le mathématicien Laplace, fut l'œuvre d'Adolphe Quételet (1796-1874), statisticien belge qui fut à l'initiative du premier congrès international de statistiques en 1853. Dès lors, la statistique se développa dans la plupart des sciences.

L'apparition d'une réelle méthodologie statistique a été initiée par des statisticiens anglais autour de 1900. C'est-à-dire une théorie bien formalisée du raisonnement qui permet, à partir des données observées, de tirer des conclusions sur les lois de probabilité des phénomènes. C'est la statistique mathématique, qui s'est développée entre 1900 et 1950 et dont les succès ont imposé, au cours de cette période, une interprétation particulière du concept de probabilité.

À partir des années cinquante, l'apparition de calculateurs puissants a donné naissance aux méthodes d'analyse des données multidimensionnelles, qui ont connu une grande vogue, parfaitement justifiée par leur efficacité. Ces méthodes permettent de décrire, de classer et de simplifier des données, les résultats auxquels elles conduisent peuvent suggérer des lois, des modèles ou des explications des phénomènes.

Aujourd'hui, les statistiques sont considérées comme des outils fiables qui peuvent fournir une représentation exacte des valeurs de données économiques, politiques, sociales, psychologiques, biologiques ou physiques. Elles permettent de mettre en corrélation de telles données et de les analyser. Le travail du statisticien ne se limite plus à recueillir des données et à les présenter sous forme de tableaux, mais il consiste principalement à interpréter l'information.

**DEFINITION.**

**Statistique**, une discipline qui a pour objet la collecte, le traitement et l'analyse de données numériques relatives à un ensemble d'individus ou d'éléments. Elle constitue un outil précieux pour l'expérimentation, la gestion des entreprises ou encore l'aide à la décision.

Une étude statistique se décompose en quatre étapes : la définition et la collecte des données, leur présentation en tableaux, leur analyse et enfin la comparaison des résultats avec des lois statistiques connues.

**1 - Définition et collecte des données**

La matière première des méthodes statistiques est constituée d'ensembles de nombres, obtenus en comptant ou en mesurant des éléments. Il est donc indispensable, lors de la collecte de données statistiques, de s'assurer de l'exhaustivité et de la fiabilité des informations recueillies.

Avant la collecte des données, on commence par définir la nature et la quantité des données à recueillir. Cette collecte s'effectue par recensement ou par sondage. Les données recueillies peuvent faire l'objet d'une vérification partielle par mesure de sécurité.

**2 - Représentation des données**

Les données recueillies sont classées et rangées dans des tableaux de façon à permettre une analyse et une interprétation directes. Ensuite, On peut représenter graphiquement les données du tableau.

### 3 - Analyse des données

Une fois les données recueillies et présentées sous forme de tableaux, le travail d'analyse commence par le calcul d'un paramètre statistique qui puisse résumer à lui seul l'ensemble des données. On distingue trois types de paramètres statistiques :

- Tendance centrale : elle sert à caractériser l'ordre de grandeur des observations ;
- Dispersion : elle sert à savoir si les mesures sont étroitement regroupées autour de la moyenne ou si elles sont dispersées ;
- Corrélation : elle sert à étudier la relation qui peut exister entre deux phénomènes.

### 4 - Comparaison des résultats avec des lois statistiques

Les statisticiens se sont aperçus que de nombreux ensembles de mesures avaient le même type de distribution. Ils ont donc été amenés à concevoir des modèles mathématiques qui soient le reflet des lois statistiques souvent rencontrées. La comparaison des résultats avec ces lois statistiques permet de donner une explication du phénomène observé et en vérifier le bien fondé.

Dans le présent ouvrage, nous nous proposons de montrer comment représenter les données recueillies et comment en faire le traitement en exposant successivement les méthodes de calcul des 3 paramètres statistiques que sont les paramètres de tendance centrale, ceux de dispersion et ceux de corrélation.

Les méthodes de collecte de donnée et celles de la comparaison des résultats de leurs traitements avec des lois statistiques feront l’objet d’un autre ouvrage.

# PARTIE 1
# STATISTIQUE DESCRIPTIVE A UNE VARIABLE

La statistique descriptive à une variable est l'ensemble des méthodes qui permet d'obtenir et de faire un 1$^{er}$ traitement des informations relatives à un caractère particulier d'individus d'une population donnée.

La statistique descriptive a plusieurs objectifs :

- recueillir l'ensemble des données relatives à un caractère particulier d'individus d'une population donnée ;

- classer l'ensemble de ces données selon des séries statistiques afin de permettre d'en faire :
  * des représentations graphiques pour en visualiser l'allure ;
  * des traitements mathématiques pour en déterminer certaines caractéristiques.

Dans cette partie, nous axerons notre propos, d'abord sur la définition des différents concepts que nous venons d'introduire, ensuite sur les premiers traitements mathématiques en vue de la détermination de certaines caractéristiques.

# CHAPITRE 1
# TABLEAUX ET GRAPHIQUES

## 1.1. TABLEAUX STATISTIQUES.

Nous donnons dans ce qui suit la définition des principaux concepts de la statistique.

**Population :** ensemble d'éléments ou d'individus ayant un caractère commun à étudier.

**Exemples 1** :

- Ensemble des étudiants d'une école ;
- Ensemble des habitants d'une ville ;
- Ensemble des livres d'une bibliothèque ;
- Ensemble de la production d'une entreprise pendant un an ;
- Etc.

**Echantillon :** partie de la population. Du fait de la taille importante de la population et de l'impossibilité d'en faire l'étude exhaustive, on se contente, le plus souvent, d'étudier le caractère d'après un échantillon.

L'échantillon doit être choisi de façon qu'il soit représentatif, pour ce faire il existe des méthodes de tri en vue de la constitution d'échantillon. Elles font l'objet d'études spécifiques.

**Exemples 2** :

- L'ensemble des étudiants d'une salle de classe d'une école ;
- L'ensemble d'un millier d'habitants choisi parmi tous les habitants d'une ville ;
- L'ensemble d'une centaine de livre trié parmi tous les livres d'une bibliothèque ;
- La production d'une entreprise pendant quelques jours ;
- Etc.

**Individu :** élément de base constituant la population ou l'échantillon, on dit aussi, unité statistique.

**Exemples 3** :

- L'étudiant d'une école ;
- L'habitant d'une ville ;
- Le livre d'une bibliothèque ;
- l'unité produite par une entreprise ;
- Etc.

**Effectif :** nombre d'individus observés constituant l'échantillon, il est noté n.

**Exemples 4** :

- n = 30 s'il y a 30 étudiants dans l'échantillon ;
- n = 2000 s'il y a 2000 habitants dans l'échantillon ;
- n = 125 s'il y a 125 livres dans l'échantillon ;
- n = 15 000 s'il y a 15 000 unités produites constituant l'échantillon ;
- etc.

**Caractère :** Aspect particulier commun à tous les individus de la population et donc de l'échantillon. Le caractère peut être qualitatif ou quantitatif et dans ce cas il peut être discret ou continu.

**Exemples 5** :

- Notes des étudiants d'une école ;
- Situations familiales des habitants d'une ville ;
- Thèmes des livres d'une bibliothèque ;
- Poids des unités produites par une entreprise ;
- Etc.

**Modalités :** Ce sont les différentes possibilités que peut prendre le caractère, par exemple féminin ou masculin si le caractère est le sexe et 1,50 m ou 1,70 m si le caractère est la taille, etc.

**Caractère qualitatif :** Un caractère est dit qualitatif quand il ne peut pas être mesuré.

**Exemples 6** : Le tableau ci-dessous liste quelques exemples de caractères qualitatifs et de modalités :

| Caractères | Modalités | Genres |
|---|---|---|
| Couleur | Rose, rouge, blanc, bleu, … | Qualitatif |
| Nationalité | Marocain, Français, Suisse, … | Qualitatif |
| Situation matrimoniale | Marié, célibataire, veuf, divorcé, … | Qualitatif |
| Disponibilité | Oui, non. | Qualitatif |

Pour chaque modalité i du caractère à étudier, on détermine, pour l'échantillon considéré, de taille n :

$n_i$ : effectif d'individus chez qui a été observée la modalité i.

$f_i = n_i / n$ : fréquence relative de la modalité i.

Avec $\sum_{i=1}^{k} n_i = n$ et $\sum_{i=1}^{k} f_i = 1$

**Exemples 7** : Dans un échantillon de 2000 habitants d'une ville, en relève que 900 personnes sont mariées, on a ainsi, pour la modalité « habitants mariés » :

$n_i$ = 900 et $f_i$ = 900/2000 = 45 % ;

- Dans une bibliothèque constituée de 5000 livres on relève que 120 livres ont pour thème les mathématiques, on a ainsi pour la modalité « livres de mathématiques » :

$n_i$ = 120 et $f_i$ = 120/5000 = 2,4 %

- On considère l'ensemble des touristes qui visitent le Maroc pendant une période donnée et on considère comme caractère la nationalité. Si l'on relève qu'il y a 300 Français parmi un ensemble de 900 touristes on a pour la modalité « nationalité française » :

$n_i$ = 300 et $f_i$ = 300/900 = 33,33 %

**Caractère quantitatif :** Un caractère est dit quantitatif quand il peut être mesuré. Il peut alors être continu ou discret :
- il est discret dans le cas d'opérations de dénombrement ou de comptage ;
- il est continu dans le cas d'opérations de mesures.

**Exemples 8** : Le tableau ci-dessous donne quelques exemples de caractères quantitatifs et de modalités.

| Caractères | Modalités | Genres |
|---|---|---|
| Poids | 60,5 Kg; 59,2 Kg; 65,3 Kg; … | Continu |
| Ancienneté en entreprise | 10 ans et 2 mois ; 9ans ; … | Continu |
| Volume | 1 $m^3$ ; 2,3 $m^3$ ; 3 $m^3$ ; … | Continu |
| Longueur | 1 m ; 2,75 km ; 350 dm ; … | Continu |
| Notation | 10/20 ; 9,5/10 ; … | Continu |
| Années d'études | 2 ans ; 3 ans ; 6 ans ; … | Discret |
| Nombre de frères et sœurs | 1 ; 2 ; 3 ; … | Discret |
| Nombre d'enfants | 0 ; 1 ; 2 ; … | Discret |

On détermine pour chaque caractère quantitatif :

**Si le caractère est discret :** $x_i$ la valeur de la modalité ;

$n_i$ : effectif d'individus chez qui, la modalité i, a été observée.

$f_i = n_i / n$ : fréquence relative de la modalité i.

$F_i : \sum_{j=1}^{j=i} f_j$ fréquence relative cumulée croissante.

F(x) : Fonction de répartition, proportion d'individus ayant des modalités du caractère étudié inférieures ou égales à x.

**Exemple 9** : On considère le poids des habitants d'une ville comme caractère, on a, pour un échantillon, la distribution suivante :

| **Poids xi** | **Effectifs concernés ni** | **Fréquences relatives fi** | **Fréquences relatives cumulées Fi** |
|---|---|---|---|
| 65Kg | 54 | 21.86% | 21.86% |
| 70Kg | 132 | 53.44% | 75.30% |
| 75Kg | 27 | 10.93% | 86.23% |
| 80Kg | 34 | 13.77% | 100.00% |
| **Total** | **247** | **100%** | **-** |

Unité statistique : habitant de la ville ;
Population : l'ensemble des habitants de la ville ;
Caractère étudié : le poids ;
Type de caractère : variable statistique discrète. (dans le cas de l'exemple).

La fonction de répartition F(x) se définit comme suit :

Pour $x \leq 65$ Kg, on a : F(x) = 21,86% ou 21,86 % de l'échantillon ont un poids inférieur ou égal à 65 Kg.

Pour $x \leq 70$ Kg, on a : F(x) = 75.30% ou 75,30 % de l'échantillon ont un poids inférieur ou égal à 70 Kg.

Pour $x \leq 75$ Kg, on a : F(x) = 86.23% ou 86,23 % de l'échantillon pèsent au plus 75 Kg.

Pour $x \leq 80$ Kg, on a : F(x) = 100.00 % ou la totalité de l'échantillon a un poids inférieur ou égal à 80 Kg.

**Exemple 10** : une enquête auprès de 1000 commerçants portant sur le nombre de leurs employés, a donné les résultats suivants :

| xi | $n_i$ | $f_i$ | Fréquence absolue cumulée croissante | Fréquence absolue cumulée décroissante | Fréquence relative cumulée croissante | Fréquence relative cumulée décroissante |
|---|---|---|---|---|---|---|
| 0 | 50 | 5 % | 50 | 1000 | 5 % | 100 % |
| 1 | 100 | 10 % | 150 | 950 | 15 % | 95 % |
| 2 | 200 | 20 % | 350 | 850 | 35 % | 85 % |
| 3 | 150 | 15 % | 500 | 650 | 50 % | 65 % |
| 4 | 120 | 12 % | 620 | 500 | 62 % | 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 %** | **-** | **-** | **-** | **-** |

Unité statistique : Un commerçant ;
Population : l'ensemble des 1000 commerçants ;
Caractère étudié : Nombre d'employés ;
Type de caractère : Variable statistique discrète.

Le nombre de commerçants n'employant aucun employé est 50, ce qui représente 5 % des commerçants.

Les fréquences absolues ou relatives cumulées croissantes sont calculées en cumulant les fréquences absolues ou relatives du haut du tableau vers le bas. Elles permettent de répondre aux questions du genre : quel est le nombre ou la proportion au plus ?

Par contre, les fréquences absolues ou relatives cumulées décroissantes sont calculées en cumulant les fréquences absolues ou relatives du bas du tableau vers le haut. Elles permettent de répondre aux questions du genre : quel est le nombre ou la proportion au moins (au minimum ou plus de) ?

Le nombre de commerçants employant au plus 5 employés (au maximum 5 employés ou moins de 6 employés) est 780, ils représentent 78 % des commerçants.

Le nombre de commerçants employant au moins 3 employés (au minimum 3 employés ou plus de 2 employés) est 650, ils représentent 65% des commerçants.

**Si le caractère est continu :** $[C_i ; C_{i+1}[$ est l'intervalle ou classe des modalités avec :

- $C_i$ et $C_{i+1}$ les bornes de la classe ;
- $c_i$ : centre de la classe ;
- $a_i$ : amplitude de la classe ;
- $d_i$ : densité de la classe.
- $n_i$ : effectif de la classe i, nombre d'individus dont la modalité du caractère est comprise entre $C_i$ et $C_{i+1}$.

$c_i = \dfrac{C_{i+1} + C_i}{2}$, $\quad a_i = C_{i+1} - C_i$ et $\quad d_i = n_i/a_i$

De la même manière que dans le cas discret, on définit :

$f_i = n_i / n$ : fréquence relative de la modalité i.

$F_i : \sum_{j=1}^{j=i} f_j$ fréquence relative cumulée croissante.

**Exemple 11** : On considère la taille comme caractère, on a pour un échantillon de 169 personnes, la distribution suivante :

| **Tailles (en m)** $[C_i ; C_{i+1}[$ | $c_i$ **(en m)** | **Effectifs concernés** $n_i$ | **Fréquences relatives** $f_i$ | **Fréquences relatives cumulées** $F_i$ |
|---|---|---|---|---|
| [1,50 ; 1,60[ | 1.55 | 35 | 20,71 % | 20,71 % |
| [1,60 ; 1,70[ | 1.65 | 42 | 24,85 % | 45,56 % |
| [1,70 ; 1,80[ | 1.75 | 53 | 31,36 % | 76,92 % |
| [1,80 ; 1,90[ | 1.85 | 39 | 23,08 % | 100 % |
| **Total** | **-** | **169** | **100 %** | **-** |

Parmi les 169 personnes, 35 mesurent entre 1,50 m et moins de 1,60 m, ce qui représente 20,71 % de l'ensemble de l'échantillon.

76,92 % de l'échantillon mesurent moins de 1,80 m.

Le fait de remplacer la classe $[C_i ; C_{i+1}[$ par $c_i$ permet de faire des calculs car on ne sait pas faire des calculs sur des intervalles.

**Série statistique :** Une série statistique est l'ensemble constitué des $x_i$ et $n_i$. On parle aussi de distribution statistique à une seule variable, comme par exemple :

- Tailles et effectifs ;
- Situations matrimoniales et effectifs ;
- Ages et effectifs.
- Etc.

**Question 1 :** Comment passer d'une série statistique relative à un caractère discret ou continu donnée sous forme d'une suite de classes $[C_i ; C_{i+1}[$ et d'effectifs $n_i$ de ces classes à une série statistique sous forme d'une suite de valeurs $x_i$ et d'effectifs $n_i$ relatifs à ces valeurs ?

On doit considérer 2 cas possibles :

**1è cas** : Classes à amplitudes égales.

Il suffit, dans ce cas, de remplacer chaque classe $[C_i ; C_{i+1}[$ par son élément central $c_i = ( C_i + C_{i+1} ) / 2$ auquel il faut affecter l'effectif $n_i$.

**Exemple 12** : On considère la série statistique relative aux poids d'un échantillon de 120 habitants d'une ville, elle se présente comme l'indique le tableau suivant :

| **Poids (kg)** $[C_i ; C_{i+1}[$ | **Effectifs** $n_i$ |
|---|---|
| [55 ; 60[ | 5 |
| [60 ; 65[ | 14 |
| [65 ; 70[ | 20 |
| [70 ; 75[ | 40 |
| [75 ; 80[ | 18 |
| [80 ; 85[ | 15 |
| [85 ; 90[ | 8 |
| **Total** | **120** |

Unité statistique : Habitant d'une ville ;
Population : L'ensemble des habitants d'une ville ;
Caractère étudié : Le poids de l'habitant ;
Type de caractère : Variable statistique continue.

On remplace chaque classe par le centre de cette classe, on obtient alors la série équivalente suivante :

| **Poids (kg)** $c_i$ | **Effectifs** $n_i$ | **Fréquence relative** $f_i$ |
|---|---|---|
| 57,5 | 5 | 4.17% |
| 62,5 | 14 | 11.67% |
| 67,5 | 20 | 16.67% |
| 72,5 | 40 | 33.33% |
| 77,5 | 18 | 15.00% |
| 82,5 | 15 | 12.50% |
| 87,5 | 8 | 6.67% |
| **Total** | **120** | **100%** |

**Exemple 13** : On considère la série statistique relative aux notes obtenues dans une matière, par les étudiants d'une classe d'école :

| **Notes** $[C_i ; C_{i+1}[$ | **Effectifs** **ni** |
|---|---|
| [6 ; 8[ | 2 |
| [8 ; 10[ | 6 |
| [10 ; 12[ | 12 |
| [12 ; 14[ | 7 |
| [14 ; 16[ | 3 |
| **Total** | **30** |

Unité statistique : Un étudiant ;
Population : L'ensemble des étudiants d'une classe d'école
Caractère : Note d'étudiant
Type de caractère : Variable statistique continue

On remplace chaque classe par le centre de cette classe, on obtient alors la série équivalente suivante :

| **Notes** **ci** | **Effectifs** **xi** | **Fréquences relatives** **fi** |
|---|---|---|
| 7 | 2 | 6.67% |
| 9 | 6 | 20% |
| 11 | 12 | 40% |
| 13 | 7 | 23.33% |
| 15 | 3 | 10% |
| **Total** | **30** | **100%** |

**2è cas** : Classes à amplitudes différentes.

Il suffit, dans ce cas :
- de considérer les amplitudes des différentes classes ;
- de calculer leur Plus Grand Commun Diviseur (PGCD) ;
- de diviser chaque classe par le PGCD pour obtenir plusieurs sous classes qui deviennent de nouvelles classes ;
- D'affecter à chaque nouvelle classe, le quotient de l'effectif de la classe mère par le nombre de sous classes.

Remarquons que cette méthode repose sur l'hypothèse simple suivante qui consiste à admettre que les effectifs se répartissent de façon régulière dans une classe.

**Exemple 14** : Reprenons l'exemple 13 et considérons la série statistique relative aux notes obtenues dans une autre matière, par les étudiants d'une classe d'école :

| **Notes** $[C_i ; C_{i+1}[$ | **Effectifs** $n_i$ |
|---|---|
| [0 ; 6[ | 6 |
| [6 ; 8[ | 4 |
| [8 ; 14[ | 12 |
| [14 ; 18[ | 4 |

Unité statistique : Un étudiant ;
Population : L'ensemble des étudiants d'une classe d'école
Caractère : Note d'étudiant
Type de caractère : Variable statistique continue

Dans cette série, les amplitudes des différentes classes sont : 6 ; 2 ; 6 ; 4. Leur PGCD est 2. On remplace chaque classe par plusieurs autres classes et on obtient alors la série équivalente suivante :

| **Notes** $[C_i ; C_{i+1}[$ | **Effectifs** $n_i$ |
|---|---|
| [0 ; 2[ | 2 |
| [2 ; 4[ | 2 |
| [4 ; 6[ | 2 |
| [6 ; 8[ | 4 |
| [8 ; 10[ | 4 |
| [10 ; 12[ | 4 |
| [12 ; 14[ | 4 |
| [14 ; 16[ | 2 |
| [16 ; 18[ | 2 |

On remplace, après cette opération, chaque classe par le centre de cette classe, on obtient alors la série équivalente suivante :

| **Notes** $x C_i ; C_{i+1} x$ | $c_i$ | **Effectifs** **ni** |
|---|---|---|
| [0 ; 2[ | 1 | 2 |
| [2 ; 4[ | 3 | 2 |
| [4 ; 6[ | 5 | 2 |
| [6 ; 8[ | 7 | 4 |
| [8 ; 10[ | 9 | 4 |
| [10 ; 12[ | 11 | 4 |
| [12 ; 14[ | 13 | 4 |
| [14 ; 16[ | 15 | 2 |
| [16 ; 18[ | 17 | 2 |

**Remarque** : Ainsi on peut considérer que toute série statistique est donnée, selon les besoins du traitement numérique :

- Soit sous forme d'une suite de classes $[C_i ; C_{i+1}[$ et d'effectifs $n_i$.
- Soit sous forme d'une suite de valeurs $x_i$ et d'effectifs $n_i$

**Question 2 :** Comment passer d'une série statistique relative à un caractère discret ou continu donnée sous forme d'une suite de valeurs $x_i$ à une série donnée sous forme d'une suite de classes $[C_i , C_{i+1}[$ et d'effectifs $n_i$ par classe ?

Pour ce faire, on utilise la règle de STURGES donnant le nombre k de classes en fonction du nombre n des données :

$$k = 1 + 3{,}322 \log_{10} n$$

Ce calcul donne un nombre réel, on prend alors pour k le nombre entier très proche du résultat de calcul de la formule précédente.

Et étant l'étendue E de toute la série statistique, on détermine e, étendue de chaque classe :

$$e = E / k \quad \text{avec} \quad E = x_{max} - x_{min}$$

$x_{max}$ et $x_{min}$ étant la valeur maximale et la valeur minimale prises par le caractère, les différentes classes seront alors :

La borne inférieure de la première classe $C_1$ est égale à $x_{min}$ ou à une valeur légèrement inférieure à $x_{min}$.

$[C_1 ; C_1+e[$
$[ C_1+e ; C_1+2e[$
$[ C_1+2e ; C_1+3e[$
…
$[ C_1+(k-1)e ; C_1+ke[$

**Exemple 15** : En prenant la taille comme caractère des habitants d'une ville on a les résultats relatifs à un échantillon de 169 habitants :

| **Tailles (en m)** $x_i$ | **Effectifs concernés** $n_i$ |
|---|---|
| 1.45 | 5 |
| 1.55 | 30 |
| 1.65 | 42 |
| 1.75 | 53 |
| 1.85 | 39 |
| **Total** | **169** |

Unité statistique : Habitants d'une ville;
Population : L'ensemble des habitants de la ville
Caractère : La taille de l'habitant
Type de caractère : Variable statistique continue

On applique la méthode de STURGES avec les conditions :

- N = 169
- E = 1,85 – 1,45 = 0.40

Ce qui donne, après calcul, $k=1+3.322 \log_{10} 169 = 8.40$

On prendra k = 8 et e = E/8 = 0.40/8 = 0.05

La série précédente peut être transformée en la série équivalente suivante

| **Tailles (en m)** $[C_i ; C_{i+1}[$ | **Effectifs concernés** **ni** |
|---|---|
| [1.45 ; 1.50[ | 5 |
| [1.50 ; 1.55[ | 0 |
| [1.55 ; 1.60[ | 30 |
| [1.60 ; 1.65[ | 0 |
| [1.65 ; 1.70[ | 42 |
| [1.70 ; 1.75[ | 0 |
| [1.75 ; 1.80[ | 53 |
| [1.80 ; 1.85[ | 0 |
| [1.85 ; 1.90[ | 39 |
| **Total** | **137** |

**Remarque** : on aboutit à 9 classes au lieu de 8 du fait de la configuration des intervalles définissant les classes.

**Exemple 16 :** On a mesuré le poids en kilogramme comme caractère pour un échantillon de 80 élèves d'une école. Les données brutes sont reportées dans le tableau suivant :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 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 |

Unité statistique : Elève d'une école ;
Population : L'ensemble des élèves d'une école ;
Caractère : Le poids ;
Type de caractère : Variable statistique discrète.

La plus grande valeur est : 97
La plus petite valeur est : 53
L'étendue est : E = 97 - 53 = 44

On applique la méthode de STURGES avec les conditions :

- n = 80
- E = 44

Ce qui donne, après calcul, $k = 1 + 3{,}322 \log_{10} 80 = 7{,}322$
On prendra $k = 7$ et $e = E / 7 = 44 / 7 = 6$

La série précédente peut être transformée en la série équivalente suivante :

| **Poids** | **ci (1)** | **ni (2)** | **$N_i$ cr (3)** | **$N_i$ dé (4)** | **$f_i$ (5)** | **$F_i$ cr (6)** | **$F_i$ dé (7)** |
|---|---|---|---|---|---|---|---|
| [52 ; 58[ | 55 | 2 | 2 | 80 | 2,5% | 2,5% | 100% |
| [58 ; 64[ | 61 | 12 | 14 | 78 | 15% | 17,5% | 97,5% |
| [64 ; 70[ | 67 | 10 | 24 | 66 | 12,5% | 30% | 82,5% |
| [70 ; 76[ | 73 | 19 | 43 | 56 | 23,75% | 53,75 | 70% |
| [76 ; 82[ | 79 | 16 | 59 | 37 | 20% | 73,75 | 46,25% |
| [82 ; 88[ | 85 | 9 | 68 | 21 | 11,25% | 85% | 26,25% |
| [88 ; 94[ | 91 | 7 | 75 | 12 | 8,75% | 93,75% | 15% |
| [94 ; 100[ | 97 | 5 | 80 | 5 | 6,25% | 100% | 6,25% |
| **Total** | **- - -** | **80** | **- - -** | **- - -** | **100%** | **- - -** | **- - -** |

**Légende du tableau :**

- (1) : point central de la classe ;
- (2) : effectif de la classe, fréquence absolue ;
- (3) : fréquence absolue cumulée croissante ;
- (4) : fréquence absolue cumulée décroissante ;
- (5) : pourcentage de la classe, fréquence relative ;
- (6) : fréquence relative cumulée croissante ;
- (7) : fréquence relative cumulée décroissante.

Le nombre de personnes pesant entre 64 et moins de 70 kilogrammes est 10, ils représentent 12,5 % des personnes pesées.

Le nombre de personnes pesant au moins 70 kilogrammes est 56, ils représentent 70 % des personnes pesées.

Le nombre de personnes pesant moins de 82 kilogrammes est 59, ils représentent 73,75 % des personnes pesées.

Pour récapituler toute cette première partie, donnons, dans un tableau synthétique, grâce à des exemples, l'ensemble des concepts que nous avons introduits jusque là :

| Population | Echantillon | Caractères | Modalités | Effectifs |
|---|---|---|---|---|
| Habitants d'une ville | 200 habitants choisis | -taille<br>-poids<br>-etc. | - 1m65<br>- 65kg<br>- etc. | 200 |
| Elèves d'une école | 30 élèves triés | -notes | - 13,5 | 30 |
| Livres d'une bibliothèque | 125 livres triés | -thèmes des livres | - math | 125 |
| Production d'une usine | 1500 unités produites triées | -poids de l'unité<br>-dimension de l'unité | - 8g<br>- 37cm | 1500 |

Le tri ou le choix pour constituer un échantillon se fait selon des processus bien précis.

## 1.2. REPRESENTATIONS GRAPHIQUES.

Il est très courant, dans un premier traitement, pour bien visualiser l'allure d'une série statistique, de la représenter par un graphe. Cette représentation peut être faite selon plusieurs manières, en effet on peut citer les différentes représentations suivantes :

- le diagramme à bandes ;
- le diagramme à secteurs ;
- le diagramme à bâtons ;
- l'histogramme des fréquences simples ;
- le polygramme des fréquences simples ;
- la courbe des fréquences cumulées.

Chaque type de représentation convient à un type de caractère (qualitatif ou quantitatif, quantitatif discret ou quantitatif continu) et à un type de série.

Nous donnons dans ce qui suit un ensemble de possibilités de représentations d'une série statistique en indiquant, chaque fois, le choix du graphe adéquat selon le type de caractère ou de la série ainsi que les raisons de ce choix.

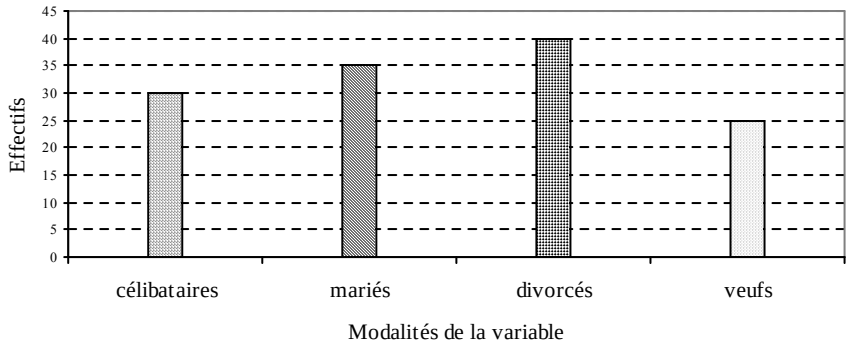
**1.2.1. Caractère qualitatif :** Rappelons qu'un caractère qualitatif est un caractère qu'on ne peut pas mesurer. Dans ce cas, deux types de représentations sont conseillés :

**Diagramme à bandes :**

**Exemple 17 :** On considère la série statistique relative à la situation familiale d'un échantillon de 130 personnes :

| **Situations familiales** | $x_i$ | **Effectifs concernés : $n_i$** |
|---|---|---|
| Célibataires | 1 | 30 |
| Mariés | 2 | 35 |
| Divorcés | 3 | 40 |
| Veufs | 4 | 25 |
| **Total** | **---** | **130** |

La représentation graphique d'une telle série peut être très bien faite par un diagramme à bandes.




**Remarques** : La largeur des bandes est quelconque mais identique pour toutes les bandes.

Seules les hauteurs des bandes indiquent les effectifs ou les fréquences relatives.

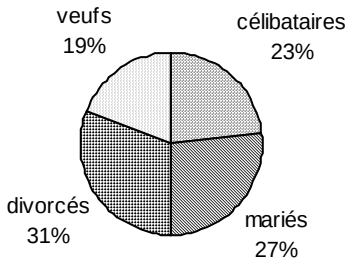
La numérotation des classes de modalités de 1 à 4 est faite uniquement dans le but de faciliter les représentations graphiques.

**Diagramme à secteurs :**

**Exemple 18 :** On reprend l'exemple 7 et l'on considère la même série statistique relative à la situation familiale d'un échantillon de 130 personnes pour laquelle nous avons converti les effectifs en pourcentage :

| **Situations familiales** | $x_i$ | **Effectifs concernés : $n_i$**<br>**Fréquences relatives : $f_i$** |
|---|---|---|
| Célibataires | 1 | 30 = 23% |
| Mariés | 2 | 35 = 27% |
| Divorcés | 3 | 40 = 31% |
| Veufs | 4 | 25 = 19% |
| **Total** | **- - -** | **130 = 100%** |

La représentation graphique d'une telle série peut être très bien faite par un diagramme à secteurs.




**Remarque 1** : le même caractère, situation familiale a pu être représenté par deux types de diagrammes.

**Remarque 2** : La surface de chaque secteur représente, en pourcentage, la fréquence relative de la modalité indiquée.

Le rayon du cercle est quelconque.

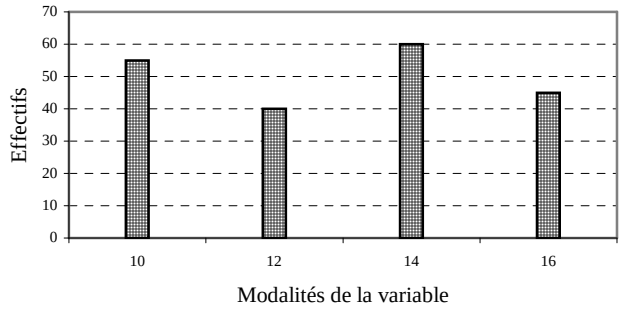
**1.2.2. Caractère quantitatif discret :** Rappelons qu'un caractère quantitatif est discret dans le cas d'opérations de comptage, dans ce cas, plusieurs types de représentation sont possibles.

**Diagramme à bâtons :**

**Exemple 19 :** On considère la série statistique des notes obtenues dans une matière, par un échantillon de 200 étudiants d'un amphithéâtre de 500.

| **Notes : $x_i$** | **Effectifs : $n_i$** |
|---|---|
| 10 | 55 |
| 12 | 40 |
| 14 | 60 |
| 16 | 45 |
| **Total** | **200** |

Pour représenter une telle série, on a habituellement recours aux diagrammes à bâtons.



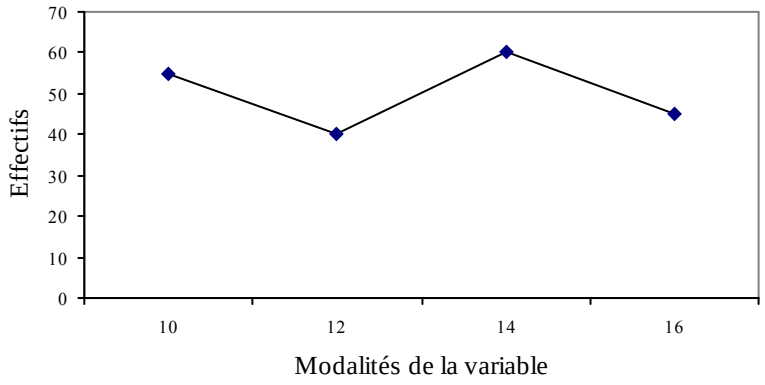

**Remarque** : La hauteur de chaque bâton est proportionnelle à $n_i$ ou $f_i$ pour la valeur $x_i$ du caractère.

Sur l'axe des x, on reporte les valeurs de $x_i$ attribuées aux caractères afin de pouvoir traiter la série statistique.

La largeur du bâton n'a aucune importance.

**Polygone de fréquences :**

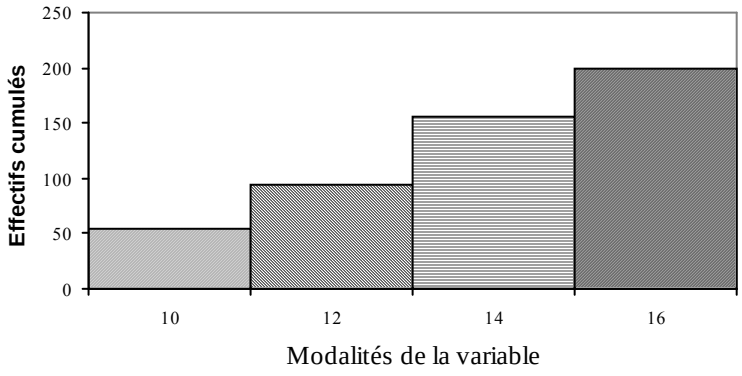
Les polygones de fréquences sont construits en joignant par une ligne les sommets des bâtons du diagramme en bâtons.




**Polygone de fréquences cumulées ou diagramme en escalier :**

**Exemple 20 :** On reprend l'exemple 18 et l'on considère les notes obtenues dans une matière, par un échantillon de 200 élèves, calculons les effectifs cumulés.

| **Notes : $x_i$** | **Effectifs : $n_i$** | **Effectifs cumulés $F_i$** |
|---|---|---|
| 10 | 55 | 55 |
| 12 | 40 | 95 |
| 14 | 60 | 155 |
| 16 | 45 | 200 |
| **Total** | **200** | **- - -** |



**1.2.3. Caractère quantitatif continu :** Rappelons qu'un caractère quantitatif est continu dans le cas d'opérations de mesures, dans ce cas, plusieurs types de représentation sont possibles.

**Histogramme :**

Un histogramme est un graphique constitué de bandes verticales jointives. On délimite en abscisses les classes successives de la variable continue, en principe de même amplitude, et sur chaque base ainsi délimitée, on élève un rectangle de hauteur proportionnelle à la fréquence correspondante de telle sorte que la surface du rectangle soit proportionnelle à l'effectif correspondant.

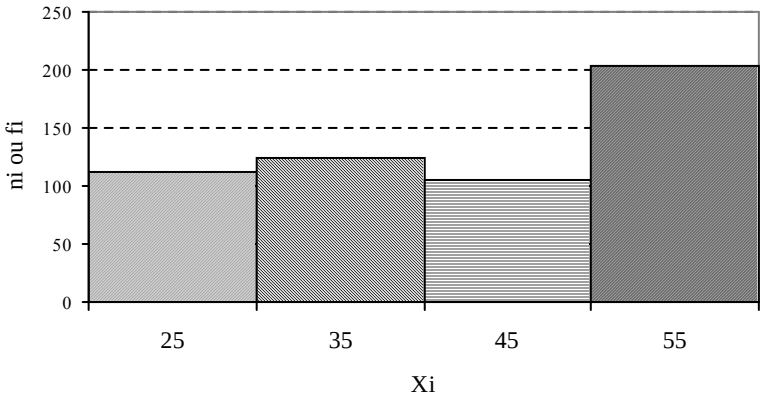
Quand les classes sont de même amplitude, la hauteur des rectangles est proportionnelle aux fréquences des classes, elle est égale numériquement à la fréquence correspondante. Si les classes n'ont pas la même amplitude, il est nécessaire d'ajuster la hauteur des rectangles de telle sorte que la surface du rectangle soit proportionnelle à l'effectif correspondant, la hauteur des rectangles est égale dans ce cas à la densité de la classe.

**Histogramme des fréquences à classes d'amplitudes égales :**

**Exemple 21 :** On considère un échantillon de 530 personnes et l'on prend pour caractère la somme en DH qu'elles ont dans leur poche.

| **Montant d'argent DH** | **Effectifs $n_i$** |
|---|---|
| [20 ; 30[ | 110 |
| [30 ; 40[ | 120 |
| [40 ; 50[ | 100 |
| [50 ; 60[ | 200 |
| **Total** | **530** |

Pour représenter une telle série statistique on a habituellement recours à l'histogramme des fréquences à classes d'amplitudes égales.




**Remarque :** On peut regrouper les valeurs discrètes par classes de même amplitude, il suffit alors que la hauteur de chaque rectangle soit proportionnelle à $n_i$ ou $f_i$ .

Sur l'axe des x, on reporte les valeurs $C_i$, bornes des classes du caractère x.

**Histogramme des fréquences à classes d'amplitudes inégales :**

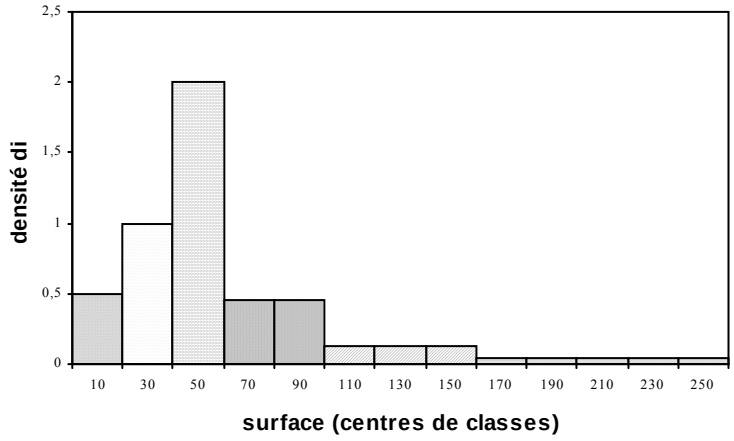
On peut regrouper les valeurs discrètes par classes d'amplitudes différentes, il suffit alors que la hauteur de chaque rectangle soit proportionnelle à $d_i$, densité de la classe considérée.

Sur l'axe des x, on reporte les valeurs $C_i$, bornes des classes du caractère x.

**Exemple 22 :** La répartition de la surface, en m², de 100 logements est représentée dans le tableau suivant :

| **Surface en m²** | **Nombre de logements** | **Densités** |
|---|---|---|
| 0 à 20 | 10 | 0,5 |
| 20 à 40 | 20 | 1 |
| 40 à 60 | 40 | 2 |
| 60 à 100 | 18 | 0,45 |
| 100 à 160 | 8 | 0,13 |
| 160 à 260 | 4 | 0,04 |
| **Total** | **100** | |

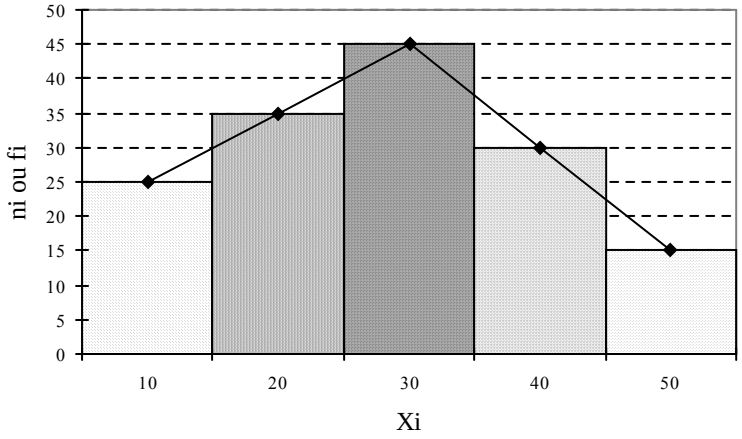
Les amplitudes des classes étant inégales, il convient de calculer les densités afin de représenter l'histogramme.




**Polygone des fréquences :** $d_i$ ou $f_i$

**Exemple 23 :** La répartition des soldes d'un échantillon de 150 comptes bancaires est donnée par le tableau suivant :

| **$[C_i ; C_{i+1}[$ en 1000 DH** | **$c_i$ en 1000 DH** | **Effectif $n_i$** |
|---|---|---|
| [5 ; 15[ | 10 | 25 |
| [15 ; 25[ | 20 | 35 |
| [25 ; 35[ | 30 | 45 |
| [35 ; 45[ | 40 | 30 |
| [45 ; 55[ | 50 | 15 |
| **Total** | **- - -** | **150** |



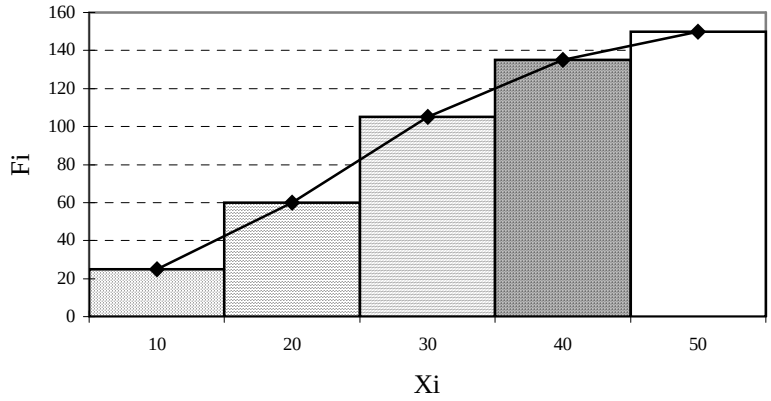

On construit, à partir de l'histogramme des fréquences.

- Le polygone des fréquences, en joignant les milieux des segments.
- La surface du polygone des fréquences est la même que celle de l'histogramme.

**Exemple 24 :** On reprend l'exemple 23 et on se propose de représenter la courbe des fréquences cumulées croissantes.

**Courbe des fréquences cumulées croissantes**

| **$[C_i ; C_{i+1}[$ en 1000 DH** | **$C_i$ en 1000 DH** | **Effectif $n_i$** | **$F_i$** |
|---|---|---|---|
| [5 ; 15[ | 10 | 25 | 25 |
| [15 ; 25[ | 20 | 35 | 60 |
| [25 ; 35[ | 30 | 45 | 105 |
| [35 ; 45[ | 40 | 30 | 135 |
| [45 ; 55[ | 50 | 15 | 150 |
| **Total** | **- - -** | **150** | **- - -** |




Les individus sont classés en classes, la fréquence cumulée associée à la classe numéro i correspond à la proportion d'individus dont la valeur du caractère est strictement inférieure à la limite supérieure de la classe numéro i.

### 1.3. EXERCICES D'APPLICATIONS.

#### 1.3.1. Exercice.

A partir des tableaux suivants préciser :

a) l'unité statistique et la population ;
b) le caractère étudié ;
c) la nature du caractère étudié ;
d) représenter graphiquement la distribution ;

Structure de l'emploi au Maroc :

| **Secteurs d'activités** | **Part en %** |
|---|---|
| Agricole, forêt, pêche et mine | 4,9 |
| Industrie, bâtiment | 34,5 |
| Commerce | 19 |
| Hôtels et restaurants | 2,7 |
| Transport et communications | 7,9 |
| Finances et banques | 6,6 |
| Emploi domestique | 20,3 |
| Secteur public | 4,1 |
| **Total** | **100** |

Effectif des stagiaires en formation :

| Niveau | 1ère année | 2ème année | Total |
|---|---|---|---|
| Technicien spécialisé | 1031 | - | 1031 |
| Technicien | 9727 | 8487 | 18214 |
| Qualification | 12542 | 9293 | 21835 |
| Spécialisation | 6573 | 1335 | 7908 |
| **Total** | **29873** | **19115** | **48988** |

Répartition du nombre de pièces d'un ensemble de logements :

| Nombre de pièces | Part en % |
|---|---|
| 1 pièce | 24,68 |
| 2 pièces | 21,45 |
| 3 pièces | 20,50 |
| 4 pièces | 16,54 |
| 5 pièces et plus | 16,83 |
| **Total** | **100** |

Durée de vie des tubes électroniques :

| Durée (heures) | Nombre de tubes |
|---|---|
| 400-499 | 90 |
| 500-599 | 88 |
| 600-699 | 120 |
| 700-799 | 105 |
| 800-899 | 102 |
| 900-999 | 75 |
| 1000-1099 | 20 |
| **Total** | **600** |

### 1.3.2. Exercice.

Une étude de marché a mesuré le degré de satisfaction d'un échantillon de 500 clients d'une banque. Les résultas sont présentés dans le tableau suivant :

| Degré de satisfaction | Effectifs |
|---|---|
| Pas du tout satisfait | 223 |
| Insatisfait | 187 |
| Indifférent | 32 |
| Satisfait | 55 |
| Très satisfait | 3 |
| **Total** | **500** |

a) Quelle est la population étudiée ?

b) Quel est le caractère étudié ? quelle est sa nature ?
c) Calculer les fréquences relatives ?
d) Représenter graphiquement cette distribution.

### 1.3.3. Exercice.

Soit la distribution suivante du nombre de pièces dans 300 logements :

| **Nombre de pièces** | **Effectifs** |
|---|---|
| 1 | 35 |
| 2 | 51 |
| 3 | 68 |
| 4 | 55 |
| 5 | 49 |
| 6 | 42 |
| **Total** | **300** |

a) Présenter dans un tableau les différentes fréquences cumulées.
b) Quel est le nombre de logements possédant au moins 3 pièces ?
c) Quelle est la proportion des logements possédant moins de 5 pièces?
d) Quel est le nombre de logements possédant au plus 4 pièces ?
e) Quelle est la proportion des logements possédant plus de 3 pièces ?
f) Représenter graphiquement :
- la distribution des fréquences ;
- la distribution des fréquences cumulées croissantes ;

### 1.3.4. Exercice.

On a relevé la recette hebdomadaire en milliers de dirhams de 40 commerces. Les données brutes sont :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 57 | 60 | 52 | 49 | 56 | 46 | 51 | 63 | 49 | 57 |
| 86 | 93 | 77 | 67 | 81 | 70 | 71 | 91 | 67 | 82 |
| 47 | 87 | 92 | 55 | 48 | 90 | 49 | 50 | 58 | 62 |
| 67 | 89 | 69 | 72 | 75 | 48 | 85 | 90 | 83 | 66 |

a) Présenter les données dans un tableau statistique sous forme de classes.
b) Représenter graphiquement la distribution de fréquences établie.

### 1.3.5. Exercice.

Le tableau suivant présente le nombre de femmes en activité selon l'âge de 500 femmes actives

| Tranche d'âges | Effectif |
|---|---|
| [15 à 20[ | 14 |
| [20 à 25[ | 70 |
| [25 à 30[ | 100 |
| [30 à 35[ | 65 |
| [35 à 40[ | 69 |
| [40 à 45[ | 56 |
| [45 à 50[ | 63 |
| [50 à 55[ | 61 |
| 55 et plus | 2 |

a) Représenter graphiquement cette distribution de fréquences.
b) Représenter le diagramme des fréquences cumulées croissantes.
c) Quel est le nombre de femmes actives âgées au moins de 25 ans?
d) Quelle est la proportion des femmes actives âgées de plus de 30 ans ?

### 1.3.6. Exercice.

Le tableau suivant donne le niveau de scolarité en nombre d'années passées à l'école d'un échantillon de 200 personnes.

| Niveau de scolarité | Effectif |
|---|---|
| [0 ; 6[ | 40 |
| [6 ; 12[ | 80 |
| [12 ; 14[ | 50 |
| [14 ; 16[ | 30 |
| **Total** | **200** |

a) Représenter graphiquement cette distribution de fréquences.
b) Représenter le diagramme des fréquences cumulées croissantes.
c) Quel est le nombre de personnes ayant un niveau de moins de 12 années passées à l'école?
d) Quel est la proportion des personnes ayant un niveau d'au moins 12 années passées à l'école?

### 1.3.7. Exercice.

Soit la répartition des travailleurs d'une entreprise selon l'âge :

11 % d'entre eux ont moins de 20 ans ;
31 % d'entre eux ont de 20 à 25 ans ;
26 % d'entre eux ont de 25 à 30 ans ;
16 % d'entre eux ont de 30 à 35 ans ;
7 % d'entre eux ont de 35 à 40 ans ;
9 % d'entre eux ont 40 ans et plus.

a) Représenter graphiquement cette distribution.
b) Représenter le diagramme des fréquences cumulées croissantes.

### 1.3.8. Exercice.

Un organisme chargé de réaliser des enquêtes statistiques gère un réseau de 125 enquêteurs. La direction de cet organisme décide d'étudier la répartition de ses enquêteurs selon le nombre d'enquêtes qu'ils ont réalisées. Les données collectées à ce sujet sont résumées dans le tableau ci-après :

| Nombre d'enquêtes réalisées | Effectifs |
|---|---|
| 5 | 8 |
| 10 | 12 |
| 15 | 35 |
| 20 | 40 |
| 25 | 20 |
| 30 | 10 |

Représenter graphiquement cette série statistique.
a- Par un polygone des fréquences relatives
b- Par une courbe des fréquences relatives cumulées croissantes.

### 1.3.9. Exercice.

Le tableau suivant donne la distribution de fréquences du nombre d'enfants dans 300 familles.

| Nombre d'enfants | Nombre de familles |
|---|---|
| 0 | 13 |
| 1 | 22 |
| 2 | 46 |
| 3 | 49 |
| 4 | 58 |
| 5 | 42 |
| 6 | 39 |
| plus de 6 | 31 |

| **Total** | **300** |
|---|---|

a) Calculer les différents types de fréquences cumulées.
b) Etablir le diagramme de fréquences et le diagramme de fréquences cumulées.
c) Quel est le nombre de familles ayant au plus 4 enfants ?
d) Quel est le nombre de familles ayant au moins 2 enfants ?
e) Quel est le pourcentage des familles qui n'ont pas d'enfants ?
f) Quel est le pourcentage des familles qui ont des enfants ?
g) Quel est le pourcentage des familles qui ont moins de 4 enfants?

### 1.3.10. Exercice.

Une coopérative laitière fabrique un fromage qui doit contenir, selon les étiquettes, 45 % de matières grasses. Un institut de consommation dont le rôle est de vérifier que la qualité des produits est bien celle qui est affirmée par l'étiquette, fait prélever et analyser un échantillon de 100 fromages. Les résultats de l'analyse sont consignés dans le tableau suivant :

| **Taux de matières grasses** | **Nombre de fromages** |
|---|---|
| [41,5 - 42,5[ | 1 |
| [42,5 - 43,5[ | 11 |
| [43,5 - 44,5[ | 24 |
| [44,5 - 45,5[ | 38 |
| [45,5 - 46,5[ | 22 |
| [46,5 - 47,5[ | 3 |
| [47,5 - 48,5[ | 1 |

a) Représenter graphiquement cette distribution.
b) Représenter le diagramme des fréquences cumulées croissantes.

# CHAPITRE 2
# CARACTERISTIQUES DE TENDANCE CENTRALE

**Les caractéristiques de tendance centrale, appelées aussi paramètres de position, servent à caractériser l'ordre de grandeur des observations. Les principaux paramètres de position sont : les moyennes, le mode, la médiane, et la médiale.**

Pour les caractéristiques centrales, nous ne nous intéressons qu'aux séries statistiques relatives à des caractères quantitatifs discrets ou continus, c'est-à-dire des séries statistiques données sous les formes : $(x_i)$ , $(x_i ; n_i)$ ; $(x_i ; f_i)$ ; $(ci ; n_i)$ ou $(c_i ; f_i)$.

## 2.1. LES MOYENNES.

**On peut réduire un ensemble d'observations en une seule observation constante appelée moyenne. La moyenne est donc une valeur qui se présente comme si toutes les observations lui étaient égales.**

On distingue plusieurs types de moyennes :

- la moyenne arithmétique ;
- la moyenne géométrique ;
- la moyenne harmonique ;
- la moyenne quadratique.

### 2.1.1. Moyenne arithmétique.

#### 2.1.1.1. Moyenne arithmétique simple.

**La moyenne arithmétique simple, qu'on appelle couramment moyenne, d'une série de plusieurs observations est égale à la somme de toutes les observations divisée par le nombre de ces observations.**

Dans le cas d'une suite de n observations : $x_1$, $x_2$, …, $x_i$, …, $x_n$ la moyenne est égale, par définition à :

$$\bar{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

L'introduction du terme $\sum_{i=1}^{n} x_i$ doit être explicitée, en effet on convient habituellement d'écrire, en mathématique :

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \ldots + x_i + \ldots + x_n$$

Dans le cas d'une série statistique donnée par un ensemble $(x_i , n_i)$, c'est-à-dire lorsque chaque valeur $x_i$ est répétée $n_i$ fois et qu'il y a k valeurs $x_i$ différentes, la moyenne arithmétique simple d'une telle série se déduit de la formule précédente :

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i} \quad \text{avec} \quad n = \sum_{i=1}^{k} n_i$$

De même dans le cas d'une série statistique donnée par un ensemble $(x_i , f_i)$ la moyenne arithmétique simple se déduit de la formule précédente :

$$\bar{x} = \sum_{i=1}^{k} f_i x_i$$

$$\text{avec } n = \sum_{i=1}^{k} n_i \quad ; \quad f_i = \frac{n_i}{n} \quad \text{et} \quad \sum_{i=1}^{k} f_i = 1$$

Dans le cas d'une variable statistique continue groupée en classes, la moyenne arithmétique simple est donnée par les formules suivantes :

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i c_i}{\sum_{i=1}^{k} n_i} = \sum_{i=1}^{k} f_i c_i$$

$c_i$ est le point central de la classe i, il est tel que : $c_i = \frac{C_i + C_{i+1}}{2}$

**Exemple 1 :** On considère l'ensemble des notes obtenues par les étudiants d'une classe d'une école, dans une matière ; on a la série statistique suivante donnée sous la forme simple ($x_i$) et pour laquelle on demande de calculer la moyenne arithmétique simple.

| | | | | |
|---|---|---|---|---|
| 12 | 11 | 13 | 12 | 13 |
| 15 | 13 | 12 | 13 | 11 |
| 13 | 15 | 11 | 11 | 12 |
| 12 | 12 | 10 | 12 | 15 |

La moyenne arithmétique simple de cette série est facile à calculer, elle est égale à :

$$\bar{x} = \frac{\sum_{i=1}^{20} x_i}{n} = \frac{248}{20} = 12{,}4$$

**Exemple 2 :** On considère la même série statistique qu'on représente maintenant sous la forme ($x_i$ ; $n_i$) pour laquelle on demande de calculer la moyenne arithmétique simple.

| $\mathbf{x_i}$ | $\mathbf{n_i}$ |
|---|---|
| 10 | 1 |
| 11 | 4 |
| 12 | 7 |
| 13 | 5 |
| 15 | 3 |

Le calcul de la moyenne arithmétique simple peut être facilement fait selon le tableau suivant :

| $x_i$ | $n_i$ | $n_i x_i$ |
|---|---|---|
| 10 | 1 | 10 |
| 11 | 4 | 44 |
| 12 | 7 | 84 |
| 13 | 5 | 65 |
| 15 | 3 | 45 |
| Total | 20 | 248 |
| **Moyenne** | **- - -** | **12,4** |

**Exemple 3 :** On considère la même série statistique qu'on représente sous la forme ($x_i$ ; $f_i$) pour laquelle on demande de calculer la moyenne arithmétique simple.

| $x_i$ | $n_i$ | $f_i$ |
|---|---|---|
| 10 | 1 | 5% |
| 11 | 4 | 20% |
| 12 | 7 | 35% |
| 13 | 5 | 25% |
| 15 | 3 | 15% |
| **Total** | **20** | **100%** |

Le calcul de la moyenne arithmétique simple peut être facilement fait selon le tableau suivant :

| $x_i$ | $n_i$ | $f_i$ | $f_i x_i$ |
|---|---|---|---|
| 10 | 1 | 0,05 | 0,5 |
| 11 | 4 | 0,20 | 2,2 |
| 12 | 7 | 0,35 | 4,2 |
| 13 | 5 | 0,25 | 3,25 |
| 15 | 3 | 0,15 | 2,25 |
| Total | 20 | 100% | 12,4 |
| **Moyenne** | **- - -** | **- - -** | **12,4** |

On voit, sur ces 3 exemples, que pour calculer la moyenne arithmétique simple, on utilise l'une des 3 formules selon la forme dans laquelle est donnée la série statistique.

**Exemple 4 :** On a procédé au recensement des 50 salariés de la société STM en relevant les salaires horaires qu'ils perçoivent. Les données brutes sont :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 34 | 36 | 45 | 62 | 37 | 43 | 42 | 102 | 31 | 42 |
| 51 | 30 | 61 | 63 | 47 | 105 | 52 | 43 | 81 | 95 |
| 92 | 77 | 60 | 36 | 48 | 49 | 65 | 71 | 78 | 81 |
| 43 | 52 | 63 | 71 | 43 | 42 | 51 | 55 | 61 | 41 |
| 93 | 82 | 83 | 47 | 54 | 61 | 102 | 33 | 48 | 55 |

**La moyenne arithmétique simple d'une telle série est égale à :**

$$\bar{x} = \frac{\sum_{i=1}^{50} x_i}{50} = \frac{2939}{50} = 58{,}78\ \text{DH / h}$$

Chaque salarié de la société touche, en moyenne, 58,78 DH par heure.

**Exemple 5 :** Une enquête, chez 1000 commerçants, porte sur le nombre d'agents qu'ils emploient. Les résultats obtenus sont représentés dans le tableau suivant :

| **Nombre d'employés** $x_i$ | **Nombre de commerçants** $n_i$ | **proportion des commerçants** $f_i$ |
|---|---|---|
| 0 | 50 | 5 % |
| 1 | 100 | 10 % |
| 2 | 200 | 20 % |
| 3 | 150 | 15 % |
| 4 | 120 | 12 % |
| 5 | 160 | 16 % |
| 6 | 130 | 13 % |
| 7 | 90 | 9 % |
| **Total** | **1000** | **100 %** |

**La moyenne arithmétique simple d'une telle série est égale à :**

$$\bar{x} = \frac{\sum_{i=1}^{8} n_i x_i}{\sum_{i=1}^{8} n_i} = \sum_{i=1}^{8} f_i x_i = \frac{3640}{1000} = 3{,}64\ \text{employés par commerçant}$$

Chaque commerçant emploie, en moyenne, trois à quatre employés.

**Exemple 6 :** La répartition de la surface, en m², de 100 logements est représentée dans le tableau suivant :

| Surface en m² | Nombre de logements | Point central |
|---|---|---|
| 0 à 20 | 10 | 10 |
| 20 à 40 | 20 | 30 |
| 40 à 60 | 40 | 50 |
| 60 à 100 | 18 | 80 |
| 100 à 160 | 8 | 130 |
| 160 à 260 | 4 | 210 |

**La moyenne arithmétique simple d'une telle série est égale à :**

$$\bar{x} = \frac{\sum_{i=1}^{6} n_i c_i}{\sum_{i=1}^{6} n_i} = \sum_{i=1}^{6} f_i c_i = \frac{6020}{100} = 60{,}20 \text{ m² par logement}$$

La superficie moyenne d'un logement est de 60,20 m².

### 2.1.1.2. Moyenne arithmétique pondérée.

**La moyenne arithmétique simple suppose que toutes les observations ont la même importance, ce qui n'est pas toujours le cas. La moyenne arithmétique pondérée intervient dans le cas où les observations n'ont pas la même importance. Il s'agit d'associer à chaque observation un coefficient de pondération indiquant son poids parmi les autres observations.**

$$\bar{x} = \frac{\sum_{i=1}^{k} \sim_i x_i}{\sum_{i=1}^{k} \sim_i}$$

$\alpha_i$ est le poids affecté à l'observation i.

**Exemple 7** : Un étudiant a eu 14 sur 20 au contrôle continu, 12 sur 20 à l'examen partiel et 13 sur 20 à l'examen final. Les trois notes n'ont pas la même importance. On associe un coefficient de 1 à la note du contrôle, un coefficient de 2 à la note de l'examen partiel, et un coefficient de 4 à la note de l'examen final. La note moyenne de l'année obtenue par cet étudiant est :

$$\bar{x} = \frac{\sum_{i=1}^{3} \alpha_i x_i}{\sum_{i=1}^{3} \alpha_i} = \frac{1\times14 + 2\times12 + 4\times13}{1 + 2 + 4} = 12{,}86$$

### 2.1.1.3 Propriétés de la moyenne arithmétique.

*** Propriété 1 : Transformation linéaire**

**La transformation linéaire d'une variable statistique x en une autre variable y telle que :**

$y = ax + b$ avec a et b deux constantes quelconques

**La moyenne de y peut être obtenue directement à partir de la moyenne de x :**

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n} (ax_i + b)}{n} = \frac{a\sum_{i=1}^{n} x_i + n\times b}{n}$$

$$\bar{y} = a\times\frac{\sum_{i=1}^{n} x_i}{n} + b \quad = \quad a\,\bar{x} + b$$

**La moyenne d'une transformation linéaire est donc une transformation linéaire de la moyenne.**

**Exemple 8 :** Le tableau suivant présente les prix en DH de 100 ordinateurs portables achetés dans différents points de vente :

| Prix | Nombre d'ordinateurs |
|---|---|
| 10000 – 11000 | 9 |
| 11000 – 12000 | 10 |
| 12000 – 13000 | 10 |
| 13000 – 14000 | 14 |
| 14000 – 15000 | 16 |
| 15000 – 16000 | 14 |
| 16000 – 17000 | 12 |
| 17000 – 18000 | 15 |
| **Total** | **100** |

Pour calculer la moyenne des prix des ordinateurs, on peut utiliser la propriété de la transformation linéaire dans le but de simplifier les calculs.

On effectue un changement de variable, c'est-à-dire, on remplace la variable prix par une autre variable y de telle sorte que le prix soit une transformation linéaire de y.

$p = a\,y + b$ Donc : $y = \dfrac{p-b}{a}$

Il faut choisir les constantes a et b qui donnent des valeurs très simples de y. On choisit la constante b parmi les valeurs de p, de préférence une valeur du milieu, pour avoir une valeur nulle de y au milieu. On choisit la constante a comme étant le plus grand diviseur commun des valeurs de (p - b) (le plus souvent a est l'amplitude constante des classes) pour avoir des valeurs entières de y.

Pour notre exemple, on choisit :

b = 13500 et a = 1000

$$Y = \frac{p-13500}{1000}$$

Les valeurs de y sont très simples, on peut calculer facilement la moyenne de y.

| **Prix** | **Nombre d'ordinateurs ($n_i$)** | **Point central ($c_i$)** | **$y_i$** | **$n_i\ y_i$** |
|---|---|---|---|---|
| 10000 – 11000 | 9 | 10500 | -3 | -27 |
| 11000 – 12000 | 10 | 11500 | -2 | -20 |
| 12000 – 13000 | 10 | 12500 | -1 | -10 |
| 13000 – 14000 | 14 | 13500 | 0 | 0 |
| 14000 – 15000 | 16 | 14500 | 1 | 16 |
| 15000 – 16000 | 14 | 15500 | 2 | 28 |
| 16000 – 17000 | 12 | 16500 | 3 | 36 |
| 17000 – 18000 | 15 | 17500 | 4 | 60 |
| **Total** | **100** | | | **83** |

$$\bar{y} = \frac{\sum_{i=1}^{8} n_i y_i}{\sum_{i=1}^{8} n_i} = \frac{83}{100} = 0{,}83$$

On calcule facilement la moyenne grâce aux formules de la transformation linéaire :

$$\bar{p} = 1000 \times \bar{y} + 13500 = 1000 \times 0{,}83 + 13500 = 14330 \text{ DH}$$

* Propriété 2 **:** La moyenne des écarts par rapport à la moyenne est nulle.

La somme des différences par rapport à la moyenne est toujours nulle.

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

*** Propriété 3 : La somme des carrées des écarts par rapport à la moyenne est minimale.**

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

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

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

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

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

Cette expression est positive, elle est donc minimale lorsque :

$$(\bar{x} - a)^2 = 0 \text{ c'est à dire lorsque } a = \bar{x}$$

### 2.1.2. Moyenne géométrique.

#### 2.1.2.1. Moyenne géométrique simple.

La moyenne géométrique simple est calculée pour des observations positives. Elle est égale à la racine $n^{ème}$ du produit de l'ensemble des n observations. Elle est utilisée principalement lorsqu'on raisonne en taux de croissance.

La moyenne géométrique est égale, par définition, dans le cas d'une suite de n observations $x_1, x_2, \ldots, x_i, \ldots, x_n$ à :

$$\bar{x}_g = \sqrt[n]{x_1 \times x_2 \times \cdots \times x_n} = (x_1 \times x_2 \times \cdots \times x_n)^{\frac{1}{n}} = [\prod_{i=1}^{n} x_i]^{\frac{1}{n}}$$

**Exemple 9** : On considère une action qui a accusé, en bourse, durant le 1er semestre de l'année 2005, les taux d'augmentation mensuels suivants : +2,1% ; 1,3% ; 0,5% ; 0,9% ; 1,4% ; 3,8%. Calculer le taux d'augmentation mensuel moyen de l'action durant le 1er semestre 2005.

C'est l'exemple type de l'application de la moyenne géométrique simple :

**Remarque** : Rappelons que pour une variable qui a accusé un taux d'augmentation de 2% par exemple, on multiplie cette variable par 1,02 pour trouver la nouvelle valeur de la variable.

Ainsi si l'action a comme valeur 25,35 DH en Janvier et qu'elle subisse un taux

d'augmentation de 2,1% entre janvier et février, sa valeur, en février est égale à :

25,35 x 1,021 = 25,88 DH.

Donc nous allons, tout le temps, utiliser cette remarque lorsqu'il s'agit de taux.

Revenons à l'exemple 9 et calculons le taux d'augmentation mensuel moyen de l'action :

$$\bar{t} = \sqrt[6]{1{,}021 \times 1{,}013 \times 1{,}005 \times 1{,}009 \times 1{,}014 \times 1{,}038} - 1 = 1{,}66\,\%$$

**Exemple 10** : La population marocaine est passée, entre 1994 et 2004 de 26 019 000 à 29 800 000.

Quel est le taux global d'augmentation de la population pendant les 10 années ?

Quel est le taux annuel moyen d'augmentation de la population ?

Entre 1994 et 2004, le taux global d'accroissement de la population marocaine est :

$$t = \frac{29800 - 26019}{26019} \times 100 = 14{,}53\%$$

Le taux d'accroissement annuel moyen est $\bar{t}$ tel que :

$$26019 \times (1+\bar{t})^{10} = 29800$$

$$(1+\bar{t})^{10} = \frac{29800}{26019} = 1{,}1453$$

$$\bar{t} = \sqrt[10]{1{,}1453} - 1 = 0{,}0137 = 1{,}37\ \%$$

Entre 1994 et 2004, la population marocaine a augmenté en moyenne, de 1,37 % par an.

**2.1.2.2. Moyenne géométrique pondérée.**

**De même que pour la moyenne arithmétique simple qui suppose que toutes les observations aient la même importance, ce qui n'est pas toujours le cas, la moyenne géométrique pondérée intervient dans le cas où les observations n'ont pas la même importance. Il s'agit d'associer à chaque observation un coefficient de pondération indiquant son poids parmi les autres observations.**

$$\bar{x}_g = \sqrt[\alpha]{x_1^{\alpha_1} \times x_2^{\alpha_2} \times \cdots \times x_n^{\alpha_n}}$$

$$\bar{x}_g = (x_1^{\alpha_1} \times x_2^{\alpha_2} \times \cdots \times x_n^{\alpha_n})^{\frac{1}{n}}$$

$$\bar{x}_g = \left[\prod_{i=1}^{k} x_i^{\alpha_i}\right]^{\frac{1}{\alpha}} = \prod_{i=1}^{k} x_i^{f_i}$$

$\alpha_i$ est le poids affecté à l'observation i.

Avec $\alpha = \sum_{i=1}^{k} \alpha_i$

C'est le cas de séries statistiques discrètes données sous la forme ($x_i$ ; $n_i$) ou ($x_i$ ; $f_i$), lorsque, dans les séries, la variable $x_i$ est répétée $n_i$ fois (ou $f_i$ en %) et qu'il y a k observations distinctes.

**Dans le cas d'une série statistique continue, on définit la moyenne géométrique pondérée comme suit :**

$$\bar{x}_g = \sqrt[n]{c_1^{n_1} \times c_2^{n_2} \times \ldots \times c_k^{n_k}} = (c_1^{n_1} \times c_2^{n_2} \times \ldots \times c_k^{n_k})^{\frac{1}{n}}$$

$$= [\prod_{i=1}^{k} c_i^{n_i}]^{\frac{1}{n}} = \prod_{i=1}^{k} c_i^{f_i}$$

$c_i$ est le point central de la classe i, il est tel que : $c_i = \frac{C_i + C_{i+1}}{2}$

**Exemple 11** : Etude du taux de variation d'une action en bourse.

Le tableau suivant donne l'évolution du taux d'augmentation de la valeur d'une action, en bourse, entre janvier et décembre 2005.

| **Périodes** | **Taux d'augmentation mensuel moyen** |
|---|---|
| Entre janvier et avril | 2,03% par mois en moyenne |
| Entre mai et juillet | 0,69% par mois en moyenne |
| Entre août et décembre | 2,13% par mois en moyenne |

Quel est le taux global de variation de la valeur de l'action entre janvier et décembre 2005 ?

Quel est le taux mensuel moyen de variation de la valeur de l'action entre janvier et décembre 2005 ?

S'agissant de taux d'augmentation mensuels relatifs à des périodes différentes, de nombres de mois différents, il y a lieu d'affecter chaque taux d'un poids égal aux nombres de mois contenu dans la période ;

Le taux d'augmentation global de la valeur de l'action est :

$$t = 1{,}02034 \times 1{,}00693 \times 1{,}02135 - 1 = 22{,}92\%$$

Le taux d'augmentation mensuel moyen de la valeur de l'action entre janvier et décembre 2005 est alors égal à :

$$\bar{t} = \sqrt[12]{1{,}2294} - 1 = 1{,}73\%$$

**2.1.2.3. Propriétés de la moyenne géométrique.**

**La moyenne géométrique est aussi égale à l'exponentielle de la moyenne arithmétique des logarithmes des variables statistiques.**

$$\text{Log}\,\bar{x}_g = \text{Log}[\prod_{i=1}^{n} x_i]^{\frac{1}{n}} = \frac{1}{n}\text{Log}[\prod_{i=1}^{n} x_i] = \frac{\sum_{i=1}^{n} \text{Log}\, x_i}{n}$$

$$\overline{x_g} = \exp\left(\frac{\sum_{i=1}^{n} Log\, x_i}{n}\right)$$

**2.1.3. Moyenne harmonique.**

**2.1.3.1. Moyenne harmonique simple.**

**La moyenne harmonique simple est égale à l'inverse de la moyenne arithmétique des inverses des observations. Son usage s'impose lorsque la variable statistique est un quotient (coût moyen, vitesse moyenne, etc.).**

Dans le cas d'une suite de n observations $x_1$, $x_2$, …, $x_i$, …, $x_n$, toutes distinctes et de poids identiques, la moyenne harmonique simple est égale à :

$$\bar{x}_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

**Exemple 12** : Calcul de la vitesse moyenne.

On considère un automobiliste qui fait 80 km et qui parcourt chaque 20 km avec des vitesses moyennes différentes, soient successivement 90 km/h, puis 75 km/h, ensuite 85 km/h et enfin 115 km/h.

Quelle est la vitesse moyenne de l'automobiliste ?

Comme il s'agit de vitesses moyennes, toutes relatives à la même distance de 20 km, elles doivent avoir le même poids. Montrons donc que la vitesse moyenne sur les 80 km est la moyenne harmonique des vitesses.

En effet, le temps t mis pour parcourir une distance d à la vitesse v est donné par la formule simple : t = d / v.

Ainsi le temps global t est la somme des quatre temps $t_i$ :

$$t = t_1 + t_2 + t_3 + t_4$$

$$\frac{d}{\overline{v}} = \frac{d_1}{V_1} + \frac{d_2}{V_2} + \frac{d_3}{V_3} + \frac{d_4}{V_4} = \sum_{i=1}^{4} \frac{d_i}{V_i}$$

En divisant les 2 membres de cette égalité par d et en constatant que di / d = 1 / 4 on trouve facilement le résultat recherché :

$$\frac{1}{\overline{v}} = \frac{\sum_{i=1}^{4} \frac{1}{V_i}}{4}$$

C'est-à-dire d'une façon plus générale : $\frac{1}{\overline{v}} = \frac{\sum_{i=1}^{n} \frac{1}{V_i}}{n}$

$$\frac{1}{\overline{v}} = \frac{1}{4}\left(\frac{1}{90} + \frac{1}{75} + \frac{1}{85} + \frac{1}{115}\right) = 0{,}01123$$

Soit après calcul : $\overline{v}$ = 89,077 km/h

**2.1.3.2. Moyenne harmonique pondérée.**

**La moyenne harmonique pondérée intervient dans le cas où les observations n'ont pas la même importance. Il s'agit d'associer à chaque observation un coefficient de pondération indiquant son poids parmi les autres observations.**

*** Cas d'une série statistique discrète** : dans laquelle la variable statistique $x_i$ est répétée $n_i$ fois (ou $f_i$ en %), lorsque la série est de la forme ($x_i$ ; $n_i$) ou ($x_i$ ; $f_i$).

$$\overline{x}_h = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} \frac{n_i}{x_i}} = \frac{1}{\sum_{i=1}^{k} \frac{f_i}{x_i}}$$

*** Cas d'une série statistique continue** :

$$\bar{x}_h = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} \frac{n_i}{c_i}} = \frac{1}{\sum_{i=1}^{k} \frac{f_i}{c_i}}$$

$c_i$ est le point central de la classe i, il est tel que : $c_i = \frac{C_i + C_{i+1}}{2}$

**Exemple 13** : Calcul de la vitesse moyenne.

**Reprenons l'exemple de l'automobiliste et supposons que maintenant il ait roulé sur un trajet de 100 Km à une vitesse de 90 Km/h, sur les 10 premiers kilomètres; de 100 Km/h sur un trajet de 30 Km, et de 120 Km/h sur les 60 derniers kilomètres.**

L'automobiliste a parcouru le trajet de 100 Km avec trois vitesses moyennes différentes sur des trajets de différentes longueurs :

| **Vitesses moyennes** | **Trajets** |
|---|---|
| $V_1$ = 90 km/h | $d_1$ = 10 km |
| $V_2$ = 100 km/h | $d_2$ = 30 km |
| $V_3$ = 120 km/h | $d_3$ = 60 km |
| **Total** | **100 km** |

Comme il s'agit de vitesses moyennes relatives à des distances différentes, elles doivent être affectées de poids différents. Montrons donc que la vitesse moyenne sur les 100 km est la moyenne harmonique pondérée des vitesses.

En effet, le temps t mis pour parcourir une distance d à la vitesse v est donné par la formule simple : t = d / v.

Ainsi le temps global t est la somme des quatre temps $t_i$ :

$$t = t_1 + t_2 + t_3 + t_4$$

$$\frac{d}{\bar{v}} = \frac{d_1}{V_1} + \frac{d_2}{V_2} + \frac{d_3}{V_3} + \frac{d_4}{V_4} = \sum_{i=1}^{4} \frac{d_i}{V_i}$$

En divisant les 2 membres de cette égalité par d et en posant $\sim_i = d_i / d$, on trouve facilement le résultat recherché :

$\frac{1}{v} = \sum_{i=1}^{4} \alpha_i \frac{1}{V_i}$ avec par exemple $\alpha_1 = \frac{10}{100} = 0{,}10$

C'est-à-dire d'une façon plus générale : $\frac{1}{v} = \sum_{i=1}^{n} \alpha_i \frac{1}{V_i}$

Après calcul, on trouve $\bar{v}$ = 109,8 km/h.

Exemple 14 **: Calcul du coût moyen d'un stock.**

**Calculer le coût moyen d'une pièce de rechange stockée dans le magasin de l'entreprise si l'on suppose que le stock ait été approvisionné, à différents prix, en plusieurs étapes.**

| Etapes | Nombre de pièces achetées | Prix unitaires des pièces |
|---|---|---|
| **N° 1** | **10** | **12,35 DH** |
| **N° 2** | **25** | **13,12 DH** |
| **N° 3** | **20** | **13,46 DH** |
| **N° 4** | **45** | **14,07 DH** |

**Comme le coût est un rapport, montrons que le coût moyen est la moyenne harmonique pondérée des différents coûts. En effet, les coûts moyens auxquels les pièces de rechange ont été achetées sont relatifs à des lots de différentes tailles, ce qui fait que ces coûts doivent être affectés de différents poids.**

**Convenons d'appeler, dans ce qui suit, pour le lot i, $cu_i$ le coût unitaire, $ct_i$ le coût total et $n_i$ le nombre de pièces de rechange achetées.**

**Nous avons l'égalité suivante évidente relative aux nombres de pièces de rechange :**

$$n = n_1 + n_2 + n_3 + n_4 = \sum_{i=1}^{4} n_i$$

**Or comme** $n_i = \frac{ct_i}{cu_i}$ **on a :** $n = \frac{ct}{\overline{cu}} = \sum_{i=1}^{4} \frac{ct_i}{cu_i}$

**En divisant les 2 membres de la dernière égalité par ct et en posant $\sim_i = ct_i / ct$ on trouve la formule recherchée, à savoir :**

$$\frac{1}{\overline{cu}} = \sum_{i=1}^{4} \frac{ct_i}{ct} \frac{1}{cu_i} = \sum_{i=1}^{4} \alpha_i \frac{1}{cu_i}$$

**Avec par exemple :**

$$\alpha_2 = \frac{ct_2}{\sum_{i=1}^{4} ct_i} = \frac{25 \times 13{,}12}{10 \times 12{,}35 + 25 \times 13{,}12 + 20 \times 13{,}46 + 45 \times 14{,}07}$$

$$= 0{,}2422$$

**Le coût moyen d'approvisionnement de la pièce de rechange est, après calculs, égal à : 13,51 DH/unité.**

### 2.1.4. Moyenne quadratique.

**La moyenne quadratique est la racine carrée de la moyenne arithmétique des carrées. Elle est très rarement utilisée.**

*** Cas d'une suite de n observations** : $x_1, x_2, \ldots, x_i, \ldots, x_n$

$$\bar{x}_q = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

*** Cas d'une série statistique discrète** : lorsque chaque variable $x_i$ est répétée $n_i$ (ou $f_i$ en %) fois dans la série et qu'il y a k valeurs différentes.

$$\bar{x}_q = \sqrt{\frac{\sum_{i=1}^{k} n_i x_i^2}{\sum_{i=1}^{k} n_i}} = \sqrt{\sum_{i=1}^{k} f_i x_i^2}$$

avec $n = \sum_{i=1}^{k} n_i$ et $\sum_{i=1}^{k} f_i = 1$

*** Cas d'une série statistique continue** :

$$\bar{x}_q = \sqrt{\frac{\sum_{i=1}^{k} n_i c_i^{\,2}}{\sum_{i=1}^{k} n_i}} = \sqrt{\sum_{i=1}^{k} f_i c_i^{\,2}}$$

avec $n = \sum_{i=1}^{k} n_i$ et $\sum_{i=1}^{k} f_i = 1$

$c_i$ est le point central de la classe i, il est tel que : $c_i = \frac{C_i + C_{i+1}}{2}$

**Exemple 15** : Dans une entreprise produisant des pièces pour l'assemblage d'une machine on veut contrôler si la longueur moyenne des pièces est conforme à la norme de 12 cm. La production est jugée comme conforme si l'écart moyen par rapport à la norme ne dépasse pas 1 cm. À cette fin on a mesuré la longueur d'un échantillon de 16 pièces dont les résultats sont :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 11 | 10 | 12,5 | 10,8 | 13,5 | 11,5 | 13 | 12,5 |
| 13 | 13,5 | 11,5 | 13,2 | 10,5 | 12,5 | 11 | 11,5 |

Peut-on admettre que le produit de l'entreprise est conforme à la norme ?

Calculons les écarts par rapport à la norme :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| -1 | -2 | 0,5 | -1,2 | 1,5 | -0,5 | 1 | 0,5 |
| +1 | 1,5 | -0,5 | 1,2 | -1,5 | 0,5 | -1 | -0,5 |

On voit bien que certains écarts sont positifs et d'autres sont négatifs ; le calcul de la moyenne arithmétique n'est pas approprié car les écarts négatifs vont compenser les écarts positifs. La moyenne qu'il faut calculer est la moyenne quadratique.

$$\bar{x}_q = \sqrt{\frac{\sum_{i=1}^{n} x_i^{\,2}}{n}} = \sqrt{\frac{(-1)^2 + (-2)^2 + 0{,}5^2 + (-1{,}2)^2 + 1{,}5^2 + (-0{,}5)^2 + 1^2 + 0{,}5^2}{16} + \frac{1^2 + 1{,}5^2 + (-0{,}5)^2 + 1{,}2^2 + (-1{,}5)^2 + 0{,}5^2 + (-1)^2 + (-0{,}5)^2}{16}}$$

$\bar{x}_q$ = 1,09 cm

L'écart moyen par rapport à la norme est de 1,09 cm, il dépasse l'écart moyen toléré qui est de 1 cm, on ne peut donc admettre que le produit de l'entreprise est conforme à la norme.

Remarque **: On peut montrer que la moyenne harmonique est inférieure ou égale à la moyenne géométrique qui est inférieure ou égale à la moyenne arithmétique qui est inférieure ou égale à la moyenne quadratique.**

$$\overline{x_h} \leq \overline{x_g} \leq \overline{x} \leq \overline{x_q}$$

Exemple 16 **: On peut aisément vérifier de telles inégalités dans l'exemple simple suivant.**

**On considère la série statistique simple constituée des cinq observations suivantes : 2 ; 5 ; 6 ; 8 et 10.**

**On trouve, après un calcul facile que :**

$$\frac{1}{\overline{x_h}} = \frac{1}{4} \times (\frac{1}{2} + \frac{1}{5} + \frac{1}{6} + \frac{1}{8} + \frac{1}{10}) = 0{,}2729 \Rightarrow \overline{x_h} = 3{,}664$$

$$\overline{x_g} = \sqrt[5]{2 \times 5 \times 6 \times 8 \times 10} = 5{,}448$$

$$\overline{x} = \frac{2+5+6+8+10}{5} = 6{,}2$$

$$\overline{x_q} = \sqrt{\frac{2^2+5^2+6^2+8^2+10^2}{5}} = 6{,}767$$

**Et l'on a bien :**

$$(\overline{x_h} = 3{,}664) \leq (\overline{x_g} = 5{,}448) \leq (\overline{x} = 6{,}2) \leq (\overline{x_q} = 6{,}767)$$

**2.2. LE MODE.**

Le mode est l'observation la plus fréquente dans une série statistique.

*** Cas d'une suite de n observations :** Le mode d'une série statistique est l'observation que l'on rencontre le plus fréquemment. Le mode peut ne pas exister, et s'il existe, il peut ne pas être unique.

**Exemple 17** : On considère les séries d'observations suivantes :

a) 3; 5 ; 8 ; 8 ; 8 ; 10 ; 10 ; 10 ; 10 ; 10 ; 14 ; 18 ; 20 ; 24 ; 24
b) 4 ; 8 ; 10 ; 10 ; 10 ; 10 ; 14 ; 18 ; 22 ; 22 ; 22 ; 22 ; 26
c) 5, 11, 14, 17, 18, 21, 23, 26, 29, 30, 32, 35, 38
d) 12 ; 23 ; 34 ; 23 ; 35 ; 23 ; 52 ; 23 ; 33 ; 56 ; 23 ; 23 ; 40

Dans ces exemples, on a successivement :

- pour le cas a : Le mode est 10.
- pour le cas b : Il y a deux modes, 10 et 22.
- pour le cas c : Le mode n'existe pas.
- pour le cas d : Le mode est 23

*** Cas d'une série statistique discrète** : Le mode correspond à la valeur qui possède la plus grande fréquence.

**Exemple 18** : Soit la distribution du nombre d'employés observés chez 1000 commerçants.

| **Nombre d'employés $x_i$** | **Nombre de commerçants ($n_i$)** | **proportion des commerçants ($f_i$)** |
|---|---|---|
| 0 | 50 | 5 % |
| 1 | 100 | 10 % |
| 2 | 200 | 20 % |
| 3 | 150 | 15 % |
| 4 | 120 | 12 % |
| 5 | 160 | 16 % |
| 6 | 130 | 13 % |
| 7 | 90 | 9 % |
| **Total** | **1000** | **100 %** |

**La variable $x_i$ nombre d'employés a pour mode 2, c'est-à-dire la plupart des commerçants ont deux employés.**

*** Cas d'une série statistique continue** : Dans le cas d'une variable statistique continue groupée en classes, on parle de classe modale, elle correspond à la classe dont la fréquence est la plus élevée. Le mode correspond à la valeur de la variable qui correspond au maximum de l'histogramme. C'est le point central de la classe modale si les classes ont la même amplitude, dans le cas contraire, il faut travailler avec les densités.

**Exemple 19** : La répartition de la surface, en m², de 100 logements est représentée dans le tableau suivant :

| **Surface en m²** | **Nombre de logements** |
|---|---|
| 0 à 20 | 10 |
| 20 à 40 | 20 |
| 40 à 60 | 40 |
| 60 à 100 | 18 |
| 100 à 160 | 8 |
| 160 à 260 | 4 |

**Les amplitudes des classes étant inégales, il convient de calculer les densités.**

| Surface en m² | Nombre de logements | Densités |
|---|---|---|
| 0 à 20 | 10 | 0,5 |
| 20 à 40 | 20 | 1 |
| 40 à 60 | 40 | 2 |
| 60 à 100 | 18 | 0,45 |
| 100 à 160 | 8 | 0,13 |
| 160 à 260 | 4 | 0,04 |
| **Total** | **100** | |

**En cherchant la plus grande densité, la classe modale est la classe 40 à 60 m², le mode est égal au centre de la classe modale, à savoir : 50 m².**

### 2.3. LA MEDIANE.

**La médiane d'une variable statistique est une valeur pour laquelle, la moitié des observations lui sont inférieure ou égales et la moitié supérieures ou égales. La médiane partage donc le nombre total d'observations en deux parties égales. La médiane est un paramètre statistique qui ne dépend que du nombre d'observations.**

**Pour déterminer la médiane, il faut raisonner en terme de fréquences cumulées, la médiane est alors la valeur de la variable qui correspond à la moitié de l'effectif total.**

*** Cas d'une série statistique discrète.**

Si le nombre d'observations est impair, la médiane est l'observation de rang $\frac{n+1}{2}$.

$$Me = x_{\frac{n+1}{2}}$$

Si le nombre d'observations est pair, la médiane est comprise entre l'observation de rang $\frac{n}{2}$ et l'observation de rang $\frac{n}{2}+1$. On prend comme valeur de la médiane la moyenne arithmétique simple des deux observations.

$$x_{\frac{n}{2}} \leq Me \leq x_{\frac{n}{2}+1}$$

$$Me = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

**Exemple 20** : Soit la distribution du nombre d'employés observés chez 1000 commerçants.

| **Nombre d'employés** $x_i$ | **Nombre de commerçants** $(n_i)$ | **Fréquences cumulées croissantes** $F_{ic}$ |
|---|---|---|
| 0 | 50 | 50 |
| 1 | 100 | 150 |
| 2 | 200 | 350 |
| 3 | 150 | 500 |
| 4 | 120 | 620 |
| 5 | 160 | 780 |
| 6 | 130 | 910 |
| 7 | 90 | 1000 |
| **Total** | **1000** | |

Le nombre d'observations, 1000, est pair, la médiane est comprise entre l'observation de rang 500 et l'observation de rang 501. On prend comme valeur de la médiane la moyenne arithmétique simple des deux observations.

$$x_{500} \leq Me \leq x_{501}$$

$$Me = \frac{x_{500} + x_{501}}{2}$$

En consultant les fréquences absolues cumulées croissantes, $x_{500}$ correspond à 3 et $x_{501}$ correspond à 4. La médiane est donc :

$$Me = \frac{3+4}{2} = 3{,}5$$

**La moitié des commerçants emploient 3 employés ou moins, et la moitié emploient 4 employés ou plus.**

*** Cas d'une série statistique continue.**

**Pour des données groupées en classes, la classe médiane est la classe qui contient la médiane. On détermine la médiane par interpolation linéaire.**

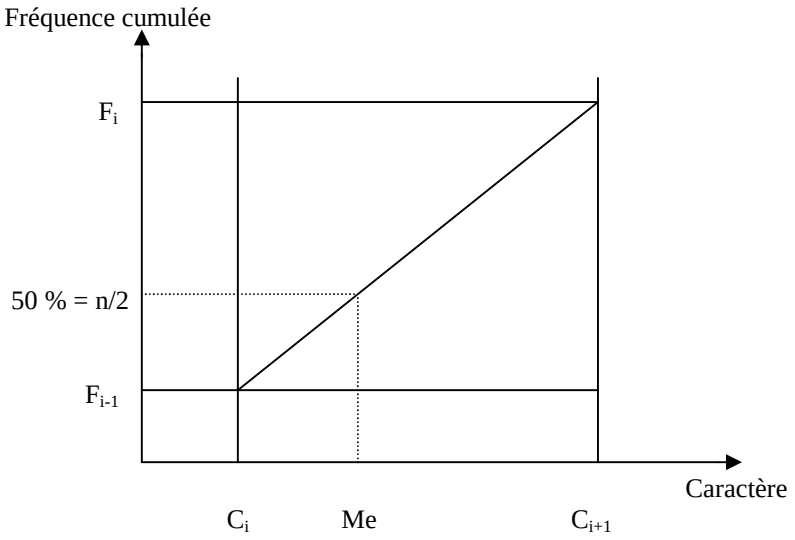
Désignons par :

$[C_i ; C_{i+1}[$ : la classe médiane ;
n : le nombre total des observations ;
$F_i$ : la fréquence absolue cumulée croissante ;
$n_i$ : la fréquence absolue de la classe médiane.

La médiane est comprise entre $C_i$ et $C_{i+1}$

$$C_i < Me < C_{i+1}$$

De même : $F_{i-1} < \frac{n}{2} < F_i$




**On suppose que la distribution au sein de la classe médiane soit régulière.**

Ainsi : $$\frac{C_{i+1} - C_i}{F_i - F_{i-1}} = \frac{Me - C_i}{\frac{n}{2} - F_{i-1}}$$

Ce qui donne : $$Me = C_i + \frac{C_{i+1} - C_i}{F_i - F_{i-1}} \times (\frac{n}{2} - F_{i-1})$$

**Exemple 21** : La répartition de la surface, en m², de 100 logements est représentée dans le tableau suivant :

| Surface en m² | Nombre de logements | F. cumulées croissantes |
|---|---|---|
| 0 à 20 | 10 | 10 |
| 20 à 40 | 10 | 20 |
| 40 à 60 | 50 | 70 |
| 60 à 100 | 18 | 88 |
| 100 à 160 | 8 | 96 |
| 160 à 260 | 4 | 100 |
| **Total** | **100** | |

**En consultant les fréquences absolues cumulées croissantes, la classe médiane est la classe 40 à 60 m². La médiane est donc :**

$$40 < Me < 60$$
$$20 < 50 < 70$$

$$\frac{60-40}{70-20} = \frac{Me-40}{50-20}$$

$$Me = 40 + \frac{20}{50} \times 30 = 52 \text{ m}^2$$

La moitié des logements ont une superficie inférieure ou égale à 52 m² et la moitié des logements ont une superficie supérieure ou égale à 52 m².

### 2.4. LA MEDIALE.

**La médiale est une valeur telle que la somme des observations qui lui sont inférieures est égale à la somme des observations qui lui sont supérieures. La médiale partage donc la somme des observations en deux parties égales. La médiale est un paramètre statistique qui dépend de la somme de toutes les observations.**

**Pour déterminer la médiale, il faut raisonner en terme de sommes cumulées, la médiale est alors la valeur de la variable qui correspond à la moitié de la somme des observations.**

La médiale calculée pour une variable statistique groupée en classes, la classe médiale est la classe qui contient la médiale. On détermine la médiale par interpolation linéaire.

Désignons par :

$[C_i ; C_{i+1}[$ : la classe médiale ;

$S = \sum_{i=1}^{k} n_i c_i$ : la somme des observations ;

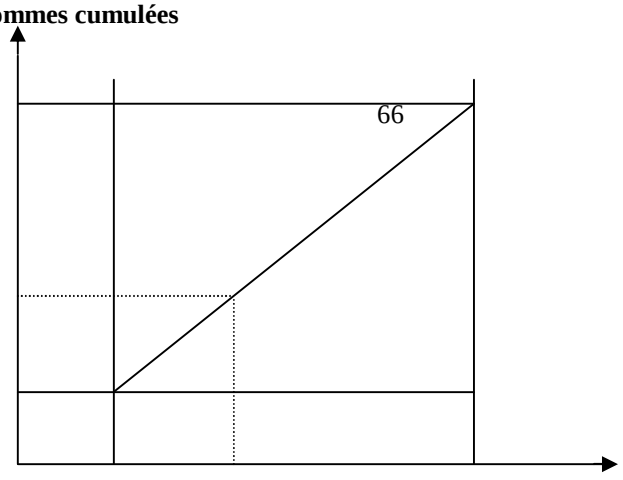
$S_i = \sum_{j=1}^{j=i} n_j c_j$ : la somme des observations cumulée croissante;

$n_i\ c_i$ : la somme des observations de la classe médiale.

La médiale est comprise entre $C_i$ et $C_{i+1}$

$$C_i < Ml < C_{i+1}$$
$$S_{i-1} < \frac{S}{2} < S_i$$

**Sommes cumulées**



$S_i$

50 % = S/2

$S_{i-1}$

Caractère

$C_i$ Ml $C_{i+1}$

**On suppose que la distribution au sein de la classe médiale soit régulière.**

Ainsi : $$\frac{C_{i+1} - C_i}{S_i - S_{i-1}} = \frac{Ml - C_i}{\frac{S}{2} - S_{i-1}}$$

$$Ml = C_i + \frac{C_{i+1} - C_i}{S_i - S_{i-1}} \times \left(\frac{S}{2} - S_{i-1}\right)$$

**Exemple 22** : La répartition de la surface, en m², de 100 logements est représentée dans le tableau suivant :

| **Surface en m²** | **Nombre de logements $n_i$** | **Point central $c_i$** | **Sommes $n_i x_i$** | **Sommes cumulées croissantes** |
|---|---|---|---|---|
| 0 à 20 | 10 | 10 | 100 | 100 |
| 20 à 40 | 20 | 30 | 600 | 700 |
| 40 à 60 | 40 | 50 | 2000 | 2700 |
| 60 à 100 | 18 | 80 | 1440 | 4140 |
| 100 à 160 | 8 | 130 | 1040 | 5180 |
| 160 à 260 | 4 | 210 | 840 | 6020 |
| **Total** | **100** | | **6020** | |

La moitié de la somme des observations est :

$$\frac{\sum_{i=1}^{6} n_i c_i}{2} = \frac{6020}{2} = 3010$$

**En consultant les sommes cumulées croissantes, la classe médiale est la classe 60 à 100 m². La médiale est donc :**

$$60 < Ml < 100$$
$$2700 < 3010 < 4140$$

$$\frac{100-60}{4140-2700} = \frac{Ml-60}{3010-2700}$$

$$Ml = 60 + \frac{40}{1440} \times 310 = 68{,}61 \text{ m}^2$$

**La moitié de la superficie totale des 100 logements est répartie sous forme de logements dont la superficie est inférieure ou égale à 68,61 m² et l'autre moitié sous forme de logements dont la superficie est supérieure ou égale à 68,61 m².**

### 2.5. LES FRACTILES.

**De même que la médiane nous a permis de partager la population en deux parties égales, le fractile d'ordre p permet de partager la population en p parties égales, chaque partie contient $\frac{100}{p}\%$ du nombre total des observations. Ainsi les quartiles, déciles, centiles vont respectivement nous permettre de partager la population respectivement en quatre, dix et cent parties égales.**

#### 2.5.1. Les quartiles.

**Les quartiles partagent le nombre total des observations en quatre parties égales, chaque partie contient 25% des observations. On définit trois quartiles.**

**Le premier quartile $Q_1$ : C'est une valeur pour laquelle un quart des observations (25%) lui sont inférieures ou égales et trois quarts des observations (75%) lui sont supérieures ou égales.**

**Le deuxième quartile $Q_2$ : C'est une valeur pour laquelle deux quarts des observations (50%) lui sont inférieures ou égales et deux quarts des observations (50%) lui sont supérieures ou égales. Il est aussi égal à la médiane.**

**Le troisième quartile $Q_3$ : C'est une valeur pour laquelle trois quarts des observations (75%) lui sont inférieures ou égales et un quart des observations (25%) lui sont supérieures ou égales.**

**Pour le calcul des quartiles, on utilise la même méthode de calcul que pour la médiane.**

**Pour des données groupées en classes, on détermine un quartile par interpolation linéaire.**

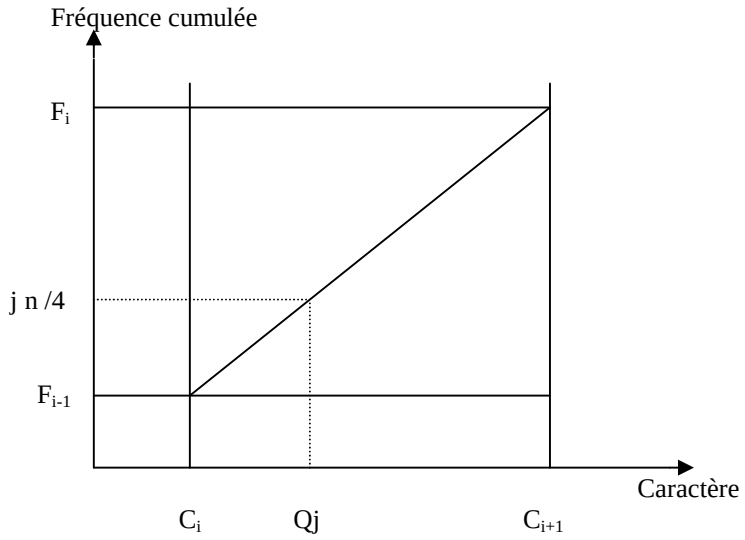
Désignons par :

$[C_i ; C_{i+1}[$ : la classe qui contient le quartile ;
n : le nombre total des observations ;
$F_i$ : la fréquence absolue cumulée croissante ;
$n_i$ : la fréquence absolue de la classe qui contient le quartile ;

Le quartile numéro j, Qj est compris entre $C_i$ et $C_{i+1}$

$$C_i < Qj < C_{i+1}$$

$$F_{i-1} < \frac{j \times n}{4} < F_i$$




**On suppose que la distribution au sein de la classe est régulière.**

Ainsi : $$\frac{C_{i+1} - C_i}{F_i - F_{i-1}} = \frac{Qj - C_i}{\frac{j \times n}{4} - F_{i-1}}$$

$$Qj = C_i + \frac{C_{i+1} - C_i}{F_i - F_{i-1}} \times \left(\frac{j \times n}{4} - F_{i-1}\right)$$

**Les trois quartiles sont :**

$$Q_1 = C_i + \frac{C_{i+1} - C_i}{F_i - F_{i-1}} \times \left(\frac{n}{4} - F_{i-1}\right)$$

$$Q_2 = C_i + \frac{C_{i+1} - C_i}{F_i - F_{i-1}} \times \left(\frac{n}{2} - F_{i-1}\right) = Me$$

$$Q_3 = C_i + \frac{C_{i+1} - C_i}{F_i - F_{i-1}} \times \left(\frac{3 \times n}{4} - F_{i-1}\right)$$

**Exemple 23** : La répartition de la surface, en m², de 100 logements est représentée dans le tableau suivant :

| Surface en m² | Nombre de logements | Fréquences cumulées croissantes |
|---|---|---|
| 0 à 20 | 10 | 10 |
| 20 à 40 | 20 | 30 |
| 40 à 60 | 40 | 70 |
| 60 à 100 | 18 | 88 |
| 100 à 160 | 8 | 96 |
| 160 à 260 | 4 | 100 |
| **Total** | **100** | |

**En consultant les fréquences absolues cumulées croissantes, q1, qui correspond à la 25ème observation, se trouve dans la classe 20 à 40 m². $q_3$, qui correspond à la 75ème observation, se trouve dans la classe 60 à 100 m².**

$$q_1 = 20 + 20 \times \frac{\frac{100}{4} - 10}{20} = 35\,m^2$$

$$q_3 = 60 + 40 \times \frac{\frac{3 \times 100}{4} - 70}{18} = 71{,}11\,m^2$$

**25 % des logements ont une superficie inférieure ou égale à 35 m².**

**75 % des logements ont une superficie inférieure ou égale à 71,11 m².**

**50 % des logements ont une superficie comprise entre 35 m² et 71,11 m².**

2.5.2. Les déciles.

**Les déciles partagent le nombre total des observations en dix parties égales, chaque partie contient 10% des observations. On définit neuf déciles.**

**Le premier décile $d_1$ : C'est une valeur pour laquelle un dixième des observations (10%) lui sont inférieures ou égales et neuf dixièmes des observations (90%) lui sont supérieures ou égales.**

**Le deuxième décile $d_2$ : C'est une valeur pour laquelle deux dixièmes des observations (20%) lui sont inférieures ou égales et huit dixièmes des observations (80%) lui sont supérieures ou égales.**

**Le $k^{ème}$ décile $d_k$ : C'est une valeur pour laquelle k dixième des observations lui sont inférieures ou égales et (10 - k) dixième des observations lui sont supérieures ou égales.**

**Le cinquième décile correspond aussi à la médiane et au deuxième quartile.**

**Pour le calcul des déciles, on utilise la même méthode de calcul que pour la médiane et les quartiles. Pour des données groupées en classes, on détermine un décile par interpolation linéaire.**

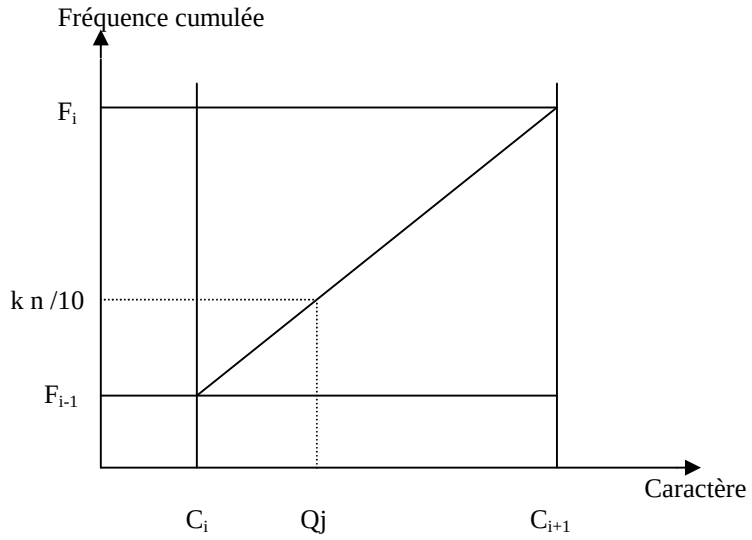
Désignons par :

$[C_i ; C_{i+1}[$ : la classe qui contient le décile ;
n : le nombre total des observations ;
$F_i$ : la fréquence absolue cumulée croissante ;
$n_i$ : la fréquence absolue de la classe qui contient le décile ;

Le décile $d_k$ est compris entre $C_i$ et $C_{i+1}$

$$C_i < d_k < C_{i+1}$$

$$F_{i-1} < \frac{k \times n}{10} < F_i$$




On suppose que la distribution au sein de la classe est régulière.

Ainsi : $$\frac{C_{i+1} - C_i}{F_i - F_{i-1}} = \frac{d_k - C_i}{\frac{k \times n}{10} - F_{i-1}}$$

$$d_k = C_i + \frac{C_{i+1} - C_i}{F_i - F_{i-1}} \times \left(\frac{k \times n}{10} - F_{i-1}\right)$$

**Exemple 24** : La répartition de la surface, en m², de 100 logements est représentée dans le tableau suivant :

| **Surface en m²** | **Nombre de logements** | **Fréquences cumulées croissantes** |
|---|---|---|
| 0 à 20 | 10 | 10 |
| 20 à 40 | 20 | 30 |
| 40 à 60 | 40 | 70 |
| 60 à 100 | 18 | 88 |
| 100 à 160 | 8 | 96 |
| 160 à 260 | 4 | 100 |
| **Total** | **100** | |

**En consultant les fréquences absolues cumulées croissantes, d1, qui correspond à la 10ème observation, se trouve dans la classe 0 à 20 m². $d_9$, qui correspond à la 90ème observation, se trouve dans la classe 100 à 160 m².**

$$d_1 = 0 + 20 \times \frac{\frac{100}{10} - 0}{10} = 20\,\text{m}^2$$

$$d_9 = 100 + 60 \times \frac{\frac{9 \times 100}{10} - 88}{8} = 115\,\text{m}^2$$

- 10 % des logements ont une superficie inférieure ou égale à 20 m².
**- 90 % des logements ont une superficie inférieure ou égale à 115 m².**

**- 80 % des logements ont une superficie comprise entre 20 m² et 115 m².**

## 2.6. EXERCICES D'APPLICATION.

### 2.6.1. Exercice.

Soit la distribution suivante du nombre de pièces dans 300 logements :

| Nombre de pièces | Effectifs |
|---|---|
| 1 | 35 |
| 2 | 51 |
| 3 | 68 |
| 4 | 55 |
| 5 | 49 |
| 6 | 42 |
| **Total** | **300** |

On demande de déterminer pour cette série statistique la moyenne arithmétique, le mode la médiane et les quartiles.

**Solution** : $\bar{x} = \sum_{i=1}^{k} f_i x_i$ = 3,53 pièces ; Mode = 3 pièces ; Me = 2,95 soit 3 pièces

$q_1$ = 1,76 soit 2 pièces ; $q_2$ = Me = 3 pièces et $q_3$ = 4,31 soit 4 pièces

### 2.6.2. Exercice.

On a relevé la recette hebdomadaire en milliers de dirhams de 40 commerces. Les données brutes sont :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 57 | 60 | 52 | 49 | 56 | 46 | 51 | 63 | 49 | 57 |
| 86 | 93 | 77 | 67 | 81 | 70 | 71 | 91 | 67 | 82 |
| 47 | 87 | 92 | 55 | 48 | 90 | 49 | 50 | 58 | 62 |
| 67 | 89 | 69 | 72 | 75 | 48 | 85 | 90 | 83 | 66 |

On demande de déterminer pour cette série statistique la moyenne arithmétique et la médiane :

- A partir de la série brute ;
- A partir de la distribution des fréquences établies à l'exercice 1.3.4.
- Comparer les résultats obtenus.

**Solution** : Série brute : Moyenne = 67 675 DH ; Me = 67 000 DH

Série des fréquences : $\bar{x} = \sum_{i=1}^{k} f_i x_i$ = 68 200 DH ; Me = 66 444 Dh

**Comparaison des résultats : Les résultats obtenus à partir de la distribution des fréquences sont des résultats approximatifs.**

### 2.6.3. Exercice.

Le tableau suivant présente le nombre de femmes en activité selon l'âge de 500 femmes actives :

| **Tranche d'âges** | **Effectif** |
|---|---|
| [15 à 20[ | 14 |
| [20 à 25[ | 70 |
| [25 à 30[ | 100 |
| [30 à 35[ | 65 |
| [35 à 40[ | 69 |
| [40 à 45[ | 56 |
| [45 à 50[ | 63 |
| [50 à 55[ | 61 |
| 55 et plus | 2 |

On demande de déterminer pour cette série statistique la moyenne arithmétique, le mode, la médiane et les quartiles.

Déterminer l'intervalle central qui contient 60 % des femmes actives.

**Solution** : $\bar{x} = \sum_{i=1}^{k} f_i x_i$ = 35,92 ans ; Mo = 27,5 ans ; Me = 35,07 ans

$q_1$ = 27,05 ans ; $q_2$ = Me = 35,07 ans et $q_3$ = 45,08 ans
$d_2$ = 25,8 ans ; $d_8$ = 47,06 ans => 60 % des femmes actives sont âgées entre 25,8 et 47,06 ans.

### 2.6.4. Exercice.

Le tableau suivant donne le niveau de scolarité, en nombre d'années passées à l'école, d'un échantillon de 200 personnes.

| Niveau de scolarité | Effectif |
|---|---|
| [0 ; 6[ | 40 |
| [6 ; 12[ | 80 |
| [12 ; 14[ | 50 |
| [14 ; 16[ | 30 |
| **Total** | **200** |

On demande de déterminer pour la série statistique la moyenne arithmétique, le mode, la médiane et les quartiles.

**Solution** : $\bar{x} = \sum_{i=1}^{k} f_i x_i$ = 9,72 années soit 10 années environ ; Mode = 13 et
Me = 10,5 années
q1 = 6,75 années ; q2 = Me = 10,5 années et q3 = 13,2 années

### 2.6.5. Exercice.

Un organisme chargé de réaliser des enquêtes statistiques gère un réseau de 125 enquêteurs. La direction de cet organisme décide d'étudier la répartition de ses enquêteurs selon le nombre d'enquêtes qu'ils ont réalisées. Les données collectées à ce sujet sont résumées dans le tableau ci-après :

| Nombre d'enquêtes réalisées | Effectifs |
|---|---|
| 5 | 8 |
| 10 | 12 |
| 15 | 35 |
| 20 | 40 |
| 25 | 20 |
| 30 | 10 |

On demande de déterminer pour cette série statistique la moyenne arithmétique, le mode et la médiane.

**Solution** : $\overline{x} = \sum_{i=1}^{k} f_i x_i$ = 18,28 enquêtes ; Mode = 20 enquêtes et Me = 15,94 soit 16 enquêtes environ.

**2.6.6. Exercice.**

Une coopérative laitière fabrique un fromage qui doit contenir, selon les étiquettes, 45 % de matières grasses. Un institut de consommation dont le rôle est de vérifier que la qualité des produits est bien celle qui est affirmée par l'étiquette, fait prélever et analyser un échantillon de 100 fromages. Les résultats de l'analyse sont consignés dans le tableau suivant :

| Taux de matières grasses | Nombre de fromages |
|---|---|
| [41,5 - 42,5[ | 1 |
| [42,5 - 43,5[ | 11 |
| [43,5 - 44,5[ | 24 |
| [44,5 - 45,5[ | 38 |
| [45,5 - 46,5[ | 22 |
| [46,5 - 47,5[ | 3 |
| [47,5 - 48,5[ | 1 |

On demande de déterminer pour la série statistique la moyenne arithmétique, le mode, la médiane, la médiale et les quartiles.

**Solution** : $\overline{x} = \sum_{i=1}^{k} f_i x_i$ = 44,82 % ; Mode = 45 % ; Me = 44,87 % et Ml = 44,89 %

$q_1$ = 44,04 % ; $q_2$ = Me = 44,87 % et $q_3$ = 45,55 %

**2.6.7. Exercice.**

Si le prix d'un article double tous les quatre ans, quel est le taux moyen d'augmentation du prix par an ?

**Solution** : Moyenne géométrique : Taux moyen = $\sqrt[4]{2}$ -1 = 0,189 = 18,9 %

**2.6.8. Exercice.**

Une enquête, abordant la crise de logement, a été réalisée auprès d'un échantillon de 1000 personnes choisies dans quatre régions différentes. Parmi les résultats de cette enquête on a relevé le nombre moyen de personnes par pièce pour chaque région.

| Région | Nombre moyen de personnes par pièce | Nombre d'habitants (en milliers) |
|---|---|---|
| Nord | 2,2 | 5146 |
| Est | 2,6 | 5600 |
| Ouest | 3,1 | 6350 |
| Sud | 3,3 | 7000 |

**Quel est le nombre moyen de personnes par pièce pour l'ensemble des quatre régions ?**

**Solution** : Moyenne harmonique = 2,78 personnes par pièces soit 278 personnes pour 100 pièces.

### 2.6.9. Exercice.

Le coefficient budgétaire de la consommation des ménages en services de santé est passé de 6,9 % en 1990 à 8,5 % en 1995, puis à 9,8 % en 2000, à 10,6 % en 2004 et enfin à 10,9 % en 2005.

a) Calculer les taux annuels moyens de croissance pour les périodes suivantes : (1990 – 1995) ; (1995 – 2000) ; (2000 – 2004) et (2004 – 2005).
b) Déterminer le taux de croissance annuel moyen de 1990 à 2005.
c) Donner une estimation du coefficient budgétaire en 2010 si la tendance relative de la période 2000 - 2005 se maintenait.

**Solution** : a) $t_{1995/1990}$ = 4,26 % par an ; $t_{2000/1995}$ = 2,89 % par an ; $t_{2004/2000}$ = 1,98 % par an et $t_{2005/2004}$ = 2,83 %. b) $t_{2005/1995}$ = 3,10 % par an. c) Coefficient budgétaire estimé en 2010 = 12,1 %.

### 2.6.10. Exercice.

Le prix à la tonne d'une matière première a évolué au cours de la période allant de 2001 à 2005, comme suit :

| **Année** | 2001 | 2002 | 2003 | 2004 | 2005 |
|---|---|---|---|---|---|
| **Prix unitaire** | 310 | 266 | 220 | 200 | 150 |

a) Sachant que chaque année une société achète la même quantité de cette matière première, calculer le coût moyen pour les cinq années.
b) Quel est le coût moyen si la société dépense, chaque année, la même somme : 1 00 000 DH, pour l'achat de cette matière première ?

**Solution** : a) Coût moyen = 229,2 DH/t. b) Coût moyen = 215,54 DH/t.

# CHAPITRE 3
# CARACTERISTIQUES DE DISPERSION

Les paramètres de dispersion d'une série statistique permettent de chiffrer la variation des valeurs observées autour d'un paramètre de position. Les principaux paramètres de dispersion sont : l'écart absolu moyen, la variance, l'écart type, le coefficient de variation et le coefficient de concentration.

Comme pour les caractéristiques centrales, nous ne nous intéressons ici qu'aux séries statistiques relatives à des caractères quantitatifs discrets ou continus, c'est-à-dire à des séries statistiques données sous les formes : $(x_i)$ , $(x_i ; n_i)$ , $(x_i ; f_i)$ ou $\{[C_i ; C_{i+1}[ ; f_i \}$.

## 3.1. L'ECART ABSOLU MOYEN.

L'écart absolu d'une variable $x_i$ par rapport à la moyenne de la série est donné par la formule simple : $\left|x_i - \bar{x}\right|$ où $\bar{x}$ est la moyenne de la série.

L'écart absolu moyen $E_m$ est la moyenne de tous les écarts ainsi définis, il est donné par la formule simple suivante :

$$E_m = \frac{\sum_{i=1}^{n} \left|x_i - \bar{x}\right|}{n}$$

Dans le cas d'une série statistique donnée par un ensemble $(x_i , n_i)$, c'est-à-dire lorsque chaque valeur $x_i$ est répétée $n_i$ fois et qu'il y a k valeurs $x_i$ différentes, l'écart absolu moyen se déduit simplement de la formule précédente :

$$E_m = \frac{\sum_{i=1}^{k} n_i \left|x_i - \bar{x}\right|}{\sum_{i=1}^{k} n_i}$$

En effet : n = $\sum_{i=1}^{k} n_i$ et $\sum_{i=1}^{n} |x_i - \bar{x}| = \sum_{i=1}^{k} n_i |x_i - \bar{x}|$ lorsque chaque valeur $x_i$ est répétée $n_i$ fois dans la série.

De même, dans le cas d'une série statistique donnée par un ensemble ($x_i$ , $f_i$) l'écart absolu moyen se déduit simplement de la formule précédente :

$$E_m = \sum_{i=1}^{k} f_i \left| x_i - \bar{x} \right|$$

En effet lorsque chaque valeur $x_i$ est répété $n_i$ fois dans la série, c'est-à-dire $f_i$ % , on peut écrire :

$$n = \sum_{i=1}^{k} n_i \quad ; \quad f_i = \frac{n_i}{n} \quad \text{et} \quad \sum_{i=1}^{k} f_i = 1$$

$$E_m = \frac{\sum_{i=1}^{k} n_i \left| x_i - \bar{x} \right|}{\sum_{i=1}^{k} n_i} = \sum_{i=1}^{k} \frac{n_i}{n} \left| x_i - \bar{x} \right| = \sum_{i=1}^{k} f_i \left| x_i - \bar{x} \right|$$

Dans le cas d'une série statistique donnée sous la forme de classes $[C_i ; C_{i+1}[$, sachant que pour faire des calculs, on doit remplacer cette série par une série équivalente en remplaçant chaque classe $[C_i ; C_{i+1}[$ par le point central $c_i$ , la formule de l'écart absolu moyen devient :

$$E_m = \frac{\sum_{i=1}^{k} n_i \left| c_i - \bar{x} \right|}{\sum_{i=1}^{k} n_i} = \sum_{i=1}^{k} \frac{n_i}{n} \left| c_i - \bar{x} \right| = \sum_{i=1}^{k} f_i \left| c_i - \bar{x} \right|$$

Avec ci = $\frac{C_i + C_{i+1}}{2}$ centre de la classe : $[C_i ; C_{i+1}[$

**Remarque** : On parle d'écart absolu plutôt que d'écart tout court car l'écart moyen est nul.

**Exemple 1 :** On considère l'ensemble des notes obtenues par les étudiants d'une école, dans une matière ; on a la série statistique suivante donnée sous la forme simple ($x_i$) et pour laquelle on demande de calculer l'écart absolu moyen.

| | | | | |
|---|---|---|---|---|
| 12 | 11 | 13 | 12 | 13 |
| 15 | 13 | 12 | 13 | 11 |
| 13 | 15 | 11 | 11 | 12 |
| 12 | 12 | 10 | 12 | 15 |

La moyenne de cette série est facile à calculer, elle est égale à 12,4. De là nous pouvons calculer les écarts absolus de chaque variable par rapport à la moyenne :

| | | | | |
|---|---|---|---|---|
| 0,4 | 1,4 | 0,6 | 0,4 | 0,6 |
| 2,6 | 0,6 | 0,4 | 0,6 | 1,4 |
| 0,6 | 2,6 | 1,4 | 1,4 | 0,4 |
| 0,4 | 0,4 | 2,4 | 0,4 | 2,6 |

La somme de tous ces écarts absolus est 21,6 et la moyenne est 1,08 qui est l'écart absolu moyen.

**Exemple 2 :** On considère la même série statistique qu'on représente sous la forme ($x_i$ ; $n_i$) pour laquelle on demande de calculer l'écart absolu moyen.

| $\mathbf{x_i}$ | $\mathbf{n_i}$ |
|---|---|
| 10 | 1 |
| 11 | 4 |
| 12 | 7 |
| 13 | 5 |
| 15 | 3 |

La moyenne de la série étant toujours égale à 12,4, le calcul des écarts absolus puis de l'écart absolu moyen peut être facilement fait selon le tableau suivant :

| $\mathbf{x_i}$ | $\mathbf{n_i}$ | $\mathbf{n_i\ x_i}$ | $\left\|x_i - \overline{x}\right\|$ | $\mathbf{n_i}\ \left\|x_i - \overline{x}\right\|$ |
|---|---|---|---|---|
| 10 | 1 | 10 | 2,4 | 2,4 |
| 11 | 4 | 44 | 1,4 | 5,6 |
| 12 | 7 | 84 | 0,4 | 2,8 |
| 13 | 5 | 65 | 0,6 | 3 |
| 15 | 3 | 45 | 2,6 | 7,8 |
| Total | 20 | 248 | - - - | 21,6 |
| **Moyenne** | - - - | **12,4** | - - - | **1,08** |

**Exemple 3 :** On considère la même série statistique qu'on représente sous la forme ($x_i$ ; $f_i$) pour laquelle on demande de calculer l'écart absolu moyen.

| $x_i$ | $n_i$ | $f_i$ |
|---|---|---|
| 10 | 1 | 5% |
| 11 | 4 | 20% |
| 12 | 7 | 35% |
| 13 | 5 | 25% |
| 15 | 3 | 15% |
| **Total** | **20** | **100%** |

La moyenne de la série étant toujours égale à 12,4, le calcul des écarts absolus des variables $x_i$ par rapport à la moyenne puis de l'écart absolu moyen peut être facilement fait selon le tableau suivant :

| $x_i$ | $n_i$ | $f_i$ | $f_i x_i$ | $\lvert x_i - \bar{x} \rvert$ | $f_i \lvert x_i - \bar{x} \rvert$ |
|---|---|---|---|---|---|
| 10 | 1 | 0,05 | 0,5 | 2,4 | 0,12 |
| 11 | 4 | 0,20 | 2,2 | 1,4 | 0,28 |
| 12 | 7 | 0,35 | 4,2 | 0,4 | 0,14 |
| 13 | 5 | 0,25 | 3,25 | 0,6 | 0,15 |
| 15 | 3 | 0,15 | 2,25 | 2,6 | 0,39 |
| Total | 20 | 100% | 12,4 | - - - | 1,08 |
| **Moyenne** | **- - -** | **- - -** | **12,4** | **- - -** | **1,08** |

On remarque que sur ces 3 exemples, pour calculer l'écart absolu moyen, on utilise l'une des 3 formules selon la forme sous laquelle la série statistique est donnée.

**Exemple 4** : On considère un échantillon de 30 personnes pour lesquelles on mesure la taille. On demande de calculer l'écart absolu moyen sachant que les résultats des mesures sont donnés dans le tableau suivant.

| **Tailles $[C_i ; C_{i+1}[$ en m** | **Effectifs $n_i$** |
|---|---|
| [1,50 ; 1,60[ | 2 |
| [1,60 ; 1,70[ | 4 |
| [1,70 ; 1,80[ | 18 |
| [1,80 ; 1,90[ | 5 |
| [1,90 ; 2,00[ | 1 |
| **Total** | **30** |

Après avoir remplacé la série donnée sous la forme ($[C_i ; C_{i+1}[$) en une série équivalente représentée sous la forme ($c_i ; n_i$), avec $c_i = (C_i + C_{i+1}) / 2$, les calculs de l'écart absolu moyen peuvent être résumés dans le tableau synthétique suivant :

| $c_i$ | $n_i$ | $n_i c_i$ | $\left\|x_i - \bar{x}\right\|$ | $n_i \left\|x_i - \bar{x}\right\|$ |
|---|---|---|---|---|
| 1,55 | 2 | 3,10 | 0,20 | 0,40 |
| 1,65 | 4 | 6,60 | 0,10 | 0,40 |
| 1,75 | 18 | 31,50 | 0,00 | 0,00 |
| 1,85 | 5 | 9,25 | 0,10 | 0,50 |
| 1,95 | 1 | 1,95 | 0,20 | 0,20 |
| Total | 30 | 52,40 | - - - | 1,50 |
| **Total / n** | | **1,75** | | **0,050** |

La moyenne de la série est 1,75 m et l'écart absolu moyen est 0,05 m.

### 3.2. LA VARIANCE.

La variance V(x), notée aussi $S^2$, est la moyenne des carrés des écarts par rapport à la moyenne ; elle est donnée, par définition, par la formule simple suivante :

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

Dans le cas d'une série statistique donnée par un ensemble $(x_i , n_i)$, c'est-à-dire lorsque chaque valeur $x_i$ est répétée $n_i$ fois et qu'il y a k valeurs $x_i$ différentes, la variance se déduit simplement de la formule précédente :

$$S^2 = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x})^2}{\sum_{i=1}^{k} n_i}$$

En effet : $n = \sum_{i=1}^{k} n_i$ et $\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{k} n_i (x_i - \bar{x})^2$ lorsque chaque valeur $x_i$ est répétée $n_i$ fois dans la série.

De même dans le cas d'une série statistique donnée par un ensemble $(x_i , f_i)$ la variance se déduit simplement de la formule précédente :

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

En effet lorsque chaque valeur $x_i$ est répété $n_i$ fois dans la série, c'est-à-dire $f_i$ % , on peut écrire :

$$n = \sum_{i=1}^{k} n_i \quad ; \quad f_i = \frac{n_i}{n} \quad \text{et} \quad \sum_{i=1}^{k} f_i = 1$$

$$S^2 = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x})^2}{\sum_{i=1}^{k} n_i} = \sum_{i=1}^{k} \frac{n_i}{n} (x_i - \bar{x})^2 = \sum_{i=1}^{k} f_i (x_i - \bar{x})^2$$

Dans le cas d'une série statistique donnée sous la forme de classes $[C_i ; C_{i+1}[$, sachant que pour faire des calculs, on doit remplacer cette série par une série équivalente en remplaçant chaque classe $[C_i ; C_{i+1}[$ par le point central $c_i$ , la formule de la variance devient :

$$S^2 = \frac{\sum_{i=1}^{k} n_i (c_i - \bar{x})^2}{\sum_{i=1}^{k} n_i} = \sum_{i=1}^{k} \frac{n_i}{n} (c_i - \bar{x})^2 = \sum_{i=1}^{k} f_i (c_i - \bar{x})^2$$

Avec ci = $\frac{C_i + C_{i+1}}{2}$ centre de la classe : $[C_i ; C_{i+1}[$

Formule développée de la variance **: elle est donnée, selon la forme de la série statistique :**

* Cas d'une série statistique de n observations $x_i$ distinctes :

$$S^2 = \frac{\sum_{i=1}^{n} x_i^{\,2}}{n} - \bar{x}^{\,2} = \overline{x^2} - \bar{x}^2$$

* Cas d'une série statistique de k observations $x_i$ distinctes dont chacune est répétées $n_i$ fois :

$$S^2 = \frac{\sum_{i=1}^{k} n_i x_i^{\,2}}{\sum_{i=1}^{k} n_i} - \bar{x}^{\,2} = \overline{x^2} - \bar{x}^2$$

* Cas d'une série statistique de k observations $x_i$ distinctes dont chacune est présentes $f_i$ fois (en %) :

$$S^2 = \sum_{i=1}^{k} f_i x_i^{\,2} - \bar{x}^{\,2} = \overline{x^2} - \bar{x}^2$$

* Cas d'une série statistique donnée sous la forme de k classes $[C_i ; C_{i+1}[$ ayant chacune un effectif $n_i$ ou une fréquence $f_i$ :

$$S^2 = \sum_{i=1}^{k} f_i c_i^{\,2} - \bar{x}^{\,2} = \overline{x^2} - \bar{x}^{\,2}$$

Avec ci = $\dfrac{C_i + C_{i+1}}{2}$ centre de la classe : $[C_i ; C_{i+1}[$

Toutes ces formules peuvent être écrites, comme nous l'avons bien montré, sous la forme simple et condensée :

$$S^2 = \overline{x^2} - \bar{x}^{\,2}$$

Transformation linéaire

Si Y = ax + b avec a et b deux constantes quelconques alors la variance de y est : $S_y^{\,2} = a^2 \times S_x^{\,2}$.

**Exemple 5 :** Le tableau suivant présente les prix en DH de 100 ordinateurs portables achetés dans différents points de vente :

| **Prix** | **Nombre d'ordinateurs** |
|---|---|
| [10000 ; 11000[ | 9 |
| [11000 ; 12000[ | 10 |
| [12000 ; 13000[ | 10 |
| [13000 ; 14000[ | 14 |
| [14000 ; 15000[ | 16 |
| [15000 ; 16000[ | 14 |
| [16000 ; 17000[ | 12 |
| [17000 ; 18000[ | 15 |
| **Total** | **100** |

Pour calculer la variance des prix des ordinateurs, on peut utiliser la propriété de la transformation linéaire dans le but de simplifier les calculs.

On effectue un changement de variable, c'est-à-dire, on remplace la variable prix par une autre variable y de telle sorte que le prix soit une transformation linéaire de y.

p = a y + b Donc : $y = \dfrac{p - b}{a}$

Il faut choisir les constantes a et b qui donnent des valeurs très simples de y. On choisit la constante b parmi les valeurs de p, de préférence une valeur du milieu, pour avoir une valeur nulle de y au milieu. On choisit la constante a comme étant le plus grand diviseur commun des valeurs de (p - b) (le plus souvent a est l'amplitude constante des classes) pour avoir des valeurs entières de y.

Pour notre exemple, on choisit :

b = 13500 et a = 1000

$$Y = \frac{p - 13500}{1000}$$

Les valeurs de y deviennent très simples, on peut alors calculer facilement la moyenne et la variance de y.

| **Prix** | **Nombre d'ordinateurs ($n_i$)** | **Point central ($c_i$)** | **$y_i$** | **$n_i y_i$** | **$n_i y_i^2$** |
|---|---|---|---|---|---|
| [10000 ; 11000[ | 9 | 10500 | -3 | -27 | 81 |
| [11000 ; 12000[ | 10 | 11500 | -2 | -20 | 40 |
| [12000 ; 13000[ | 10 | 12500 | -1 | -10 | 10 |
| [13000 ; 14000[ | 14 | 13500 | 0 | 0 | 0 |
| [14000 ; 15000[ | 16 | 14500 | 1 | 16 | 16 |
| [15000 ; 16000[ | 14 | 15500 | 2 | 28 | 56 |
| [16000 ; 17000[ | 12 | 16500 | 3 | 36 | 108 |
| [17000 ; 18000[ | 15 | 17500 | 4 | 60 | 240 |
| **Total** | **100** | | | **83** | **551** |

$$\bar{y} = \frac{\sum_{i=1}^{8} n_i y_i}{\sum_{i=1}^{8} n_i} = \frac{83}{100} = 0{,}83$$

$$S_y^2 = \frac{\sum_{i=1}^{8} n_i y_i^2}{\sum_{i=1}^{8} n_i} - \bar{y}^2 = \frac{551}{100} - 0{,}83^2 = 4{,}82$$

On calcule facilement la moyenne et la variance grâce aux formules de la transformation linéaire :

$$\bar{p} = 1000 \times \bar{y} + 13500 = 1000 \times 0{,}83 + 13500 = 14330 \text{ DH}$$
$$S_p^{\,2} = 1000^2 \times S_y^{\,2} = 1000^2 \times 4{,}82 = 4820000$$

### 3.3. L'ECART TYPE.

L'écart type S est la racine carrée de la variance, il est donné, par définition, par les formules suivantes :

* Cas d'une série statistique de n observations $x_i$ distinctes :

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

* Cas d'une série statistique de k observations $x_i$ distinctes dont chacune est répétées $n_i$ fois :

$$S = \sqrt{\frac{\sum_{i=1}^{k} n_i (x_i - \bar{x})^2}{\sum_{i=1}^{k} n_i}}$$

* Cas d'une série statistique de k observations $x_i$ distinctes dont chacune est présentes $f_i$ fois (en %) :

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

* Cas d'une série statistique donnée sous la forme de k classes $[C_i ; C_{i+1}[$ ayant chacune un effectif $n_i$ ou une fréquence relative $f_i$ :

$$S = \sqrt{\sum_{i=1}^{k} f_i (c_i - \bar{x})^2}$$

avec ci = $\frac{C_i + C_{i+1}}{2}$ centre de la classe : $[C_i ; C_{i+1}[$

D'une façon générale, nous pouvons utiliser, quelque soit la forme sous laquelle est donnée la série statistique, la formule condensée simple :

$$S = \sqrt{\overline{x^2} - \overline{x}^2}$$

### 3.4. LE COEFFICIENT DE VARIATION.

Le coefficient de variation CV ou coefficient de dispersion est le rapport de l'écart type à la moyenne. Il est exprimé sous la forme d'un pourcentage.

$$CV = \frac{S}{\overline{x}} \times 100 \text{ en \%}$$

Le coefficient de variation est indépendant des unités choisies, il est utile pour comparer des distributions qui ont des unités différentes.

**Exemple 6** : On considère toujours la même série, des 3 premiers exemples de ce chapitre. Soit l'ensemble des notes obtenues par les étudiants d'une école, dans une matière ; on a la série statistique suivante donnée sous la forme simple ($x_i$ ; $n_i$) et pour laquelle on demande de calculer la variance, l'écart type et le coefficient de dispersion.

| $x_i$ | $n_i$ | $f_i$ | $f_i x_i$ | $x_i^2$ | $f_i x_i^2$ |
|---|---|---|---|---|---|
| 10 | 1 | 0,05 | 0,5 | 100 | 5 |
| 11 | 4 | 0,20 | 2,2 | 121 | 24,2 |
| 12 | 7 | 0,35 | 4,2 | 144 | 50,4 |
| 13 | 5 | 0,25 | 3,25 | 169 | 42,25 |
| 15 | 3 | 0,15 | 2,25 | 225 | 33,75 |
| Total | 20 | 100% | 12,4 | - - - | 155,6 |
| **Moyenne** | **- - -** | **- - -** | **12,4** | **- - -** | **155,6** |

Ainsi la variance de le série est $S^2 = 155,6 - 12,4^2 = 1,84$

L'écart type est égal à S = 1,356

Le coefficient de dispersion CV = $\frac{1,356}{12,4} \times 100 = 10,94\%$ ce qui dénote d'une légère dispersion de la série autour de sa moyenne.

**Exemple 7** : On considère la série de l'exemple 4 relative au poids de 30 personnes et on demande de calculer la variance, l'écart type et le coefficient de variation de cette série qui est donnée par le tableau suivant :

| **Tailles $[C_i ; C_{i+1}[$ en m** | **Effectifs $n_i$** |
|---|---|
| [1,50 ; 1,60[ | 2 |
| [1,60 ; 1,70[ | 4 |
| [1,70 ; 1,80[ | 18 |
| [1,80 ; 1,90[ | 5 |
| [1,90 ; 2,00[ | 1 |
| **Total** | **30** |

Après avoir remplacé la série donnée sous la forme ($[C_i ; C_{i+1}[$) en une série équivalente représentée sous la forme ($c_i ; n_i$)

avec $c_i = (C_i + C_{i+1}) / 2$, les calculs de la variance, de l'écart type et du coefficient de variation peuvent être résumés dans le tableau synthétique suivant :

| $c_i$ | $n_i$ | $n_i c_i$ | $n_i c_i^2$ |
|---|---|---|---|
| 1,55 | 2 | 3,10 | 4,8050 |
| 1,65 | 4 | 6,60 | 10,8900 |
| 1,75 | 18 | 31,50 | 55,1250 |
| 1,85 | 5 | 9,25 | 17,1125 |
| 1,95 | 1 | 1,95 | 3,8025 |
| Total | 30 | 52,40 | 91,7350 |
| **Total / n** | | **1,7467** | **3,0578** |

La moyenne de la série est égale à 52,40 / 30 = 1,75 m

La variance est V = 3,0578 – $1.7467^2$ = 0,0064 $m^2$

L'écart type est égal à $S = \sqrt{3,0578 - 1,7467^2}$ = 0,08 m

Le coefficient de dispersion est égal à 0,08 / 1.75 = 4,57% ce qui dénote d'une très faible dispersion de la série autour de sa moyenne.

**Exemple 8** : Allal est un marchand de journaux, il comptabilise le nombre de journaux qu'il vend, par jour, en un mois et dresse ses résultats dans le tableau suivant.

| | | | | | |
|---|---|---|---|---|---|
| 125 | 118 | 127 | 110 | 107 | 125 |
| 118 | 110 | 107 | 125 | 127 | 127 |
| 107 | 125 | 118 | 107 | 107 | 118 |
| 107 | 118 | 125 | 127 | 125 | 107 |
| 110 | 125 | 127 | 127 | 125 | 125 |

Calculer la variance, l'écart type et le coefficient de dispersion de cette série.

Commençons d'abord par représenter cette série, donnée sous la forme (xi) en une série équivalente sous la forme (xi ; ni) après avoir compté combien de fois chaque valeur xi est répétée.

On obtient la série équivalente suivante :

| Nombre de journaux vendus | Nombre de fois |
|---|---|
| 107 | 7 |
| 110 | 3 |
| 118 | 5 |
| 125 | 9 |
| 127 | 6 |
| **Total** | **30** |

Les calculs de la moyenne, de la variance, de l'écart type et du coefficient de variation de la série peuvent être résumés dans le tableau suivant :

| xi | ni | ni xi | ni $xi^2$ |
|---|---|---|---|
| 107 | 7 | 749 | 80143 |
| 110 | 3 | 330 | 36300 |
| 118 | 5 | 590 | 69620 |
| 125 | 9 | 1125 | 140625 |
| 127 | 6 | 762 | 96774 |
| Total | 30 | 3556 | 423462 |
| **Total/n** | | **118,53** | **14115,40** |

La moyenne de la série se situe entre 118 et 119 journaux vendus par jour.

La variance est égale à $V = 14115,40 - 118,53^2 = 66,04$

L'écart type est égal à $S = \sqrt{66,04} = 8,13$

Le coefficient de dispersion est égal à 8,13 / 118,53 = 6,86% ce qui dénote d'une légère dispersion de la série.

**Exemple 9** : Les salaires versés, par une entreprise, à ses 130 salariés sont répartis comme suit :

| Tranches de salaire | Nombre de salariés hommes | Nombre de salariés femmes | Total |
|---|---|---|---|
| [1000 ; 2000[ | 8 | 4 | 12 |
| [2000 ; 3000[ | 12 | 9 | 21 |
| [3000 ; 4000[ | 10 | 6 | 16 |
| [4000 ; 5000[ | 14 | 10 | 24 |
| [5000 ; 6000[ | 11 | 8 | 19 |
| [6000 ; 7000[ | 8 | 6 | 14 |
| [7000 ; 8000[ | 7 | 5 | 12 |
| [8000 ; 10000[ | 5 | 2 | 7 |
| [10000 ; 15000[ | 3 | 1 | 4 |
| [15000 ; 20000[ | 1 | 0 | 1 |
| **Total** | **79** | **51** | **130** |

On demande de calculer la moyenne, la variance, l'écart type et le coefficient de variation pour l'ensemble des salariés et pour chaque sexe.

## Calculs pour l'ensemble des salariés.

## Calcul de la moyenne, la variance, l'écart type et le coefficient de variation.

| Tranches de salaire | $c_i$ | $n_i$ | $n_i c_i$ | $c_i^2$ | $n_i c_i^2$ |
|---|---|---|---|---|---|
| [1000 ; 2000[ | 1500 | 12 | 18000 | 2250000 | 27000000 |
| [2000 ; 3000[ | 2500 | 21 | 52500 | 6250000 | 131250000 |
| [3000 ; 4000[ | 3500 | 16 | 56000 | 12250000 | 196000000 |
| [4000 ; 5000[ | 4500 | 24 | 108000 | 20250000 | 486000000 |
| [5000 ; 6000[ | 5500 | 19 | 104500 | 30250000 | 574750000 |
| [6000 ; 7000[ | 6500 | 14 | 91000 | 42250000 | 591500000 |
| [7000 ; 8000[ | 7500 | 12 | 90000 | 56250000 | 675000000 |
| [8000 ; 10000[ | 9000 | 7 | 63000 | 81000000 | 567000000 |
| [10000 ; 15000[ | 12500 | 4 | 50000 | 156250000 | 625000000 |
| [15000 ; 20000[ | 17500 | 1 | 17500 | 306250000 | 306250000 |
| **Total** | **-** | **130** | **650500** | **-** | **4179750000** |

La moyenne est : $\bar{x} = \frac{650500}{130}$ = 5003,85 DH.

La variance est : $S^2 = \frac{4179750000}{130} - 5003{,}85^2 = 7113446{,}75$

L'écart type est : $S = \sqrt{7113446{,}75}$ = 2667,10 DH

Le coefficient de variation est : CV = $\frac{2667,10}{5003,85}$ x 100 = 53,3 %

## Calculs pour les salariés hommes.

## Calcul de la moyenne, la variance, l'écart type et le coefficient de variation

| Tranches de salaire | $c_i$ | $n_i$ | $n_i c_i$ | $c_i^2$ | $n_i c_i^2$ |
|---|---|---|---|---|---|
| [1000 ; 2000[ | 1500 | 8 | 12000 | 2250000 | 18000000 |
| [2000 ; 3000[ | 2500 | 12 | 30000 | 6250000 | 75000000 |
| [3000 ; 4000[ | 3500 | 10 | 35000 | 12250000 | 122500000 |
| [4000 ; 5000[ | 4500 | 14 | 63000 | 20250000 | 283500000 |
| [5000 ; 6000[ | 5500 | 11 | 60500 | 30250000 | 332750000 |
| [6000 ; 7000[ | 6500 | 8 | 52000 | 42250000 | 338000000 |
| [7000 ; 8000[ | 7500 | 7 | 52500 | 56250000 | 393750000 |
| [8000 ; 10000[ | 9000 | 5 | 45000 | 81000000 | 405000000 |
| [10000 ; 15000[ | 12500 | 3 | 37500 | 156250000 | 468750000 |
| [15000 ; 20000[ | 17500 | 1 | 17500 | 306250000 | 306250000 |
| **Total** | **-** | **79** | **405000** | | **2743500000** |

La moyenne est : $\bar{x} = \frac{405000}{79}$ = 5126,58 DH.

La variance est : $S^2 = \frac{2743500000}{79} - 5126,58^2 = 8446002,24$

L'écart type est : $S = \sqrt{8446002,24}$ = 2906,20 DH

Le coefficient de variation est : CV = $\frac{2906,20}{5126,58}$ x 100 = 56,7 %

## Calculs pour les salariés femmes.

## Calcul de la moyenne, la variance, l'écart type et le coefficient de variation

| Tranches de salaire | $c_i$ | $n_i$ | $n_i c_i$ | $c_i^2$ | $n_i c_i^2$ |
|---|---|---|---|---|---|
| [1000 ; 2000[ | 1500 | 4 | 6000 | 2250000 | 9000000 |
| [2000 ; 3000[ | 2500 | 9 | 22500 | 6250000 | 56250000 |
| [3000 ; 4000[ | 3500 | 6 | 21000 | 12250000 | 73500000 |
| [4000 ; 5000[ | 4500 | 10 | 45000 | 20250000 | 202500000 |
| [5000 ; 6000[ | 5500 | 8 | 44000 | 30250000 | 242000000 |
| [6000 ; 7000[ | 6500 | 6 | 39000 | 42250000 | 253500000 |

| | | | | | |
|---|---|---|---|---|---|
| [7000 ; 8000[ | 7500 | 5 | 37500 | 56250000 | 281250000 |
| [8000 ; 10000[ | 9000 | 2 | 18000 | 81000000 | 162000000 |
| [10000 ; 15000[ | 12500 | 1 | 12500 | 156250000 | 156250000 |
| [15000 ; 20000[ | 17500 | 0 | 0 | 306250000 | 0 |
| **Total** | **-** | **51** | **245500** | | **1436250000** |

La moyenne est : $\bar{x} = \frac{244500}{51}$ = 4794,12 DH.

La variance est : $S^2 = \frac{1436250000}{51} - 4794{,}12^2 = 5178200{,}69$

L'écart type est : $S = \sqrt{5178200{,}69}$ = 2275,57 DH

Le coefficient de variation est : $CV = \frac{2275{,}57}{4794{,}12} \times 100 = 47{,}47\ \%$

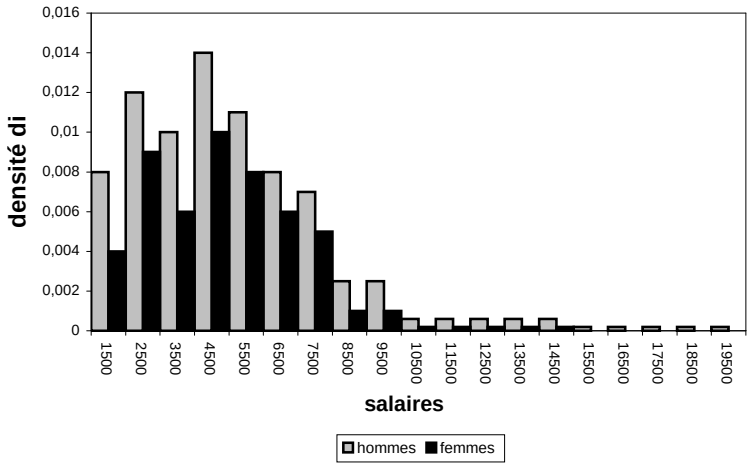
Récapitulatif des résultats

| **Salariés** | **Moyenne** | **Ecart type** | **Coefficient de variation** |
|---|---|---|---|
| **Hommes** | 5126,58 | 2906,20 | 56,7 % |
| **Femmes** | 4794,12 | 2275,57 | 47,47 % |
| **Ensemble** | 5003,85 | 2667,10 | 53,3 % |

En moyenne, un salarié de l'entreprise perçoit un salaire de 5003,85. La répartition des salaires est caractérisée par une forte dispersion.

La répartition des salaires varie selon le sexe, en effet, un salarié homme touche, en moyenne, plus qu'un salarié femme, (respectivement 5126,58 DH et 4794,12 DH). Les salaires sont plus dispersés chez les hommes que chez les femmes, 56,7 % pour les premiers et 47,47 % pour les seconds.

Répartition des salaires selon le sexe




### 3.5. INDICE DE CONCENTRATION.

L'indice de concentration est donné par la formule :

$$\text{Indice de concentration} = \frac{\text{Médiale - Médiane}}{\text{Etendu}} \times 100$$

**Exemple 10** : On considère l'ensemble des buts marqués par une équipe durant les 30 matchs du championnat de football, on a la série statistique suivante donnée sous la forme simple ($x_i$) et pour laquelle on demande de calculer la moyenne, la variance, l'écart type, le coefficient de variation et le coefficient de concentration.

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

La somme et la somme des carrées de cette série sont faciles à calculer :

$$\sum_{i=1}^{30} x_i = 55 \qquad \sum_{i=1}^{30} x_{i^2} = 155$$

La moyenne de cette série est : $\bar{x} = \frac{55}{30} = 1,83$ but par match.

La variance de cette série est : $S^2 = \frac{155}{30} - 1,83^2 = 1,82$

L'écart type de cette série est : $S = \sqrt{1,82} = 1,35$ but

Le coefficient de variation est : $CV = \frac{1,35}{1,83} \times 100 = 73,77\ \%$

Pour déterminer la médiane de cette série, on considère le nombre d'observations, 30 qui est pair, la médiane est comprise entre l'observation de rang 15 et l'observation de rang 16. On prend comme valeur de la médiane la moyenne arithmétique simple des deux observations. La série classée par ordre croissant est :

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

$x_{15} < Me\ x_{16}$

$2 \leq Me \leq 2$ ce qui donne : Me = 2 buts

Pour déterminer la médiale de cette série on cumule les valeurs de la série classée jusqu'à arriver à la moitié de la somme totale, c'est à dire 22,5. Elle correspond à la valeur 2.

La médiane est égale à la médiale, le coefficient de concentration de cette série est donc nul.

**Exemple 11 :** On considère la même série statistique de l'exemple 10 et on la représente sous la forme ($x_i$ ; $n_i$). On demande de calculer la moyenne, la variance, l'écart type, le coefficient de variation et le coefficient de concentration de cette série.

| $\mathbf{x_i}$ | $\mathbf{n_i}$ |
|---|---|
| 0 | 5 |
| 1 | 8 |
| 2 | 9 |
| 3 | 5 |
| 4 | 1 |
| 5 | 2 |
| **Total** | **30** |

La somme et la somme des carrés de cette série sont faciles à calculer :

$\sum_{i=1}^{6} n_i x_i = 55$ $\sum_{i=1}^{6} n_i x_i^2 = 155$

La moyenne de cette série est : $\bar{x} = \frac{55}{30} = 1,83$ but par match.

La variance de cette série est : $S^2 = \frac{155}{30} - 1,83^2 = 1,82$

L'écart type de cette série est : $S = \sqrt{1,82} = 1,35$ but

Le coefficient de variation de cette série est :

$$CV = \frac{1,35}{1,83} \times 100 = 73,77\ \%$$

Calcul de l'indice de concentration :

| $x_i$ | $n_i$ | $F_i$ | $n_i\ x_i$ | $n_i\ x_i$ cumulé |
|---|---|---|---|---|
| 0 | 5 | 5 | 0 | 0 |
| 1 | 8 | 13 | 8 | 8 |
| 2 | 9 | 22 | 18 | 26 |
| 3 | 5 | 27 | 15 | 41 |
| 4 | 1 | 28 | 4 | 45 |
| 5 | 2 | 30 | 10 | 55 |
| **Total** | **30** | **- - -** | **55** | **- - -** |

Pour déterminer la médiane de cette série on considère le nombre d'observations, 30 qui est pair ; la médiane est comprise entre l'observation de rang 15 et l'observation de rang 16. On prend comme valeur de la médiane la moyenne arithmétique simple des deux observations.

$x_{15} < Me\ x_{16}$

$2 \leq Me \leq 2$ ce qui donne : Me = 2 buts

La médiale de cette série est la moitié de la somme totale, c'est à dire 22,5 qui correspond à la valeur 2

La médiane est égale à la médiale, le coefficient de concentration de cette série est donc nul.

**Exemple 12** : Une coopérative laitière fabrique un fromage qui doit contenir, selon les étiquettes, 45 % de matières grasses. Un institut de consommation dont le rôle est de vérifier que la qualité des produits est bien celle qui est affichée par l'étiquette, fait prélever et analyser

un échantillon de 120 fromages. Les résultats de l'analyse sont consignés dans le tableau suivant :

| Taux de matières grasses | Nombre de fromages |
|---|---|
| [41,5 - 42,5[ | 10 |
| [42,5 - 43,5[ | 11 |
| [43,5 - 44,5[ | 24 |
| [44,5 - 45,5[ | 38 |
| [45,5 - 46,5[ | 22 |
| [46,5 - 47,5[ | 4 |
| [47,5 - 48,5[ | 11 |

On demande de calculer la moyenne, la variance, l'écart type, le coefficient de variation et le coefficient de concentration de cet échantillon.

| $c_i$ | $n_i$ | $f_i$ | $f_i c_i$ | $c_i^2$ | $f_i c_i^2$ |
|---|---|---|---|---|---|
| 42 | 10 | 8,33% | 3,4986 | 1764 | 146,9412 |
| 43 | 11 | 9,17% | 3,9431 | 1849 | 169,5533 |
| 44 | 24 | 20% | 8,8 | 1936 | 387,2 |
| 45 | 38 | 31,67% | 14,2515 | 2025 | 641,3175 |
| 46 | 22 | 18,33% | 8,4318 | 2116 | 387,8628 |
| 47 | 4 | 3,33% | 1,551 | 2209 | 73,5597 |
| 48 | 11 | 9,17% | 4,4016 | 2304 | 211,2768 |
| Total | 120 | 100% | 44,8776 | | 2017,7113 |
| **Moyenne** | | | **44,8776** | | **2017,7113** |

La moyenne est : 44,8776 % de matières grasses.

La variance est : $S^2 = 2017{,}7113 - 44{,}8776^2 = 3{,}712$

L'écart type est : $S = \sqrt{3{,}712} = 1{,}93$ % de matières grasses.

Le coefficient de variation est : $CV = \frac{1{,}93}{44{,}8776} \times 100 = 4{,}3$ %

Les calculs de l'indice de concentration peuvent être résumés dans le tableau suivant :

| Taux de matières grasses | $n_i$ | $f_i$ | $F_i$ | $f_i c_i$ | $f_i c_i$ cumulé |
|---|---|---|---|---|---|
| [41,5 – 42,5[ | 10 | 8,33% | 8,33% | 3,4986 | 3,4986 |
| [42,5 – 43,5[ | 11 | 9,17% | 17,5% | 3,9431 | 7,4417 |
| [43,5 – 44,5[ | 24 | 20% | 37,5% | 8,8 | 16,2417 |

| | | | | | |
|---|---|---|---|---|---|
| [44,5 – 45,5[ | 38 | 31,67% | 69,17% | 14,2515 | 30,4932 |
| [45,5 – 46,5[ | 22 | 18,33% | 87,5% | 8,4318 | 38,925 |
| [46,5 – 47,5[ | 4 | 3,33% | 90,83% | 1,551 | 40,476 |
| [47,5 – 48,5[ | 11 | 9,17% | 100% | 4,4016 | 44,8776 |
| **Total** | **120** | **100%** | **- - -** | **44,8776** | **- - -** |

En consultant les fréquences cumulées croissantes, la classe médiane qui correspond à 50%, est la classe [44,5 – 45,5[. La médiane est donc :

$$44,5 < Me < 45,5$$
$$37,5 < 50 < 69,17$$

Un calcul simple d'extrapolation donne pour la médiane :

$$\frac{45,5-44,5}{69,17-37,5} = \frac{Me-44,5}{50-37,5}$$

$$Me = 44,5 + \frac{1}{31,67} \text{ x } 12,5 = 44,89$$

En consultant les sommes cumulées croissantes, la moitié de la somme totale (soit 22,4388) se trouve dans la classe [44,5 – 45,5[. La médiale est donc :

$$44,5 < Ml < 45,5$$
$$16,2417 < 22,4388 < 30,4932$$

Un calcul simple d'extrapolation donne pour la médiale :

$$\frac{45,5-44,5}{30,4932-16,2417} = \frac{Ml-44,5}{22,4388-16,2417}$$

$$Ml = 44,5 + \frac{1}{14,2515} \text{ x } 6,1971 = 44,93$$

L'étendu de la série est : 48,5 – 41,5 = 7

L'indice de concentration est donné par la formule :

$$\text{Indice de concentration} = \frac{\text{Médiale - Médiane}}{\text{Etendu}} \times 100 = 0,57\%$$

**Exemple 13** : On reprend les données de l'exemple 8 relatives aux ventes de journaux faites par Allal, pour calculer l'indice de concentration de la série qui est donnée par le tableau suivant :

| | | | | | |
|---|---|---|---|---|---|
| 125 | 118 | 127 | 110 | 107 | 125 |

| | | | | | |
|---|---|---|---|---|---|
| 118 | 110 | 107 | 125 | 127 | 127 |
| 107 | 125 | 118 | 107 | 107 | 118 |
| 107 | 118 | 125 | 127 | 125 | 107 |
| 110 | 125 | 127 | 127 | 125 | 125 |

Calculons l'indice de concentration de cette série.

Rappelons les résultats que nous avons déjà trouvés lors de l'étude de l'exemple 8, à savoir :

La moyenne est : 118,53.

L'écart type est : 8,13.

Le coefficient de variation est : 6,86%.

Pour déterminer la médiane, s'agissant d'une série donnée sous la forme brute ($x_i$), il nous faudra la classer par valeurs croissantes des ventes. On obtient le tableau suivant :

| | | | | | |
|---|---|---|---|---|---|
| 107 | 107 | 107 | 107 | 107 | 107 |
| 107 | 110 | 110 | 110 | 118 | 118 |
| 118 | 118 | 118 | 125 | 125 | 125 |
| 125 | 125 | 125 | 125 | 125 | 125 |
| 127 | 127 | 127 | 127 | 127 | 127 |

Il y a 30 observations, la médiane est la moyenne arithmétique des 15è et 16è observations, soit :

Me = (118 + 125) / 2 = 121,5

Pour déterminer la médiane, on doit consulter les fréquences cumulées croissantes, la valeur médiane est exactement 118.

Pour déterminer la médiale, nous devons réorganiser la série sous la forme ($x_i$ ; $n_i$).

| **Xi** | **ni** | **$f_i$** | **$F_i$** | **Somme $n_i$ $x_i$** | **Sommes cumulées croissantes** |
|---|---|---|---|---|---|
| 107 | 7 | 0,2333 | 0,2333 | 749 | 749 |
| 110 | 3 | 0,1000 | 0,3333 | 330 | 1079 |
| 118 | 5 | 0,1667 | 0,5000 | 590 | 1669 |
| 125 | 9 | 0,3000 | 0,8000 | 1125 | 2794 |
| 127 | 6 | 0,2000 | 1,0000 | 762 | 3556 |

En consultant les sommes cumulées croissantes, la moitié de la somme totale (soit 1778) se trouve entre les valeurs 118 et 125. La médiale est donc :

$$118 < Ml < 125$$
$$1669 < 1778 < 2794$$

Un calcul simple d'extrapolation donne pour la médiale :

$$\frac{125-118}{2794-1669} = \frac{Ml-118}{1778-1669}$$

La médiale est : Ml = 118 + 109 x 0,006 222 = 118,68

L'étendu de la série est : 127 – 107 = 20

L'indice de concentration est donné par la formule :

$$\text{Indice de concentration} = \frac{\text{Médiale - Médiane}}{\text{Etendu}} \times 100 = 3{,}4\%$$

**Exemple 14** : On reprend les données de l'exemple 9 et on demande de calculer le coefficient de concentration des séries statistiques relatives aux salaires des hommes, des femmes et de l'ensemble du personnel. Conclure.

**a) Calculs pour l'ensemble des salariés**

| **Tranches de salaire** | **ci** | **ni** | **ni cumulé** | **ni ci** | **ni ci cumulé** |
|---|---|---|---|---|---|
| [1000 ; 2000[ | 1500 | 12 | 12 | 18000 | 18000 |
| [2000 ; 3000[ | 2500 | 21 | 33 | 52500 | 70500 |
| [3000 ; 4000[ | 3500 | 16 | 49 | 56000 | 126500 |
| [4000 ; 5000[ | 4500 | 24 | 73 | 108000 | 234500 |
| [5000 ; 6000[ | 5500 | 19 | 92 | 104500 | 339000 |
| [6000 ; 7000[ | 6500 | 14 | 106 | 91000 | 430000 |
| [7000 ; 8000[ | 7500 | 12 | 118 | 90000 | 520000 |
| [8000 ; 10000[ | 9000 | 7 | 125 | 63000 | 583000 |
| [10000 ; 15000[ | 12500 | 4 | 129 | 50000 | 633000 |
| [15000 ; 20000[ | 17500 | 1 | 130 | 17500 | 650500 |
| **Total** | **Total** | **130** | | **650500** | |

La médiane correspond à la 65ème observation (130/2). En consultant les fréquences cumulées croissantes, la classe médiane est la classe [4000 ; 5000[. La médiane est donc :

$$4000 < Me < 5000$$
$$49 < 65 < 73$$

$$\frac{5000 - 4000}{73 - 49} = \frac{Me - 4000}{65 - 49}$$

$$Me = 4000 + \frac{1000}{24} \times 16 = 4666,67 \text{ DH}$$

En consultant les sommes cumulées croissantes, la moitié de la somme totale (325250 DH) se trouve dans la classe [5000 ; 6000[. La médiale est donc :

$$5000 < Ml < 6000$$
$$234500 < 325250 < 339000$$

$$\frac{6000 - 5000}{339000 - 234500} = \frac{Ml - 5000}{325250 - 234500}$$

$$Ml = 5000 + \frac{1000}{104500} \times 90750 = 5868,42 \text{ DH}$$

L'étendu de la série est : 20000 – 1000 = 19000

L'indice de concentration est donné par la formule :

$$\text{Indice de concentration} = \frac{\text{Médiale - Médiane}}{\text{Etendu}} \times 100$$

$$\text{Indice de concentration} = \frac{5868,42 - 4666,67}{19000} \times 100 = 6,33\ \%$$

**b) Calculs pour les salariés hommes**

| Tranches de salaire | ci | ni | ni cumulé | ni ci | ni ci cumulé |
|---|---|---|---|---|---|
| [1000 – 2000[ | 1500 | 8 | 8 | 12000 | 12000 |
| [2000 – 3000[ | 2500 | 12 | 20 | 30000 | 42000 |
| [3000 – 4000[ | 3500 | 10 | 30 | 35000 | 77000 |
| [4000 – 5000[ | 4500 | 14 | 44 | 63000 | 140000 |

| | | | | | |
|---|---|---|---|---|---|
| [5000 – 6000[ | 5500 | 11 | 55 | 60500 | 200500 |
| [6000 – 7000[ | 6500 | 8 | 63 | 52000 | 252500 |
| [7000 – 8000[ | 7500 | 7 | 70 | 52500 | 305000 |
| [8000 – 10000[ | 9000 | 5 | 75 | 45000 | 350000 |
| [10000 – 15000[ | 12500 | 3 | 78 | 37500 | 387500 |
| [15000 – 20000[ | 17500 | 1 | 79 | 17500 | 405000 |
| **Total** | **Total** | **79** | | **405000** | |

La médiane correspond à la 39,5$^{ème}$ observation (79/2). En consultant les fréquences cumulées croissantes, la classe médiane est la classe [4000 – 5000[. La médiane est donc :

$$4000 < Me < 5000$$
$$30 < 39,5 < 44$$
$$\frac{5000 - 4000}{44 - 30} = \frac{Me - 4000}{39,5 - 30}$$
$$Me = 4000 + \frac{1000}{14} \text{ x } 9,5 = 4678,57 \text{ DH}$$

En consultant les sommes cumulées croissantes, la moitié de la somme totale (202500 DH) se trouve dans la classe [5000 – 6000[. La médiale est donc :

$$6000 < Ml < 7000$$
$$200500 < 202500 < 252500$$
$$\frac{7000 - 6000}{252500 - 200500} = \frac{Ml - 6000}{202500 - 200500}$$
$$Ml = 6000 + \frac{1000}{52000} \text{ x } 2000 = 6038,46 \text{ DH}$$

L'étendu de la série est : 20000 – 1000 = 19000

L'indice de concentration est donné par la formule :

$$\text{Indice de concentration} = \frac{\text{Médiale - Médiane}}{\text{Etendu}} \times 100$$

$$\text{Indice de concentration} = \frac{6038,46 - 4678,57}{19000} \times 100 = 7,16\ \%$$

**c) Calculs pour les salariés femmes**

| Tranches de salaire | ci | Ni | ni cumulé | ni ci | ni ci cumulé |
|---|---|---|---|---|---|
| [1000 ; 2000[ | 1500 | 4 | 4 | 6000 | 6000 |
| [2000 ; 3000[ | 2500 | 9 | 13 | 22500 | 28500 |
| [3000 ; 4000[ | 3500 | 6 | 19 | 21000 | 49500 |
| [4000 ; 5000[ | 4500 | 10 | 29 | 45000 | 94500 |
| [5000 ; 6000[ | 5500 | 8 | 37 | 44000 | 138500 |
| [6000 ; 7000[ | 6500 | 6 | 43 | 39000 | 177500 |
| [7000 ; 8000[ | 7500 | 5 | 48 | 37500 | 215000 |
| [8000 ; 10000[ | 9000 | 2 | 50 | 18000 | 233000 |
| [10000 ; 15000[ | 12500 | 1 | 51 | 12500 | 245500 |
| [15000 ; 20000[ | 17500 | 0 | 51 | 0 | 245500 |
| **Total** | **Total** | **51** | | **245500** | |

La médiane correspond à la 25,5$^{ème}$ observation (51/2). En consultant les fréquences cumulées croissantes, la classe médiane est la classe [4000 – 5000[. La médiane est donc :

$$4000 < Me < 5000$$
$$19 < 25,5 < 29$$

$$\frac{5000 - 4000}{29 - 19} = \frac{Me - 4000}{25,5 - 19}$$

$$Me = 4000 + \frac{1000}{10} \times 6,5 = 4650,00 \text{ DH}$$

En consultant les sommes cumulées croissantes, la moitié de la somme totale (122750 DH) se trouve dans la classe [5000 – 6000[. La médiale est donc :

$$5000 < Ml < 6000$$
$$94500 < 122750 < 138500$$

$$\frac{6000 - 5000}{138500 - 94500} = \frac{Ml - 5000}{122750 - 94500}$$

$$Ml = 5000 + \frac{1000}{44000} \text{ x } 28250 = 5642{,}05 \text{ DH}$$

L'étendu de la série est : 15000 – 1000 = 14000

L'indice de concentration est donné par la formule :

$$\text{Indice de concentration} = \frac{\text{Médiale - Médiane}}{\text{Etendu}} \times 100$$

$$\text{Indice de concentration} = \frac{5642{,}05 - 4650{,}00}{14000} \times 100 = 7{,}09\ \%.$$

**Récapitulatif des résultats.**

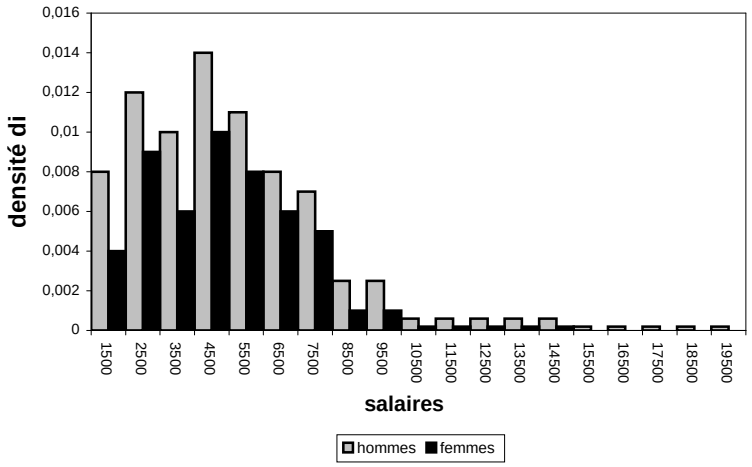
Nous reproduisons, sur un même tableau, les résultats de l'exemple 8 et ceux trouvés dans cet exemple.

| Salariés | Moyenne | Ecart type | Coefficient de variation | Indice de concentration |
|---|---|---|---|---|
| **Hommes** | 5126,58 | 2906,20 | 56,7 % | 7,16 % |
| **Femmes** | 4794,12 | 2275,57 | 47,47 % | 7,09 % |
| **Ensemble** | 5003,85 | 2667,10 | 53,3 % | 6,33 % |

En moyenne, un salarié de l'entreprise perçoit un salaire de 5003,85. La répartition des salaires est caractérisée par une forte dispersion.

La répartition des salaires varie selon le sexe, en effet, un salarié homme touche, en moyenne, plus qu'un salarié femme, (respectivement 5126,58 DH et 4794,12 DH). Les salaires sont plus dispersés chez les hommes que chez les femmes, 56,7 % pour les premiers et 47,47 % pour les seconds, alors que la concentration des salaires est légèrement moins forte chez les femmes (7,09 % contre 7,16 % pour les hommes). Ce résultat peut être illustré par cette représentation graphique :

Répartition des salaires selon le sexe




## 3.6. EXERCICES D'APPLICATION.

### 3.6.1. Exercice.

Les relevés statistiques des tailles, en mètre, de 20 personnes sont consignés dans le tableau suivant :

| | | | | |
|---|---|---|---|---|
| 1,58 | 1,62 | 1,75 | 1,58 | 1,70 |
| 1,75 | 1,70 | 1,58 | 1,62 | 1,82 |
| 1,62 | 1,82 | 1,70 | 1,58 | 1,75 |
| 1,70 | 1,58 | 1,62 | 1,70 | 1,85 |

a) Classer cette série statistique de 20 observations en série statistique équivalente sous la forme d'une série ($x_i$ ; $n_i$) ;
b) Calculer la variance et l'écart type de cette série ;
c) Calculer le coefficient de dispersion et l'indice de concentration de cette série.

**Solution** : a) Facile à faire ; b) $S^2 = 0,0075$ et $S = 8,64$ cm
c) CV = 5,14 % et Indice de concentration = 2,59 %

### 3.6.2. Exercice.

On a relevé, sur 2 mois, les chiffres d'affaires des ventes d'un magasin, les résultats sont donnés dans le tableau suivant :

| CA en DH | Nombre de jours | | CA en DH | Nombre de jours |
|---|---|---|---|---|
| [2 000 ; 4 000[ | 2 | | [10 000 ; 12 000[ | 14 |
| [4 000 ; 6 000[ | 6 | | [12 000 ; 14 000[ | 11 |
| [6 000 ; 8 000[ | 8 | | [14 000 ; 16 000[ | 5 |
| [8 000 ; 10 000[ | 10 | | [16 000 ; 18 000[ | 4 |

a) Calculer la moyenne et l'écart type de chiffre d'affaires ;
b) Calculer le coefficient de dispersion

**Solution** : a) $\bar{x} =$ 10366,67 DH et S = 3549,491356 DH ; b) CV = 34 %

### 3.6.3. Exercice.

On a recensé l'ancienneté, en années par défaut, de 45 agents d'une entreprise, elle se répartit comme suit :

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 2 | 6 | 3 | 5 | 3 | 2 | 6 | 5 | 1 |
| 3 | 3 | 2 | 5 | 6 | 1 | 4 | 5 | 6 |
| 1 | 2 | 3 | 5 | 2 | 5 | 6 | 1 | 3 |
| 5 | 5 | 1 | 2 | 2 | 3 | 3 | 6 | 6 |
| 3 | 2 | 3 | 3 | 3 | 6 | 5 | 5 | 1 |

a) Calculer la moyenne et l'écart type de l'ancienneté ;
b) Calculer le coefficient de dispersion et donner une interprétation du résultat.

**Solution** : a) Moyenne = 3.56 et S = 1,73 ; b) CV = 49 %

### 3.6.4. Exercice.

La série statistique donnant les effectifs de 40 classes d'une école est représentée par le tableau suivant :

| Nombre d'étudiants | Nombre de classes | | Nombre d'étudiants | Nombre de classes |
|---|---|---|---|---|
| [12 ; 16[ | 3 | | [24 ; 28[ | 12 |
| [16 ; 20[ | 5 | | [28 ; 32[ | 7 |
| [20 ; 24[ | 9 | | [32 ; 36[ | 4 |

a) Calculer la moyenne d'étudiant par classe et l'écart type de cette série ;
b) Calculer le coefficient de dispersion et donner une interprétation du résultat ;
c) Calculer l'indice de concentration. Qu'en déduire ?

**Solution** : a) $\bar{x} =$ 24,7 étudiants par classe et S = 5,47 ; b) CV = 22 %

c) Indice de concentration = 4,58 %

### 3.6.5. Exercice.

Les rendements à l'hectare d'une exploitation agricole composée de 200 lots, d'un hectare chacun, sont répartis comme suit :

| **Rendements en quintaux** | **Nombre de lots** | | **Rendements en quintaux** | **Nombre de lots** |
|---|---|---|---|---|
| [15 ; 17[ | 3 | | [25 ; 27[ | 26 |
| [17 ; 19[ | 5 | | [27 ; 29[ | 28 |
| [19 ; 21[ | 9 | | [29 ; 31[ | 33 |
| [21 ; 23[ | 12 | | [31 ; 33[ | 34 |
| [23 ; 25[ | 18 | | [33 ; 35[ | 32 |

a) Calculer la moyenne et l'écart type de cette série ;
b) Calculer le coefficient de dispersion et donner une interprétation du résultat ;
c) Calculer l'indice de concentration. Interpréter le résultat.

**Solution** : a) $\bar{x} = 28{,}2$ quintaux et S = 4,56 quintaux ; b) CV = 16 %
c) Indice de concentration = 3,85 %.

### 3.6.6. Exercice.

On considère les notes obtenues par les étudiants d'une classe, dans plusieurs matières, ayant chacune un coefficient de pondération différent.

| **Matières** | **Coefficients** | **Notes** | **Nombre d'étudiants** |
|---|---|---|---|
| Math | 4 | [10 ; 12[ | 8 |
| | | [12 ; 14[ | 12 |
| | | [14 ; 16[ | 10 |
| Economie | 2 | [7 ; 9[ | 4 |
| | | [9 ; 11[ | 7 |
| | | [11 ; 13[ | 13 |
| | | [13 ; 15[ | 6 |
| Compta | 3 | [6 ; 8[ | 2 |
| | | [8 ; 10[ | 6 |
| | | [10 ; 12[ | 7 |
| | | [12 ; 14[ | 10 |
| | | [14 ; 16[ | 4 |
| | | [16 ; 18[ | 1 |

a) Calculer les moyennes et les écarts types des notes des étudiants dans chaque matière ;
b) Calculer la moyenne générale et l'écart type de tous les étudiants.

**Solution** : a) Pour les math : $\bar{x} = 13{,}13$ et S = 1,54

Pour l'économie : $\bar{x} = 11{,}40$ et S = 1,87

Pour la compta : $\bar{x} = 11{,}73$ et S = 2,45

b) Moyenne générale = 12,28 Ecart type général = 0,77

### 3.6.7. Exercice.

L'entreprise SONFI commercialise du matériel, des logiciels et des consommables informatiques, la répartition des chiffres d'affaires en pourcentage des 5 dernières années est donnée par le tableau suivant :

| **Années** | **Chiffres d'affaires en pourcentage (%)** | | | |
|---|---|---|---|---|
| | **Micro** | **Logiciels** | **Consommables** | **Total CA** |
| 2001 | 40 | 30 | 30 | 100 |
| 2002 | 41 | 34 | 25 | 100 |
| 2003 | 40 | 37 | 23 | 100 |
| 2004 | 42 | 33 | 25 | 100 |
| 2005 | 43 | 32 | 25 | 100 |

a) Calculer les moyennes et les écarts types des pourcentages des chiffres d'affaires de chaque département ;
b) l'entreprise SONFI réalise, en 2006, un chiffre d'affaires de 2 524 312,36 DH dans le département micro, combien a-t-elle réalisé, en moyenne, dans les 2 autres départements ?

**Solution** : a) Pour le département micro : $\bar{x} = 41{,}2$ et S = 1,30

Pour le département logiciels : $\bar{x} = 33{,}2$ et S = 2,59

Pour le département consommables : $\bar{x} = 25{,}6$ et S = 2,61

b) Pour le département logiciels : CA = 2034154,62 DH
Pour le département consommables : CA = 1568504,77 DH.

### 3.6.8. Exercice.

La société CDG rémunère ses 25 salariés mensuellement et calcule :

$$\sum_{i=1}^{25} x_i = 125\,000{,}00\,\text{DH} \quad \text{et} \quad \sum_{i=1}^{25} x_i^2 = 652\,456\,000{,}00\,\text{DH}$$

a) Calculer la moyenne et l'écart type des salaires des 25 agents de la société ;
b) Que deviennent cette moyenne et cet écart type si l'on augmente chaque agent de 10% ?
c) Que deviennent cette moyenne et cet écart type si l'on augmente chaque agent de 1000 DH par mois ?

**Solution** : a) $\bar{x} = 5\,000{,}00$ DH et S = 1047,97 DH

b) $\bar{y} = 5\,500{,}00$ DH et Sy = 1 152,77 DH

c) $\bar{y} = 6\,000{,}00$ DH et Sy = 1 047,97 DH

### 3.6.9. Exercice.

Le relevé statistique des poids et des longueurs des barres de fer fabriquées par la société MARFER a donnée, pour une journée de production, les résultats suivants :

| **Poids (Kg)** | **Longueurs (cm)** | $\mathbf{n_i}$ | | **Poids (Kg)** | **Longueurs (cm)** | $\mathbf{n_i}$ |
|---|---|---|---|---|---|---|
| 5,80 | [490 ; 500[ | 12 | | 6,20 | [540 ; 550[ | 2 |
| | [500 ; 510[ | 25 | | | [550 ; 560[ | 4 |
| | [510 ; 520[ | 5 | | | [560 ; 570[ | 8 |
| 5,90 | [500 ; 510[ | 4 | | | [570 ; 580[ | 12 |
| | [510 ; 520[ | 36 | | | [580 ; 590[ | 6 |
| | [520 ; 530[ | 9 | | | [590 ; 600[ | 2 |
| 6,00 | [510 ; 520[ | 8 | | 6,30 | [560 ; 570[ | 1 |
| | [520 ; 530[ | 41 | | | [570 ; 580[ | 5 |
| | [530 ; 540[ | 10 | | | [580 ; 590[ | 4 |
| 6,10 | [540 ; 550[ | 3 | | | [590 ; 600[ | 20 |
| | [550 ; 560[ | 14 | | | [600 ; 610[ | 10 |
| | [560 ; 570[ | 2 | | | [610 ; 620[ | 4 |

$n_i$ : nombre de barres ayant les caractéristiques de poids et de longueur indiquées dans le tableau.

a) Calculer la longueur moyenne et l'écart type des barres de fer de 6,20 Kg de poids ;
b) Calculer le poids moyen et l'écart type des barres de fer de longueurs comprises entre 560 et 570 cm ;
c) Calculer le poids moyen et l'écart type d'une barre de fer ;
d) Calculer la longueur moyenne et l'écart type d'une barre de fer ;
e) Quels sont les modes en poids et en longueur des barres de fer fabriquées par la société MARFER ?
f) Quelle est la longueur médiane des barres de fer ?
g) Quel est le poids médiant des barres de fer ?

**Solution** : a) $\bar{x} =$ 571,47 cm et S = 12,34 cm ; b) $\bar{x} =$ 6,19 Kg et S = 0,051 Kg

c) $\bar{x} =$ 6,03 Kg et S = 0,173 Kg ; d) $\bar{x} =$ 540,79 cm et S = 33,98 cm

e) Mode en poids : Mo = 6 Kg et Classe modale en longueur : [520 ; 530[

f) Me = 526,7 cm ; g) Me = 5,9 + $\frac{0,1}{0,24}$ x 0,13 = 5,95 Kg

**3.6.10. Exercice.**

Le relevé des entrées des 5 salles d'un cinéma, relevées au cours de la semaine passée, a donné le tableau suivant :

| **Jours** | **Ciné N° 1** | **Ciné N° 2** | **Ciné N° 3** | **Ciné N° 4** | **Ciné N° 5** |
|---|---|---|---|---|---|
| Lundi | 100 | 201 | 350 | 250 | 283 |
| Mardi | 102 | 210 | 362 | 242 | 241 |
| Mercredi | 110 | 204 | 342 | 236 | 263 |
| Jeudi | 105 | 206 | 382 | 246 | 285 |
| Vendredi | 100 | 212 | 366 | 283 | 299 |
| Samedi | 102 | 220 | 354 | 255 | 201 |
| Dimanche | 121 | 231 | 328 | 222 | 204 |
| **Capacités** | **250** | **250** | **400** | **350** | **300** |

a) Calculer la moyenne et l'écart type des entrées de l'ensemble des cinémas pour chaque jour de la semaine ;
b) Calculer la moyenne et l'écart type des entrées de chaque cinéma pendant la semaine passée ;
c) Quel est le cinéma qui affiche le meilleur taux de remplissage pour la semaine passée ?
d) Quel est le jour qui affiche le meilleur taux de remplissage global pour les 5 cinémas ?

**Solution**

a)

| **Jours** | **Moyenne** | **Ecart type** |
|---|---|---|
| Lundi | 236,8 | 62,5 |
| Mardi | 231,4 | 67,2 |
| Mercredi | 231,0 | 59,0 |
| Jeudi | 244,8 | 75,4 |
| Vendredi | 252,0 | 63,2 |
| Samedi | 226,4 | 68,1 |
| Dimanche | 221,2 | 55,6 |

b)

| | **Ciné N° 1** | **Ciné N° 2** | **Ciné N° 3** | **Ciné N° 4** | **Ciné N° 5** |
|---|---|---|---|---|---|
| Moyenne | 105,7 | 212,0 | 354,9 | 247,7 | 253,7 |
| Ecart type | 7,6 | 10,4 | 17,4 | 18,9 | 39,6 |

c) C'est le cinéma N°3 qui affiche le meilleur taux de remplissage pour la semaine passée.
d) C'est le Vendredi qui affiche le meilleur taux de remplissage global pour les 5 cinémas.

# PARTIE 2
# STATISTIQUE DESCRIPTIVE A DEUX VARIABLES

La statistique descriptive à deux variables est l'ensemble des méthodes qui permet d'obtenir et de faire un 1$^{er}$ traitement des informations relatives à deux caractères particuliers d'individus d'une population donnée.

La statistique descriptive a plusieurs objectifs :

- recueillir l'ensemble des données relatives à deux caractères particuliers d'individus d'une population donnée ;

- classer l'ensemble de ces données selon des séries statistiques afin de permettre d'en faire :
  * des représentations graphiques pour en visualiser l'allure ;
  * des traitements mathématiques pour en déterminer certaines caractéristiques ;
  * des traitements mathématiques pour en déterminer les relations possibles existants entre ces caractères.

Dans cette partie, nous axerons notre propos sur le dernier point relatif à la détermination des relations de corrélation entre les caractères étudiés.

# CHAPITRE 4

# REGRESSION ET CORRELATION

## 4.1. INTRODUCTION.

On constate, très souvent, dans la pratique, qu'il existe des relations entre deux ou plusieurs variables. En analyse de régression, on cherche à expliquer une variable métrique y qui dépend d'une ou de plusieurs variables explicatives métriques $x_1$, $x_2$, $x_3$, . . . .$x_p$. A cette fin, un modèle mathématique peut représenter convenablement la relation entre y et les $x_i$, ce modèle servira aussi pour faire des prévisions.

$$Y = f(x_1, x_2, \ldots .x_p)$$

La variable Y s'appelle la variable **expliquée**, dépendante, endogène, tandis que les variables $x_1$, $x_2$, $x_3$, . . . .$x_p$ sont les variables **explicatives**, indépendantes, exogènes.

S'appuyant sur des données observées, l'analyse de régression consiste à ajuster un modèle explicatif $y = f(x_i)$.

## 4.2. REGRESSION SIMPLE.

S'il n'y a qu'une seule variable explicative, on dira que le modèle de régression est simple. Son but est de confirmer empiriquement une relation de cause à effet entre deux variables. Ensuite, si cette relation est confirmée, il y aura lieu d'en évaluer l'intensité.

4.2.1. Notion de covariance.

### 4.2.1.1. Définition.

On définit la covariance de deux variables statistiques par la moyenne arithmétique des produits des différences des observations par rapport à leur moyenne :

**- Cas d'une série statistique double** :

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

$$COV(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \times (y_i - \bar{y})}{n}$$

**- Cas d'un tableau de contingences** :

si x possède k modalités : $x_1, x_2, x_3, \ldots\ldots x_i, \ldots\ldots, x_k$

et si y possède p modalités : $y_1, y_2, y_3, \ldots\ldots y_j, \ldots\ldots, y_p$

$$COV(x,y) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{p} n_{ij}(x_i - \bar{x}) \times (y_j - \bar{y})}{n}$$

La covariance a pour but d'étudier le sens de la relation entre deux variables statistiques :

- Une covariance positive indique une relation croissante, c'est-à-dire que les deux variables statistiques varient dans le même sens ; les valeurs élevées d'une série correspondent aux valeurs élevées de l'autre ;

- Une covariance négative indique une relation décroissante, c'est-à-dire que les deux variables statistiques varient en sens inverse ; les valeurs élevées d'une série correspondent aux valeurs faibles de l'autre.

**4.2.1.1. Propriétés.**

- Formule développée de la covariance :

$$COV(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \times (y_i - \bar{y})}{n}$$

$$COV(x,y) = \frac{\sum_{i=1}^{n}(x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x}\bar{y})}{n}$$

$$COV(x,y) = \frac{\sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \bar{x}\bar{y}}{n}$$

$$COV(x,y) = \frac{\sum_{i=1}^{n} x_i y_i}{n} - \bar{y}\bar{x} - \bar{x}\bar{y} + \bar{x}\bar{y} \quad COV(x,y) = \frac{\sum_{i=1}^{n} x_i y_i}{n} - \bar{x}\bar{y}$$

La covariance est égale à la différence entre la moyenne des produits et le produit des moyennes.

Dans le cas d'un tableau de contingences :

$$COV(x,y) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{p} n_{ij} x_i y_i}{n} - \bar{x}\bar{y}$$

**- Transformation linéaire** :

Soit la transformation linéaire d'une variable statistique x :

x' = ax + b, avec a et b deux constantes quelconques.

Soit la transformation linéaire d'une variable statistique y :

y' = a'y + b', avec a' et b' deux constantes quelconques.

$$COV(x',y') = \frac{\sum_{i=1}^{n} (x_i' - \bar{x}') \times (y_i' - \bar{y}')}{n}$$

$$COV(x',y') = \frac{\sum_{i=1}^{n} (ax_i + b - a\bar{x} - b) \times (a'y_i + b' - a'\bar{y} - b')}{n}$$

$$COV(x',y') = \frac{\sum_{i=1}^{n} a(x_i - \bar{x}) \times a'(y_i - \bar{y})}{n}$$

$$COV(x', y') = \frac{a \times a' \sum_{i=1}^{n} (x_i - \bar{x}) \times (y_i - \bar{y})}{n}$$

$$COV(x', y') = a\, a' \times COV(x, y)$$

- On peut démontrer la relation suivante **:**

$$\left|COV(x, y)\right| \leq Sx \times Sy$$

**Exemple 1 :** On considère un échantillon de 12 clients choisis au hasard. On note, pour un trimestre :

- x : le nombre d'articles achetés par chacun des 12 clients ;
- y : le nombre de visites à un centre commercial, de chaque client.

On obtient les résultats suivants :

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 34 | 42 | 53 | 30 | 50 | 60 | 46 | 57 | 32 | 24 | 36 | 28 |
| $y_i$ | 12 | 14 | 15 | 10 | 15 | 17 | 12 | 14 | 10 | 09 | 11 | 10 |

**Dans le but d'étudier le sens de la relation entre X et Y, calculons la covariance (X,Y).**

| | $x_i$ | $y_i$ | $x_i^2$ | $y_i^2$ | $x_i\, y_i$ |
|---|---|---|---|---|---|
| | 34 | 12 | 1156 | 144 | 408 |
| | 42 | 14 | 1764 | 196 | 588 |
| | 53 | 15 | 2809 | 225 | 795 |
| | 30 | 10 | 900 | 100 | 300 |
| | 50 | 15 | 2500 | 225 | 750 |
| | 60 | 17 | 3600 | 289 | 1020 |
| | 46 | 12 | 2116 | 144 | 552 |
| | 57 | 14 | 3249 | 196 | 798 |
| | 32 | 10 | 1024 | 100 | 320 |
| | 24 | 9 | 576 | 81 | 216 |
| | 36 | 11 | 1296 | 121 | 396 |
| | 28 | 10 | 784 | 100 | 280 |
| **Total** | **492** | **149** | **21774** | **1921** | **6423** |

$\sum_{i=1}^{12} x_i = 492$ et $\sum_{i=1}^{12} y_i = 149$

$\sum_{i=1}^{12} x_i^2 = 21774$ et $\sum_{i=1}^{12} y_i^2 = 1921$

$\sum_{i=1}^{12} x_i y_i = 6423$

$$\bar{x} = \frac{\sum_{i=1}^{12} x_i}{n} = \frac{492}{12} = 41$$

$$\bar{y} = \frac{\sum_{i=1}^{12} y_i}{n} = \frac{149}{12} = 12{,}4166667 = 12{,}42$$

$$S^2_x = \frac{\sum_{i=1}^{12} x_i^2}{n} - \bar{x}^2 = \frac{21774}{12} - 41^2 = 133{,}5$$

$$S_x = \sqrt{133{,}5} = 11{,}55$$

$$S^2_y = \frac{\sum_{i=1}^{12} y_i^2}{n} - \bar{y}^2 = \frac{1921}{12} - 12{,}4166667^2 = 5{,}91$$

$$S_y = \sqrt{5{,}91} = 2{,}43$$

$$COV(x,y) = \frac{\sum_{i=1}^{12} x_i y_i}{n} - \bar{x}\,\bar{y} = \frac{6423}{12} - 41 \times 12{,}4166667 = 26{,}17$$

On vérifie bien que : $COV(x,y) < S_x S_y$ en effet :

$COV(x,y) = 26{,}17 < S_x S_y = 11{,}55 \times 2{,}43 = 28{,}06$

La covariance est positive, il y a donc une relation croissante entre le nombre d'articles achetés et le nombre de visites au centre commercial : c'est-à-dire que plus il y a de visites, plus il y a d'articles achetés, ce qui semble tout à fait logique.

**Exemple 2 :** Le concours d'accès à un établissement de formation porte sur deux épreuves : "Expression et communication" et "Informatique". Les candidats qui se sont présentés à ce concours se répartissent, en fonction des notes obtenues à ces deux épreuves, de la manière suivante :

| **y** / **x** | 3 | 7 | 10 | 12 | 15 |
|---|---|---|---|---|---|
| 7 | 0 | 3 | 9 | 7 | 11 |
| 9 | 10 | 13 | 18 | 16 | 13 |
| 11 | 9 | 11 | 14 | 17 | 14 |
| 14 | 12 | 9 | 7 | 5 | 2 |

x : note sur 20 obtenue en expression et communication ;
y : note sur 20 obtenue en informatique.

**Dans le but d'étudier le sens de la relation entre x et y, calculons la covariance (x,y).**

Distribution marginale de x

| **x** | 7 | 9 | 11 | 14 | Total |
|---|---|---|---|---|---|
| **Effectifs** | 30 | 70 | 65 | 35 | 200 |

$$\bar{x} = \frac{\sum_{i=1}^{4} n_i x_i}{200} = \frac{2045}{200} = 10{,}23$$

$$S^2_x = \frac{\sum_{i=1}^{4} n_i x_i^{\,2}}{200} - \bar{x}^2 = \frac{21865}{200} - 10{,}23^2 = 4{,}67$$

$$Sx = \sqrt{4{,}67} = 2{,}16$$

En moyenne, les candidats qui se sont présentés au concours ont obtenu une note de 10,23 sur 20 en expression et communication.

Les notes obtenues en expression et communication s'écartent, en moyenne, de 2,16 points de la note moyenne.

Distribution marginale de Y

| **y** | 3 | 7 | 10 | 12 | 15 | Total |
|---|---|---|---|---|---|---|
| **Effectifs** | 31 | 36 | 48 | 45 | 40 | 200 |

$$\bar{y} = \frac{\sum_{i=1}^{5} n_i y_i}{200} = \frac{1965}{200} = 9{,}83$$

$$S^2_y = \frac{\sum_{i=1}^{5} n_i y_i^2}{200} - \bar{y}^2 = \frac{22323}{200} - 9{,}83^2 = 14{,}99$$

$$Sy = \sqrt{14{,}99} = 3{,}87$$

En moyenne, les candidats qui se sont présentés au concours ont obtenu une note de 9,83 sur 20 en informatique.

Les notes obtenues en informatique s'écartent, en moyenne, de 3,87 points de la note moyenne.

Intensité de la relation linéaire entre X et Y

$$COV(x, y) = \frac{\sum_{i=1}^{4}\sum_{j=1}^{5} n_{ij}\, x_i\, y_j}{200} - \bar{x}\,\bar{y}$$

$$\sum_{i=1}^{4}\sum_{j=1}^{5} n_{ij}\, x_i\, y_j = 7x3x0+7x7x3+7x10x9+7x12x7+7x15x11$$
$$+ 9x3x10+9x7x13+9x10x18+9x12x16+9x15x13$$
$$+11x3x9+11x7x11+11x10x14+11x12x17+11x15x14$$
$$+14x3x12+14x7x9+14x10x7+14x12x5+14x15x2$$

$$\sum_{i=1}^{4}\sum_{j=1}^{5} n_{ij}\, x_i\, y_j = 19576.$$

$$COV(x, y) = \frac{19576}{200} - 10{,}23 \times 9{,}83 = -2{,}68$$

La covariance est négative, il y a donc une relation décroissante entre les notes d'expression communication et les notes d'informatique. En d'autres termes, les candidats bons en informatique sont, en moyenne, faibles en expression et communication.

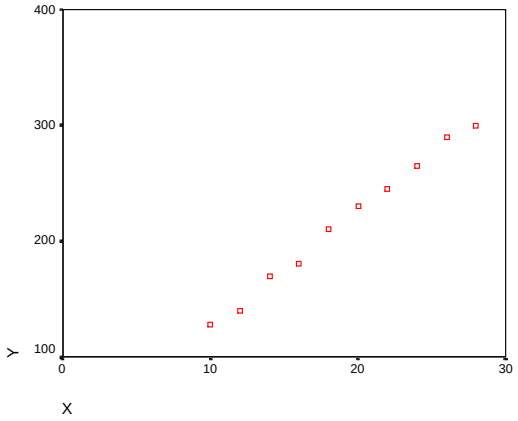
### 4.2.2. Identification du modèle.

On doit préciser la variable dont on veut expliquer les variations (variable dépendante y), puis celle qui est la cause de ces variations (variable explicative x).

Le diagramme de dispersion d'une variable y en fonction d'une autre variable x est formé des points moyens conditionnels de coordonnées ($x_i$ , $y_i$), et donne une idée de la façon dont varie, en moyenne, la variable y en fonction de la variable x.

A partir du diagramme de dispersion, on peut souvent représenter une courbe continue approchant les données. Cette courbe est appelée courbe d'ajustement (voir graphe page suivante).

**Diagramme de dispersion**



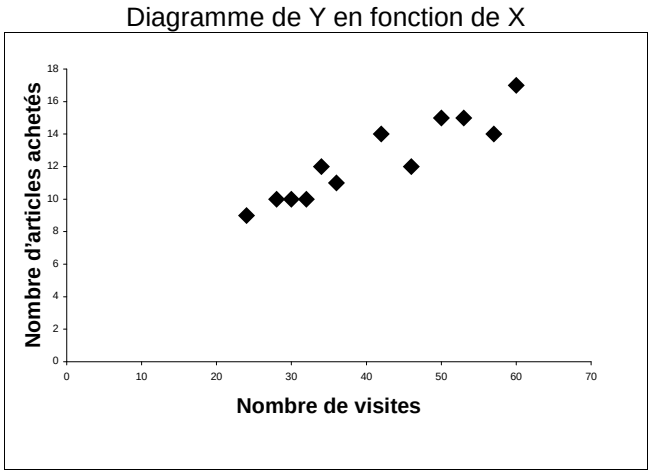

Bien que la relation entre deux variables ne soit pas toujours linéaire, on accepte, dans une première approximation, de considérer que cette relation est linéaire et ce pour les raisons simples suivantes :

- On peut toujours, dans une première approximation, approcher une courbe par la corde qui la soutient ;
- la théorie de la régression linéaire est beaucoup plus développée et surtout beaucoup plus simple à appliquer et à interpréter que celle de la régression non linéaire ;

La régression linéaire permet donc de déterminer la droite qui s'ajuste au mieux aux valeurs observées. Cette droite est appelée droite de régression de y en fonction de x.

**Exemple 3 :** Reprenons les données de l'exemple 1, et traçons le diagramme de dispersion de Y en fonction de X :

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Xi** | 34 | 42 | 53 | 30 | 50 | 60 | 46 | 57 | 32 | 24 | 36 | 28 |
| **Yi** | 12 | 14 | 15 | 10 | 15 | 17 | 12 | 14 | 10 | 09 | 11 | 10 |




### 4.2.3. Ajustement du modèle.

Le modèle théorique en régression linéaire simple s'écrit :

$$y = a\,x + b + \varepsilon$$

Le paramètre « a » donne la pente de la droite, appelée coefficient de régression ; il mesure la variation de y lorsque x augmente d'une unité. Le paramètre « b » est l'ordonnée à l'origine, c'est-à-dire la valeur prise par y lorsque x = 0.

ε Représente l'erreur aléatoire, elle est non observable et comprend à la fois les erreurs de mesure sur les valeurs observées de Y et tous les autres facteurs explicatifs non pris en compte dans le modèle.

L'analyse de régression repose sur un certain nombre d'hypothèses qui sont :

- La variable explicative x est mesurée sans erreur ;
- Les erreurs aléatoires ε sont distribuées normalement avec une moyenne nulle et une variance constante inconnue ;
- Les erreurs aléatoires ε sont indépendantes avec la variable explicative ;
- Les erreurs aléatoires ε sont indépendantes entre elles.

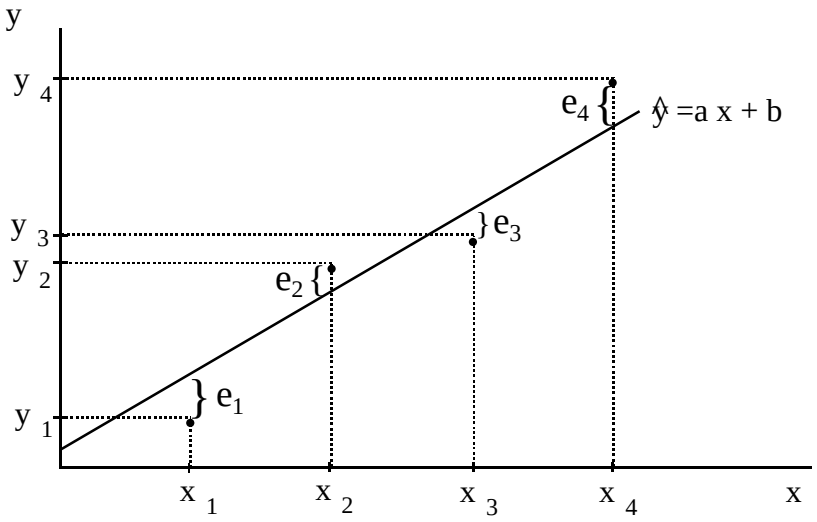
Il existe différentes méthodes pour ajuster une droite de régression. La méthode la plus utilisée est la méthode des moindres carrés.

La méthode des moindres carrés est une méthode d'ajustement qui consiste à minimiser la somme des carrés des différences entre les valeurs observées, $y_i$, et les valeurs estimées par la droite, $\hat{y}_i$ différence appelée résidu.

Le modèle empirique, estimé à partir des observations, sera désigné de cette façon :

$$\hat{y} = a_0 x + b_0$$

$a_0$ et $b_0$ sont des estimations des paramètres a et b du modèle théorique.




On définit le i-ème résidu ( noté $e_i$ ) comme étant la différence mesurée **verticalement** sur le graphique entre la valeur observée de $y_i$ et sa valeur estimée : $e_i = y_i - \hat{y}_i$ .

On remarque que :

- le résidu est positif (ei >0) si yi se trouve au-dessus de la droite au point xi.
- le résidu est négatif (ei < 0) si yi se trouve au-dessous de la droite au point xi.
- le résidu est nul (ei = 0) si yi se trouve précisément sur la droite au point xi.

On désire expliquer les variations observées sur la variable dépendante y, c'est pour cette raison qu'il faut considérer les différences mesurées verticalement.

La méthode des moindres carrées est celle qui minimise la somme des carrés des résidus; symboliquement, on cherche à :

$$\text{Minimiser l'expression} : \sum_{i=1}^{n}\left( y_i - \hat{y}_i \right)^2 = \sum_{i=1}^{n} e_i^{\,2}$$

Avec le critère des moindres carrés, tous les résidus deviennent positifs; car sinon, en nous limitant aux résidus simples, il est impossible que des résidus positifs annulent des résidus négatifs.

Les démonstrations algébriques sont facilitées par le recours aux outils du calcul différentiel. La minimisation d'une fonction quadratique à plusieurs variables s'effectue en annulant les dérivées partielles de premier ordre et en vérifiant le signe des dérivées partielles de deuxième ordre.

**4.2.3.1. Calcul des coefficients.**

Par calcul différentiel, on cherche les 2 valeurs $a_0$ et $b_0$ qui minimisent la somme des carrés des résidus, cette somme quadratique est notée $f( a_0 , b_0)$, puisqu'elle est fonction de 2 termes inconnus $a_0$ et $b_0$ :

$$f(a_0, b_0) = \sum e_i^{\,2} = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - a_0 x_i - b_0)^2$$

$f(a_0 , b_0)$ est minimum lorsque les dérivées premières partielles de $f(a_0 , b_0)$ par rapport à $a_0$ et à $b_0$ sont nulles et que les dérivées secondes partielles sont positives.

Appelons :
- $f'_{a_0}$, la dérivée première partielle de f par rapport à $a_0$;
- $f''_{a_0}$, la dérivée seconde partielle de f par rapport à $a_0$

Les 2 conditions seront vérifiées si :
- $f'_{a_0} = 0$ et $f'_{b_0} = 0$ ;
- $f''_{a_0} > 0$ et $f''_{b_0} > 0$

**1ère Condition** : écrivons que les dérivées premières partielles sont nulles, c'est-à-dire que : $f'_{a_0} = 0$ et $f'_{b_0} = 0$.

$$f(a_0, b_0) = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - a_0 x_i - b_0)^2$$

On a :

$$f'_{b_0} = \sum -(y_i - a_0 x_i - b_0) = 0$$

$$\sum (y_i - a_0 x_i - b_0) = 0$$

$$\sum y_i - n\, b_0 - a_0 \sum x_i = 0$$

$$\sum y_i = n\, b_0 + a_0 \sum x_i$$

On a aussi :

$$f'_{a_0} = \sum -2\, x_i (y_i - a_0 x_i - b_0) = 0$$

$$\sum (x_i y_i - b_0\, x_i - a_0 x_i^2) = 0$$

$$\sum x_i y_i - b_0 \sum x_i - a_0 \sum x_i^2 = 0$$

$$\sum x_i y_i = b_0 \sum x_i + a_0 \sum x_i^2$$

On a donc un système de deux équations à deux inconnues, ces deux équations qui sont appelées équations normales sont :

$$\sum y_i = n\, b_0 + a_0 \sum x_i$$
$$\sum x_i y_i = b_0 \sum x_i + a_0 \sum x_i^2$$

**Calcul de $b_0$** : En considérant la seconde équation, on a successivement les égalités suivantes :

$$\sum y_i = n\, b_0 + a_0 \sum x_i \Rightarrow \quad n\, b_0 = \sum y_i - a_0 \sum x_i$$

$$b_0 = \frac{\sum y_i}{n} - a_0 \frac{\sum x_i}{n}$$

$$b_0 = \bar{y} - a_0 \bar{x}$$

**Calcul de $a_0$** : En considérant le première équation et en y remplaçant $b_0$ par l'expression qu'on vient d'établir, on a successivement les égalités suivantes :

$$\sum x_i y_i = b_0 \sum x_i + a_0 \sum x_i^2$$
$$\sum x_i y_i = (\bar{y} - a_0 \bar{x}) \sum x_i + a_0 \sum x_i^2$$
$$\sum x_i y_i = \bar{y} \sum x_i - a_0 \bar{x} \sum x_i + a_0 \sum x_i^2$$
$$\sum x_i y_i = n \bar{x} \bar{y} + a_0 (\sum x_i^2 - n \bar{x}^2)$$

$$a_0 = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2}$$

La droite de régression de Y en fonction de X, selon la méthode des moindres carrés est la droite d'équation :

$$\hat{y} = a_0 x + b_0$$

avec : $a_0 = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2}$ et $b_0 = \bar{y} - a_0 \bar{x}$

L'estimation de a et de b par la méthode des moindres carrés conduit aux formules équivalentes suivantes :

$$a_0 = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2} = \frac{COV(x,y)}{S_x^2}$$

D'où $\hat{y} = a_0 x + b_0 = a_0 x + (\bar{y} - a_0 \bar{x}) = a_0 (x - \bar{x}) + \bar{y}$

Ces estimateurs sont des fonctions linéaires des observations $x_1, x_2, \ldots x_n$.

**2è Condition** : montrons que les dérivées secondes partielles sont positives, c'est-à-dire que : $f''_{a_0} > 0$ et $f''_{b_0} > 0$.

$$f(a_0, b_0) = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - a_0 x_i - b_0)^2$$

On a :

$$f'_{a_0} = \sum -2\,x_i(y_i - a_0 x_i - b_0)$$
$$f''_{a_0} = \left[\sum -2\,x_i(y_i - a_0 x_i - b_0)\right]' = 2\sum x_i^2 \text{ qui est bien positif.}$$

et :

$$f'_{b_0} = 2\sum -(y_i - a_0 x_i - b_0)$$
$$f''_{b_0} = \left[2\sum -(y_i - a_0 x_i - b_0)\right]' = 2 \text{ qui est bien positif.}$$

Nous pouvons donc conclure que les valeurs de $a_0$ et $b_0$ que nous avons déterminées correspondent bien à un minimum de l'expression : $\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{n} e_i^{\,2}$

**Exemple 4 :** Reprenons les données de l'exemple 1 et déterminons la droite de régression de y en fonction de x.

| | $x_i$ | $y_i$ | $x_i^2$ | $y_i^2$ | $x_i\,y_i$ |
|---|---|---|---|---|---|
| | 34 | 12 | 1156 | 144 | 408 |
| | 42 | 14 | 1764 | 196 | 588 |
| | 53 | 15 | 2809 | 225 | 795 |
| | 30 | 10 | 900 | 100 | 300 |
| | 50 | 15 | 2500 | 225 | 750 |
| | 60 | 17 | 3600 | 289 | 1020 |
| | 46 | 12 | 2116 | 144 | 552 |
| | 57 | 14 | 3249 | 196 | 798 |
| | 32 | 10 | 1024 | 100 | 320 |
| | 24 | 9 | 576 | 81 | 216 |
| | 36 | 11 | 1296 | 121 | 396 |
| | 28 | 10 | 784 | 100 | 280 |
| **Total** | **492** | **149** | **21774** | **1921** | **6423** |

$\sum_{i=1}^{12} x_i = 492$ et $\sum_{i=1}^{12} y_i = 149$

$\sum_{i=1}^{12} x_i^2 = 21774$ et $\sum_{i=1}^{12} y_i^2 = 1921$

$\sum_{i=1}^{12} x_i y_i = 6423$

$\bar{x} = \frac{\sum_{i=1}^{12} x_i}{n} = \frac{492}{12} = 41$ et $\bar{y} = \frac{\sum_{i=1}^{12} y_i}{n} = \frac{149}{12} = 12{,}42$

$S^2_x = \frac{\sum_{i=1}^{12} x_i^2}{n} - \bar{x}^2 = \frac{21774}{12} - 41^2 = 133{,}5$

$S_x = \sqrt{133{,}5} = 11{,}55$

$S^2_y = \frac{\sum_{i=1}^{12} y_i^2}{n} - \bar{y}^2 = \frac{1921}{12} - 12{,}4166667^2 = 5{,}91$

$S_y = \sqrt{5{,}91} = 2{,}43$

$COV(x, y) = \frac{\sum_{i=1}^{12} x_i y_i}{n} - \bar{x}\,\bar{y} = \frac{6423}{12} - 41 \times 12{,}4166667 = 26{,}17$

La droite de régression de y en fonction de x, selon la méthode des moindres carrés est la droite d'équation :

$$\hat{y} = a_0 x + b_0$$

avec : $a_0 = \frac{\sum x_i y_i - n \bar{x}\bar{y}}{\sum x_i^2 - n \bar{x}^2}$ et $b_0 = \bar{y} - a_0 \bar{x}$

Les calculs donnent :

$$a_0 = \frac{6423 - 12 \times 41 \times 12{,}4166667}{21774 - 12 \times 41^2} = 0{,}196005 = 0{,}20$$

On peut vérifier que :

$$a_0 = \frac{COV(x, y)}{S^2_x} = \frac{26{,}17}{133{,}5} = 0{,}196005 = 0{,}20$$

$$b_0 = 12{,}4166667 - 0{,}196005 \text{ x } 41 = 4{,}3804617$$

La droite de régression de y en fonction de x, selon la méthode des moindres carrés est la droite d'équation :

$$\hat{y} = 0{,}20\,x + 4{,}38$$

**4.2.3.2. Propriétés de la droite de régression.**

1) La droite de régression passe par le point moyen de coordonnées : $(\bar{x}, \bar{y})$

2) $\sum y_i = \sum \hat{y}_i$ et $\sum (y_i - \hat{y}_i) = 0$

3) $\sum (y_i - \hat{y}_i)^2$ est la plus petite somme des carrés des écarts que l'on peut obtenir.

**4.2.3.3. Interprétation des coefficients a et b.**

Nous donnerons, sur des exemples pratiques, les interprétations qu'il y a lieu de donner des coefficients a et b, mais d'ores et déjà, nous pouvons dire :

**Le coefficient a** est le taux de croissance de la variable expliquée chaque fois que la variable explicative augment d'une unité.

**Le coefficient b** est l'ordonnée à l'origine, son interprétation requiert, dans chaque cas, de revenir au problème posé.

Dans le cas de l'exemple 4, le modèle d'ajustement a pour expression :

$\hat{y} = 0{,}20\,x + 4{,}38$ on peut interpréter les coefficients a et b comme suit :

* a = 0,20 veut dire que pour toutes les 10 visites, il y a 2 achats qui se réalisent ;

* b = 4,38 veut dire que même sans visite, il y a entre 4 et 5 articles vendus, ce qui semble aberrant car on ne peut imaginer des achats sans qu'il y ait des visites. Cette valeur non nulle de b n'a donc pas de signification physique, dans le cas de notre cas.

**Exemple 5 :** Reprenons les données de l'exemple 1 et vérifions les 3 remarques qu'on vient de citer.

La droite de régression de y en fonction de x selon la méthode des moindres carrés est la droite d'équation :

$$\hat{y} = 0{,}20\,x + 4{,}38$$

| $x_i$ | $y_i$ | $\hat{y}_i$ | $(y_i - \hat{y}_i)$ | $(y_i - \hat{y}_i)^2$ |
|---|---|---|---|---|
| 34 | 12 | 11,04 | 0,96 | 0,91 |
| 42 | 14 | 12,61 | 1,39 | 1,92 |
| 53 | 15 | 14,77 | 0,23 | 0,05 |
| 30 | 10 | 10,26 | -0,26 | 0,07 |
| 50 | 15 | 14,18 | 0,82 | 0,67 |
| 60 | 17 | 16,14 | 0,86 | 0,74 |
| 46 | 12 | 13,40 | -1,40 | 1,95 |
| 57 | 14 | 15,55 | -1,55 | 2,41 |
| 32 | 10 | 10,65 | -0,65 | 0,43 |
| 24 | 9 | 9,08 | -0,08 | 0,01 |
| 36 | 11 | 11,44 | -0,44 | 0,19 |
| 28 | 10 | 9,87 | 0,13 | 0,02 |
| **Total** | **492** | **149** | **149** | **0,00** | **9,37** |

1) Vérifions que la droite de régression passe bien par le point moyen de coordonnées $(\bar{x}, \bar{y})$, en effet :

$\hat{y} = 0{,}20 \times 41 + 4{,}38$ = 12,42 = $\bar{y}$

2) Vérifions aussi que $\sum y_i = \sum \hat{y}_i$ = 149 ce qui donne bien

$\sum (y_i - \hat{y}_i) = 0$

3) Enfin, on vérifie bien que $\sum (y_i - \hat{y}_i)^2$ = 9,37 est la plus petite somme des carrés des écarts que l'on peut obtenir. Rappelons que ce minimum est assuré par le choix des coefficients a et b.

### 4.3. QUALITE DE L'AJUSTEMENT.

#### 4.3.1. Coefficient de détermination.

La modèle d'ajustement que nous avons déterminé est de la forme : y = a x + b ; mathématiquement parlant, cette équation peut s'écrire aussi sous la forme : x = a' y + b'. Pour que ces deux équations aient une cohérence mathématique, on doit avoir :

y = a x + b = a ( a' y + b' ) + b = a a' y + a b' + b

Ce qui donne, en identifiant les termes y dans les deux membres, les conditions nécessaires suivantes :

a a' = 1 et a b' + b = 0

Mais comme les points de coordonnées ( $x_i$ , $y_i$ ) ne sont pas tous sur la droite de régression y = a x + b, la condition a a' = 1 ne peut être satisfaite avec exactitude.

La 1ère condition donne, en déduisant la formule de a' à partir de celle de a :

$$a\,a' = R^2 = \frac{\sum (x_i - \bar{x})^2 (y_i - \bar{y})^2}{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2} = \frac{COV(x, y)^2}{S^2_x \, S^2_y}$$

Compte tenu de l'inégalité : $|COV(x, y)| \leq Sx \times Sy$ , le coefficient $R^2 \leq 1$. De ce fait, le modèle d'ajustement adopté sera d'autant plus valide que le coefficient $R^2$ sera proche de 1.

On appelle $R^2$, le coefficient de détermination du modèle d'ajustement ; il est égal au pourcentage de la variation totale dans la variable y qui est expliquée par la régression. Il synthétise la capacité de la droite de régression à retrouver les différentes valeurs de la variable dépendante $y_i$

On pourrait introduire le coefficient $R^2$ d'une autre manière, en effet, la variation totale $\sum (yi - \hat{y}_i)^2$ observée sur la variable expliquée y peut être décomposée en 2 parties :

$$\sum \left(y_i - \bar{y}\right)^2 = \sum \left(\hat{y}_i - \bar{y}\right)^2 + \sum \left(y_i - \hat{y}_i\right)^2$$

- Le premier terme $\sum \left(\hat{y}_i - \bar{y}\right)^2$ désigné par SCR mesure la variation autour de la droite de régression, on l'appelle Somme des Carrés due à la Régression ;

- Le second terme, $\sum \left(y_i - \hat{y}_i\right)^2$ désigné par SCE, mesure la variation résiduelle, on l'appelle la somme des carrés due à l'erreur.

La somme des carrés totale SCT s'écrit donc :

SCT = SCR + SCE

Puisqu'on cherche à expliquer la variation totale de y autour de sa moyenne, SCT, on peut utiliser le coefficient de détermination $R^2$ comme indice de la qualité de l'ajustement de la droite aux données.

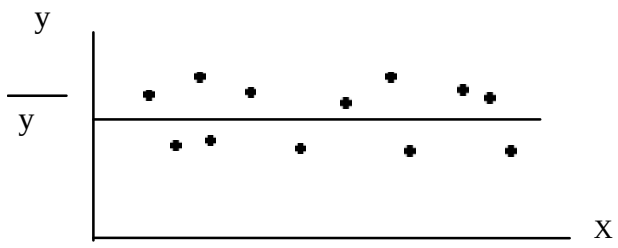
$$R^2 = \frac{SCR}{SCT} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (yi - \bar{y})^2}$$

Etudions tous les cas possibles des valeurs que peut prendre $R^2$ :

**- Cas où $R^2 = 0$ :**

Il faut pour cela que SCR = 0, alors le modèle utilisé n'explique aucune variation dans la variable dépendante y. En outre, SCR = 0 implique que toutes les valeurs prédites sont égales à la moyenne des y, soit $\hat{y}_i = \bar{y}$ pour i = 1, 2, . . . .n.

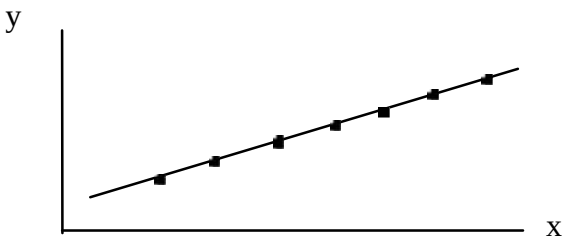
Graphiquement, dans le cas d'une régression simple, on aura la situation suivante, dans laquelle on peut voir clairement que la variable explicative x n'est d'aucune utilité pour prédire y.




**- Cas où $R^2 = 1$ :**

Il faut pour cela que SCR = SCT, ce qui revient à SCE = 0. S'il en est ainsi, le modèle utilisé explique toute la variation observée sur y. En outre, SCE = 0 implique que toutes les valeurs prédites sont égales aux valeurs observées correspondantes de y, c'est-à-dire : yi = $\hat{y}_i$ pour i = 1, 2, . . . n.

Graphiquement, on a la situation suivante dans laquelle le modèle de régression explique parfaitement les variations de y. La variable explicative x peut prédire sans erreur les valeurs de y, au moins pour les valeurs de l'échantillon.




**- Cas général : $R^2 < 1$**

En général, nous ne sommes ni dans le cas de $R^2 = 0$ ni dans celui de $R^2 = 1$ mais nous trouvons $R^2 < 1$ et plus $R^2$ est proche de 1 plus le modèle peut prétendre expliquer les valeurs de y par celles de x.

Le coefficient de détermination $R^2$ sert à définir le coefficient de corrélation de PEARSON R comme nous allons le voir juste après.

**Exemple 6 :** Reprenons les données de l'exemple 1 et décomposons la somme des carrés totale et calculons le coefficient de détermination.

| | $x_i$ | $y_i$ | $\hat{Y}i$ | $(y_i - \hat{y}_i)$ | $(\hat{y}_i - \bar{y})^2$ | $(y_i - \hat{y}_i)^2$ |
|---|---|---|---|---|---|---|
| | 34 | 12 | 11,04 | 0,18 | 1,89 | 0,91 |
| | 42 | 14 | 12,61 | 2,50 | 0,04 | 1,92 |
| | 53 | 15 | 14,77 | 6,66 | 5,52 | 0,05 |
| | 30 | 10 | 10,26 | 5,86 | 4,66 | 0,07 |
| | 50 | 15 | 14,18 | 6,66 | 3,10 | 0,67 |
| | 60 | 17 | 16,14 | 20,98 | 13,84 | 0,74 |
| | 46 | 12 | 13,40 | 0,18 | 0,95 | 1,95 |
| | 57 | 14 | 15,55 | 2,50 | 9,81 | 2,41 |
| | 32 | 10 | 10,65 | 5,86 | 3,12 | 0,43 |
| | 24 | 9 | 9,08 | 11,70 | 11,13 | 0,01 |
| | 36 | 11 | 11,44 | 2,02 | 0,97 | 0,19 |
| | 28 | 10 | 9,87 | 5,86 | 6,51 | 0,02 |
| **Total** | **492** | **149** | **149,00** | **70,92** | **61,55** | **9,37** |

**SCT** = $\sum (yi - \bar{y})^2$ = **70,92**

SCR = $\sum \left( \hat{y}_i - \bar{y} \right)^2$ = 61,55

SCE = $\sum \left( y_i - \hat{y}_i \right)^2$ = 9,37

On vérifie bien que :

SCR + SCE = 61,55 + 9,37 = 70,92 = SCT

$R^2 = \frac{SCR}{SCT} = \frac{61,55}{70,92} = 0,87$

On peut vérifier aussi que :

$R^2 = 0,93^2 = 0,87$

Le nombre de visites au centre commercial explique 87 % des variations du nombre d'articles achetés.

### 4.3.2. Coefficient de corrélation de PEARSON.

#### 4.3.2.1. Définition.

On définit à partir du coefficient de détermination $R^2$, le coefficient de corrélation linéaire R, il a pour objet de mesurer l'intensité de la liaison linéaire entre deux variables statistiques x et y.

Le coefficient de corrélation de x et y peut être estimé à l'aide d'un échantillon aléatoire de n couples d'observations par la formule suivante :

$$R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum x_i^2 - n\bar{x}^2}\sqrt{\sum y_i^2 - n\bar{y}^2}}$$

$$R = \frac{COV(x, y)}{S_x S_y}$$

Cette définition montre que le coefficient de corrélation possède le même signe que la covariance et qu'il est toujours compris entre -1 et +1 puisque comme on l'a vu : $R^2 < 1$

Le signe du coefficient de corrélation linéaire indique le sens de la relation entre x et y, ainsi :

- R = +1 : dans ce cas, les points se trouvent tous sur une même droite croissante, on parle de corrélation linéaire positive parfaite.
- R = -1 : dans ce cas, les points se trouvent tous sur une même droite décroissante, on parle de corrélation linéaire négative parfaite.
- R = 0 : dans ce cas, il n'y a aucune dépendance linéaire entre les deux variables, on parle de corrélation linéaire nulle.
- -1 < R < 0 : dans ce cas, les deux variables varient en sens inverse, la relation linéaire est faible ou forte selon que le coefficient de corrélation linéaire est proche de 0 ou de -1.
- 0 < R < 1 : dans ce cas, les deux variables varient dans le même sens, la relation linéaire est faible ou forte selon que le coefficient de corrélation linéaire est proche de 0 ou de 1.

Le problème de la régression est intimement lié à celui de la corrélation : plus la corrélation est forte entre deux variables, mieux l'on pourra prédire ou expliquer la valeur de la variable dépendante y en fonction de la variable explicative x.

On peut affirmer que la corrélation mesure l'**intensité** de la relation **linéaire** entre 2 variables aléatoires, tandis que la régression simple est une **équation** décrivant le plus adéquatement possible cette relation.

**Exemple 7 :** Reprenons les données de l'exemple 1 et calculons le coefficient de corrélation.

| | $x_i$ | $y_i$ | $x_i^2$ | $y_i^2$ | $x_i y_i$ |
|---|---|---|---|---|---|
| | 34 | 12 | 1156 | 144 | 408 |
| | 42 | 14 | 1764 | 196 | 588 |
| | 53 | 15 | 2809 | 225 | 795 |
| | 30 | 10 | 900 | 100 | 300 |
| | 50 | 15 | 2500 | 225 | 750 |
| | 60 | 17 | 3600 | 289 | 1020 |
| | 46 | 12 | 2116 | 144 | 552 |
| | 57 | 14 | 3249 | 196 | 798 |
| | 32 | 10 | 1024 | 100 | 320 |
| | 24 | 9 | 576 | 81 | 216 |
| | 36 | 11 | 1296 | 121 | 396 |
| | 28 | 10 | 784 | 100 | 280 |
| **Total** | **492** | **149** | **21774** | **1921** | **6423** |

$$\sum_{i=1}^{12} x_i = 492 \quad \text{et} \quad \sum_{i=1}^{12} y_i = 149$$

$$\sum_{i=1}^{12} x_i^2 = 21774 \quad \text{et} \quad \sum_{i=1}^{12} y_i^2 = 1921$$

$$\sum_{i=1}^{12} x_i y_i = 6423$$

$$\bar{x} = \frac{\sum_{i=1}^{12} x_i}{n} = \frac{492}{12} = 41 \quad \text{et} \quad \bar{y} = \frac{\sum_{i=1}^{12} y_i}{n} = \frac{149}{12} = 12{,}42$$

$$S^2_x = \frac{\sum_{i=1}^{12} x_i^2}{n} - \bar{x}^2 = \frac{21774}{12} - 41^2 = 133{,}5$$

$$S_x = \sqrt{133{,}5} = 11{,}55$$

$$S^2_y = \frac{\sum_{i=1}^{12} y_i^2}{n} - \bar{y}^2 = \frac{1921}{12} - 12{,}4166667^2 = 5{,}91$$

$$S_y = \sqrt{5{,}91} = 2{,}43$$

$$COV(x, y) = \frac{\sum_{i=1}^{12} x_i y_i}{n} - \bar{x}\,\bar{y} = \frac{6423}{12} - 41 \times 12{,}4166667 = 26{,}17$$

$$R = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum x_i^2 - n\bar{x}^2}\sqrt{\sum y_i^2 - n\bar{y}^2}}$$

$$R = \frac{6423 - 12 \times 41 \times 12{,}4166667}{\sqrt{21774 - 12 \times 41^2} \times \sqrt{1921 - 12 \times 12{,}4166667^2}} = 0{,}93$$

On peut aussi vérifier que :

$$R = \frac{COV(x, y)}{S_x S_y} = \frac{26{,}17}{11{,}55 \times 2{,}43} = 0{,}93$$

Il y a donc une forte corrélation linéaire croissante entre le nombre d'articles achetés et le nombre de visites des clients au centre commercial.

**4.3.2.2. Propriétés du coefficient de corrélation.**

Ces propriétés sont au nombre de deux :

- Le coefficient de corrélation linéaire est indépendant des unités de mesure.

- Le coefficient de corrélation linéaire est indépendant de toute transformation linéaire positive.

En effet, soit les transformations linéaires des variables statistiques x et y :

x' = ax + b, avec a et b deux constantes quelconques.

y' = a'y + b', avec a' et b' deux constantes quelconques.

$$R(x', y') = \frac{COV(x', y')}{Sx' \times Sy'}$$

$$R(x', y') = \frac{a \times a' \times COV(x, y)}{|a| Sx \times |a'| Sy}$$

$$R(x', y') = \pm \frac{COV(x, y)}{Sx \times Sy} \qquad => \qquad R(x ,y') = \pm R(x ,y)$$

Une transformation linéaire ne change pas l'intensité de la relation linéaire mais elle peut changer le sens de la relation.

### 4.4. CALCULS DES PREVISIONS.

Pour obtenir une prévision ponctuelle de Y pour une valeur particulière $x_0$ de X, il suffit de remplacer X par $x_0$ dans le modèle empirique, ce qui s'écrit :

$$\hat{y} = a_0 x_0 + b_0$$

**Exemple 8 :** Reprenons les données de l'exemple 1 et effectuons une prévision du nombre d'articles que pourrait acheter un client après 25 visites au centre commercial.

La droite de régression de y en fonction de x, selon la méthode des moindres carrés est la droite d'équation :

$$\hat{y} = 0{,}20\, x + 4{,}38$$

Si $x_0 = 25$ alors $\hat{y} = 0{,}20 \times 25 + 4{,}38 = 9{,}38$ soit 9 ou 10 articles achetés après 25 visites au centre commercial.

## 4.5. REGRESSION NON LINEAIRE SIMPLE.

Dans certaines situations, il arrive que le nuage de points du diagramme ne ressemble pas à une relation linéaire. La régression linéaire n'est donc pas adaptée. On doit donc ajuster une courbe non linéaire. On parle de régression non linéaire.

Certains modèles non linéaires peuvent être ramené à des régressions linéaires grâce à des transformations de variables. C'est les cas notamment du modèle exponentiel en $a^x$ et du modèle polynomial en $x^a$ .

### 4.5.1. Modèle exponentiel.

Le modèle général exponentiel a pour équation :

$$y = a_0 \times b_0^{\,x}$$

Grâce à une transformation logarithmique, le modèle devient linéaire :

$\text{Log}(y) = \text{Log}(a_0 \times b_0^{x})$

$\text{Log}(y) = \text{Log}(a_0) + \text{Log}(b_0) \times x$

On pose :

$y' = \text{Log}(y)$, $\quad a'_0 = \text{Log}(b_0)$ et $b'_0 = \text{Log}(b_0)$

Le modèle devient :

$y' = a'_0 + b'_0 \times x$

On détermine $a'_0$ et $b'_0$ par les formules générales de la régression linéaire.

$$a'_0 = = \frac{\sum x_i y'_i - \bar{x}\,\bar{y'}}{\sum x_i^{\,2} - n\,\bar{x}^{2}} \qquad \text{et} \qquad b'_0 = \bar{y'} - b_1' \bar{x}$$

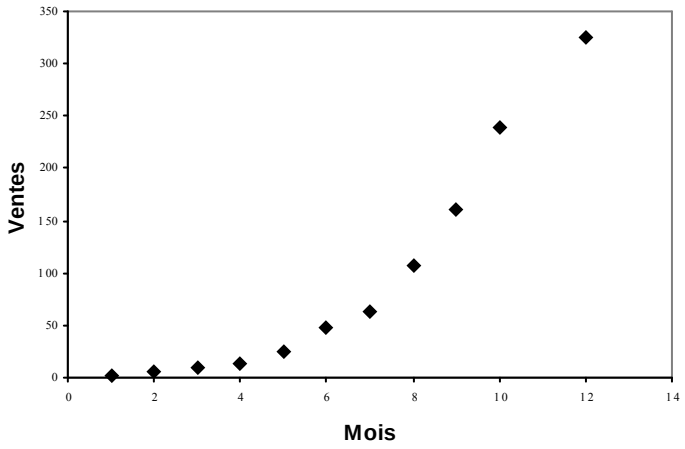
On retrouve les constantes $b_0$ et $b_1$ grâce à l'exponentiel :

$$a_0 = e^{a'_0} \qquad \text{et} \qquad b_0 = e^{b'_0}$$

Exemple 9 : **Le tableau suivant indique l'évolution des ventes d'un produit pour les 12 premiers mois de son lancement :**

| **Mois : $x_i$** | **Ventes : $y_i$** |
|---|---|
| 1 | 1 |
| 2 | 6 |
| 3 | 10 |
| 4 | 14 |
| 5 | 25 |
| 6 | 48 |
| 7 | 63 |
| 8 | 108 |
| 9 | 161 |
| 10 | 240 |
| 12 | 325 |

**Diagramme de dispersion**




Le nuage de points du diagramme de dispersion indique que la relation entre le temps et les ventes n'est pas linéaire, mais exponentielle.

On ajuste une courbe exponentielle d'équation :

$$y = a_0 \times b_0^{x}$$

Grâce à une transformation logarithmique, le modèle devient linéaire :

$$\text{Log}(y) = \text{Log}\,(a_0 \times b_0^{x})$$

$$\text{Log}(y) = \text{Log}\,(a_0) + \text{Log}(b_0) \times x$$

On pose :

$y' = Log(y)$, $a'_0 = Log(b_0)$ et $b'_0 = Log(b_0)$

Le modèle devient : $y' = a'_0 + b'_0 \times x$

| | $x_i$ | $y_i$ | $y'_i$ | $x_i^2$ | $y'^2_i$ | $x_i\, y'_i$ |
|---|---|---|---|---|---|---|
| | 1 | 1 | 0,000 | 1 | 0,000 | 0,000 |
| | 2 | 6 | 1,792 | 4 | 3,210 | 3,584 |
| | 3 | 10 | 2,303 | 9 | 5,302 | 6,908 |
| | 4 | 14 | 2,639 | 16 | 6,965 | 10,556 |
| | 5 | 25 | 3,219 | 25 | 10,361 | 16,094 |
| | 6 | 48 | 3,871 | 36 | 14,986 | 23,227 |
| | 7 | 63 | 4,143 | 49 | 17,166 | 29,002 |
| | 8 | 108 | 4,682 | 64 | 21,922 | 37,457 |
| | 9 | 161 | 5,081 | 81 | 25,821 | 45,733 |
| | 10 | 240 | 5,481 | 100 | 30,037 | 54,806 |
| | 12 | 325 | 5,784 | 144 | 33,453 | 69,406 |
| **Total** | **67** | | **38,995** | **529** | **169,223** | **296,773** |

$\sum_{i=1}^{12} x_i = 67$ $\sum_{i=1}^{12} y'_i = 38,995$

$\sum_{i=1}^{12} x_i^2 = 529$ $\sum_{i=1}^{12} y'^2_i = 169,223$

$\sum_{i=1}^{12} x_i y'_i = 296,773$

$$\bar{x} = \frac{\sum_{i=1}^{12} x_i}{n} = \frac{67}{12} = 5,58$$

$$\bar{y'} = \frac{\sum_{i=1}^{12} y_i}{n} = \frac{38,995}{12} = 3,250$$

$$COV(x, y') = \frac{\sum_{i=1}^{12} x_i y'_i}{n} - \bar{x}\,\bar{y'} = \frac{296,773}{12} - 5,58 \times 3,25 = 6,596$$

On détermine $b_0'$ et $b_1'$ par les formules de la régression linéaire.

$$a'_0 = \frac{\sum x_i y_i' - n \bar{x} \bar{y}'}{\sum x_i^2 - n \bar{x}^2} = \frac{296{,}773 - 12 \times 5{,}58 \times 3{,}25}{529 - 12 \times 5{,}58^2} = 0{,}509$$

$$b'_0 = \bar{Y}' - b_1' \bar{X} = 3{,}25 - 0{,}509 \times 5{,}58 = 0{,}410$$

On retrouve les constantes $b_0$ et $b_1$ grâce à l'exponentiel :

$$a_0 = e^{a'_0} = e^{0{,}509} = 1{,}66$$
$$b_0 = e^{b'_0} = e^{0{,}410} = 1{,}51$$

L'équation du modèle est donc :

$y = 1{,}66 \times 151^x$

### 4.5.2. Modèle polynomial.

Nous nous contenterons d'étudier, à ce niveau, le modèle polynomial simple et nous laisserons le cas du modèle polynomial général, lorsque nous aborderons la régression multiple.

Le modèle polynomial simple a pour équation générale :

$$y = a_0 \times x^{bo}$$

Grâce à une transformation logarithmique, le modèle devient linéaire :

$\text{Log}(y) = \text{Log}(a_0 \times x^{bo})$

$\text{Log}(y) = \text{Log}(a_0) + b_0 \text{Log}(x)$

On pose :

$y' = \text{Log}(y)$, $a'_0 = \text{Log}(a_0)$ et $x' = \text{Log}(x)$

Le modèle devient : $y' = a'_0 + b_0 \times x'$

On détermine $a'_0$ et $b'_0$ par les formules générales de la régression linéaire.

$$a'_0 = = \frac{\sum x'_i y'_i - \overline{x'}\,\overline{y'}}{\sum x'^2_i - n\,\overline{x'}^2} \qquad \text{et} \qquad b_0 = \overline{y'} - b_1' \overline{x'}$$

On retrouve les constantes $a_0$ grâce à l'exponentiel : $a_0 = e^{a'_0}$

Après la détermination de $a_0$ et $b_0$, le modèle se trouve entièrement déterminé.

**Remarque** : dans le cas du modèle polynomial du second degré qui a comme équation générale : $y = a\,x^2 + b\,x + c$, on peut montrer, par un simple changement de variables, qu'on peut revenir au modèle polynomial simple du paragraphe 2.5.1., en effet, si l'on pose :

$x = X - X_0$ et $y = Y - Y_0$

Le modèle devient :

$Y - Y_0 = a\,(X - X_0)^2 + b\,(X - X_0) + c$

$Y = a\,X^2 + b\,(1 - 2aX_0)\,X + a\,X^2_0 - b\,X_0 + Y_0 + c$

Il suffit de prendre :

$X_0 = 1/2a$ et $Y_0 = -\,a\,X^2_0 + b\,X_0 - c = -\,a/4 + b/2a - c$

Le modèle devient : $Y = a\,X^2$ qui est le modèle polynomial simple.

### 4.6. REGRESSION MULTIPLE.

La régression multiple a pour but d'expliquer les variations d'une variable dépendante y et p variables explicatives $x_1$ , $x_2$ , ..., $x_p$ ($p > 1$), ensuite, si cette relation est confirmée d'évaluer son intensité.

L'utilisation de plusieurs variables indépendantes, permet d'améliorer le pourcentage de variation expliquée, c'est à dire augmenter le coefficient de détermination $R^2$ , qui reflète la qualité de l'ajustement. Ce qui implique une réduction de la variance résiduelle, $S^2_e$ , ce qui a pour effet d'augmenter la précision des estimations de y.

### 4.6.1. Identification du modèle.

Le modèle théorique en régression linéaire multiple s'écrit :

$y_i = a_1x_{1i} + a_2x_{2i} + a_3x_{3i} + .... + a_px_{pi} + b_0 + \varepsilon_i$

Les paramètre a**i** sont appelés coefficients de régression partielle, ils mesurent la variation de y lorsque $x_i$ augmente d'une unité et que les autres variables explicatives sont maintenues constantes.

$\varepsilon_i$ représente l'erreur aléatoire, elle est non observable et comprend à la fois les erreurs de mesure sur les valeurs observées de $y_i$ et tous les autres facteurs explicatifs non pris en compte dans le modèle.

L'analyse de régression repose sur les mêmes hypothèses présentées dans la régression simple auxquels il faut ajouter qu'il n'y a pas de colinéarité parfaite entre les variables explicatives $x_i$, c'est-à-dire que leurs coefficients de corrélation linéaire doivent être nuls ou proches de zéro.

### 4.6.2. Ajustement du modèle.

De la même manière que la régression simple, la méthode des moindres carrés consiste à minimiser la somme des carrés des différences entre les valeurs observées, $y_i$, et les valeurs estimées par le modèle, $\hat{y}_i$ différence appelée résidu.

Le modèle empirique, estimé à partir des observations, sera désigné de cette façon :

$$\hat{y}_i = a'_1 x_{1i} + a'_2 x_{2i} + a'_3 x_{3i} + ... + a'_p x_{pi} + b'_0$$

pour : (i=1, 2, …, n)

$a'_1$ , $a'_2$ … et $a'_p$ ainsi que $b'_0$ sont des estimations des paramètres $a_1$ , $a_2$ … et $a_p$ ainsi que $b_0$ du modèle théorique.

On définit le i-ème résidu $e_i$ par : $e_i = Y_i - \hat{Y}_i$

La méthode des moindres carrés minimise la somme des carrés des résidus, somme désignée par $f(a_1, a_2, \ldots a_p, b_0)$, une fonction de ( p + 1 ) inconnues :

$$f(a_1, a_2, ..., a_p, b_0) = \sum e_i^2 = \sum \left( y_i - \hat{y}_i \right)^2 = \sum \left( y_i - a_1 x_{1i} - ... - a_p x_{pi} - b_0 \right)^2$$

En annulant simultanément les dérivées partielles par rapport à $a_1$, $a_2$, . . .$a_p$ et $b_0$, on obtient un système de ( p + 1 ) équations linéaires homogène à (p+1) inconnues qui sont justement $a_1$, $a_2$, . . .$a_p$ et $b_0$. Ce système est semblable à celui montré dans le cas de la régression linéaire simple.

Dans le cas de la régression multiple, les calculs deviennent très complexes, et pratiquement impossibles à faire sans l'aide de l'ordinateur. Il existe un nombre important de logiciels informatiques qui traitent le problème de la régression simple et de la régression multiple. Les logiciels fournissent en plus des estimations des coefficients du modèle, toutes les statistiques et tests nécessaires pour juger de la validité du modèle.

Nous allons, dans ce qui suit, étudier, dans les détails, le cas de la régression linéaire simple à deux variables explicatives.

### 4.6.3. Régression linéaire à 2 variables explicatives.

La formule générale du modèle est : $y = a_1 x_1 + a_2 x_2 + b$

La méthode des moindres carrés est celle qui minimise la somme des carrés des résidus; symboliquement, on cherche à :

$$\text{Minimiser l'expression} : \sum_{i=1}^{n} \left( y_i - \hat{y}_i \right)^2 = \sum_{i=1}^{n} e_i^{\,2}$$

De même que pour la régression simple, avec le critère des moindres carrés, tous les résidus deviennent positifs; car sinon, en nous limitant aux résidus simples, il est impossible que des résidus positifs annulent des résidus négatifs.

Les démonstrations algébriques sont facilitées par le recours aux outils du calcul différentiel. La minimisation d'une fonction quadratique à plusieurs variables s'effectue en annulant les dérivées partielles de premier ordre et en vérifiant que les signes des dérivées partielles de deuxième ordre sont tous positifs.

#### 4.6.3.1. Calcul des coefficients.

Par calcul différentiel, on cherche les valeurs $a_1$ $a_2$ et $b_0$ qui minimisent la somme des carrés des résidus, cette somme quadratique est notée $f(a_1,a_2,b_0)$, puisqu'elle est fonction des 3 termes inconnues : les 2 termes $a_1$ et $a_2$ et le 3è terme $b_0$ :

$$f(a_1, a_2, b_0) = \sum e_i^{\,2} = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - a_1 x_{1i} + a_2 x_{2i} - b_0)^2$$

$f(a_1,a_2,b_0)$ est minimum lorsque les dérivées premières partielles de $f(a_1, a_2, b_0)$ par rapport à $a_1$ , $a_2$, et à $b_0$ sont nulles et que les dérivées secondes partielles sont toutes positives.

Convenons de garder les mêmes notations pour $a_i$ et sont estimation $a_{0i}$ pour simplifier les écritures.

Appelons :

- $f'_{a_0}$ , la dérivée première partielle de f par rapport à $a_0$;
- $f''_{a_0}$ , la dérivée seconde partielle de f par rapport à $a_0$.

Les 2 conditions seront vérifiées si :

- $f'_{a_1} = 0$ , $f'_{a_2} = 0$ , et $f'_{b_0} = 0$ ;
- $f''_{a_0} > 0$ , $f''_{a_2} > 0$ et $f''_{b_0} > 0$.

**1ère Condition** : écrivons que les dérivées premières partielles sont nulles, c'est-à-dire que : $f'_{a_0} = 0,\ f'_{a_2} = 0$ et $f'_{b_0} = 0$.

$$f(a_1,a_2,b_0) = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - a_1x_{1i} - a_2x_{2i} - b_0)^2$$

On a :

$$f'_{b_0} = \sum -(y_i - a_1x_{1i} - a_2x_{2i} - b_0) = 0$$
$$\sum (y_i - a_1x_{1i} - a_2x_{2i} - b_0) = 0$$
$$\sum y_i - n\,b_0 - a_1\sum x_{1i} - a_2\sum x_{2i} = 0$$
$$\sum y_i = n\,b_0 + a_1\sum x_{1i} + a_2\sum x_{2i}$$

On a aussi :

$$f'_{a_1} = \sum -2\,x_{1i}(y_i - a_1x_{1i} - a_2x_{2i} - b_0) = 0$$
$$\sum (x_{1i}y_i - b_0\,x_{1i} - a_1x_{1i}^2 - a_2x_{1i}x_{2i}) = 0$$
$$\sum x_{1i}y_i - b_0\sum x_{1i} - a_1\sum x_{1i}^2 - a_2\sum x_{1i}x_{2i} = 0$$
$$\sum x_{1i}y_i = b_0\sum x_{1i} + a_1\sum x_{1i}^2 + a_2\sum x_{1i}x_{2i}$$

On a enfin :

$$f'_{a_2} = \sum -2x_{2i}(y_i - a_1x_{1i} - a_2x_{2i} - b_0) = 0$$
$$\sum(x_{2i}y_i - b_0 x_{2i} - a_1x_{1i}^2 - a_2x_{1i}x_{2i}) = 0$$
$$\sum x_{2i}y_i - b_0 \sum x_{2i} - a_1\sum x_{1i}x_{2i} - a_2\sum x_{2i}^2 = 0$$
$$\sum x_{2i}y_i = b_0 \sum x_{2i} + a_1\sum x_{1i}x_{2i} + a_2\sum x_{2i}^2$$

On a donc un système de trois équations à trois inconnues, ces deux équations qui sont appelées équations normales sont :

$$\sum y_i = n\,b_0 + a_1\sum x_{1i} + a_2\sum x_{2i}$$
$$\sum x_{1i}y_i = b_0 \sum x_{1i} + a_1\sum x_{1i}^2 + a_2\sum x_{1i}x_{2i}$$
$$\sum x_{2i}y_i = b_0 \sum x_{2i} + a_1\sum x_{1i}x_{2i} + a_2\sum x_{2i}^2$$

**Calcul de $b_0$** : En considérant la dernière équation, on a successivement les égalités suivantes :

$$\sum y_i = n\,b_0 + a_1\sum x_{1i} + a_2\sum x_{2i}$$
$$b_0 = \frac{\sum y_i}{n} - a_1\frac{\sum x_{1i}}{n} - a_2\frac{\sum x_{2i}}{n}$$
$$b_0 = \bar{y} - a_1\overline{x_1} - a_2\overline{x_2}$$

**Calcul de $a_1$ et de $a_2$** : En considérant les deux premières équations, on a :

$$\sum x_{1i}y_i = b_0 \sum x_{1i} + a_1\sum x_{1i}^2 + a_2\sum x_{1i}x_{2i}$$
$$\sum x_{2i}y_i = b_0 \sum x_{2i} + a_1\sum x_{1i}x_{2i} + a_2\sum x_{2i}^2$$

Nous remplaçons $b_0$ par sa valeur : $b_0 = \bar{y} - a_1\overline{x_1} - a_2\overline{x_2}$ et nous divisons les deux membres des deux égalités par n. On trouve successivement :

$$\sum x_{1i}y_i = (\bar{y} - a_1\overline{x_1} - a_2\overline{x_2})\sum x_{1i} + a_1\sum x_{1i}^2 + a_2\sum x_{1i}x_{2i}$$
$$\sum x_{2i}y_i = (\bar{y} - a_1\overline{x_1} - a_2\overline{x_2})\sum x_{2i} + a_1\sum x_{1i}x_{2i} + a_2\sum x_{2i}^2$$

Qui deviennent après remplacement de $b_0$ :

$$\overline{x_1y} = (\bar{y} - a_1\overline{x_1} - a_2\overline{x_2})\overline{x_1} + a_1\overline{x_1^2} + a_2\overline{x_1x_2}$$

$$\overline{x_2 y} = (\bar{y} - a_1 \overline{x_1} - a_2 \overline{x_2})\overline{x_2} + a_1 \overline{x_1 x_2} + a_2 \overline{x_2^2}$$

Soit, en utilisant les notations de S² et COV :

$$S^2_{x_1} a_1 + COV(x_1, x_2) = COV(x_1, y)$$
$$COV(x_1, x_2) a_1 + S^2_{x_2} a_2 = COV(x_2, y)$$

Pour résoudre ce système de deux équations à deux inconnues $a_1$ et $a_2$ on procède par addition. Ainsi, pour calculer $a_1$, on multiplie les 2 membres de la 1ère équation par $S^2_{x_2}$ et ceux de la 2è équation par $COV(x_1, x_2)$ et on soustrait, après, membre à membre, la 2è équation de la 1è. De même, pour calculer $a_2$, on multiplie les 2 membres de la 1è équations par $COV(x_1, x_2)$ et ceux de la 2è équations par $S^2_{x_1}$ et on soustrait, après, membre à membre la 1ère équation de la 2è. On trouve alors les résultats suivants :

$$a_1 = \frac{S^2_{x_2} COV(x_1, y) - COV(x_1, x_2) COV(x_2, y)}{S^2_{x_1} S^2_{x_2} - COV(x_1, x_2)^2}$$

$$a_2 = \frac{S^2_{x_1} COV(x_2, y) - COV(x_1, x_2) COV(x_1, y)}{S^2_{x_1} S^2_{x_2} - COV(x_1, x_2)^2}$$

La régression de y en fonction de $x_1$ et de $x_2$, selon la méthode des moindres carrés, est l'équation :

$$\hat{y} = a_1 x_1 + a_2 x_2 + b_0$$

L'estimation de $a_1$, de $a_2$ et de b par la méthode des moindres carrés conduit aux formules suivantes :

$$b_0 = \bar{y} - a_1 \overline{x_1} - a_2 \overline{x_2}$$

$$a_1 = \frac{S^2_{x_2} COV(x_1, y) - COV(x_1, x_2) COV(x_2, y)}{S^2_{x_1} S^2_{x_2} - COV(x_1, x_2)^2}$$

$$a_2 = \frac{S_{x_1}^2 COV(x_2, y) - COV(x_1, x_2)COV(x_1, y)}{S_{x_1}^2 S_{x_2}^2 - COV(x_1, x_2)^2}$$

d'où : $\hat{y} = a_1 x_1 + a_2 x_2 + b_0 = a_1 x_1 + a_2 x_2 + \left(\bar{y} - a_1 \overline{x_1} - a_2 \overline{x_2}\right)$

$$\hat{y} = a_1 (x_1 - \overline{x_1}) + a_2 (x_2 - \overline{x_2}) + \bar{y}$$

Ces estimateurs sont des fonctions linéaires des observations $x_{1i}, x_{2i}$ et $y_i$.

**2è Condition** : montrons que les dérivées secondes partielles sont positives, c'est-à-dire que : $f''_{a_1} > 0 \quad f''_{a_2} > 0$ et $f''_{b_0} > 0$.

$f'_{a_1} = \sum -2 x_{1i}(y_i - a_1 x_{1i} - a_2 x_{2i} - b_0)$

$f''_{a_1} = 2\sum x_{1i}^2$ ce qui est bien positif.

$f'_{a_2} = \sum -2 x_{2i}(y_i - a_1 x_{1i} - a_2 x_{2i} - b_0)$

$f''_{a_2} = 2\sum x_{2i}^2$ ce qui est bien positif.

$f'_{b_0} = \sum -(y_i - a_1 x_{1i} - a_2 x_{2i} - b_0)$

$f''_{b_0} = 1$ ce qui est bien positif.

Nous pouvons donc conclure que les valeurs de $a_1$ de $a_2$ et $b_0$ que nous avons déterminées correspondent bien à un minimum de l'expression : $x^{p-1}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{n} e_i^2$

**Remarque** : Dans le cas de la régression polynomiale générale qui a pour forme : $y = a_p x^p + a_{p-1} x^{p-1} + \ldots + a_2 x^2 + a_1 x + a_0$, il suffit de remplacer $x^k$ par $x_k$ pour revenir au modèle multilinéaire qu'on vient d'étudier.

L'équation du modèle devient, après le changement de variables :
$y = a_p x_p + a_{p-1} x_{p-1} + \ldots + a_2 x_2 + a_1 x_1 + b_0$

N'oublions pas que, dans un modèle multilinéaire, il est nécessaire que les variables $x_k$ soient indépendantes pour justifier le recours à plusieurs variables, mais cela n'est pas tout à

fait le cas, dans notre modèle polynomial général transformé en modèle multilinéaire car les $x^k$ ne sont pas indépendantes, en effet prenons le cas simple de x et $x^2$ et montrons, par un exemple, que ces deux variables ne sont pas indépendantes :

| | $x_1$ | $x^2 = x_2$ | $x_1^2$ | $x_2^2$ | $x_1x_2$ |
|---|---|---|---|---|---|
| | 1,1 | 1,21 | 1,21 | 1,4641 | 1,331 |
| | 1,3 | 1,69 | 1,69 | 2,8561 | 2,197 |
| | 1,7 | 2,89 | 2,89 | 8,3521 | 4,913 |
| | 1,8 | 3,24 | 3,24 | 10,4976 | 5,832 |
| | 2,4 | 5,76 | 5,76 | 33,1776 | 13,824 |
| | 3,2 | 10,24 | 10,24 | 104,8576 | 32,768 |
| | 3,5 | 12,25 | 12,25 | 150,0625 | 42,875 |
| | 3,9 | 15,21 | 15,21 | 231,3441 | 59,319 |
| | 4,2 | 17,64 | 17,64 | 311,1696 | 74,088 |
| | 4,7 | 22,09 | 22,09 | 487,9681 | 103,82 |
| **Somme** | **27,8** | **92,22** | **92,22** | **1341,749** | **340,97** |
| **Sommes/10** | **2,78** | **9,222** | **9,222** | **134,1749** | **34,097** |

On calcule les variances, les écarts types des variables et leur covariance :

$V(x_1) = 1,4936$ => $Sx_1 = 1,222129$

$V(x^2) = 49,129656$ => $Sx^2 = 7,009255$

$COV(x_1,x^2) = 8,45984$

$R(x_1,x^2) = 8,45984/(1,222129x7,009255) = 0,988$

Le même calcul pourra montrer que les $x^k$ et $x^l$ ne sont pas indépendantes quels que soient k et l mais nous admettons, dans une 1ère approximation, la validité du modèle malgré cette entorse à l'hypothèse d'indépendance des variables.

**Exemple 10 :** Le tableau suivant regroupe les données relatives à une variable dépendante y et 2 variables explicatives $x_1$ et $x_2$.

| $x_1$ | $x_2$ | y |
|---|---|---|
| 15 | 20 | 90 |
| 28 | 15 | 115 |
| 40 | 10 | 120 |
| 70 | 9 | 100 |
| 120 | 11 | 130 |
| 130 | 8 | 118 |
| 160 | 4 | 98 |
| 250 | 7 | 135 |

Le modèle empirique, estimé à partir des observations, sera désigné de cette façon : $y = a_1 x_1 + a_2 x_2 + b$.

| | $x_1$ | $x_2$ | y | $X^2_1$ | $X^2_2$ | $y^2$ | $x_1x_2$ | $x_1 y$ | $x_2y$ |
|---|---|---|---|---|---|---|---|---|---|
| | 15 | 20 | 90 | 225 | 400 | 8100 | 300 | 1350 | 1800 |
| | 28 | 15 | 115 | 784 | 225 | 13225 | 420 | 3220 | 1725 |
| | 40 | 10 | 120 | 1600 | 100 | 14400 | 400 | 4800 | 1200 |
| | 70 | 9 | 100 | 4900 | 81 | 10000 | 630 | 7000 | 900 |
| | 120 | 11 | 130 | 14400 | 121 | 16900 | 1320 | 15600 | 1430 |
| | 130 | 8 | 118 | 16900 | 64 | 13924 | 1040 | 15340 | 944 |
| | 160 | 4 | 98 | 25600 | 16 | 9604 | 640 | 15680 | 392 |
| | 250 | 7 | 135 | 62500 | 49 | 18225 | 1750 | 33750 | 945 |
| **Total** | **813** | **84** | **906** | **126909** | **1056** | **104378** | **6500** | **96740** | **9336** |

Les moyennes :

$\bar{x}_1 = 101{,}625$ $\bar{x}_2 = 10{,}5$ et $\bar{y} = 113{,}25$

Les variances :

$V(x_1) = 5535{,}984 V(x_2) = 21{,}75$ et $V(Y) = 221{,}688$

Les covariances :

$COV(x_1,x_2) = -254{,}563$ , $COV(x_1,y) = 583{,}469$
et $COV(x_2;y) = -22{,}125$

On calcule les coefficients $a_1$ et $a_2$ par les formules déjà trouvées :

$$a_1 = \frac{S_{x_2}^2 COV(x_1, y) - COV(x_1, x_2)COV(x_2, y)}{S_{x_1}^2 S_{x_2}^2 - COV(x_1, x_2)^2}$$

$$= \frac{21{,}75 \text{x} 583{,}469 - 254{,}563 \text{x} 22{,}125}{5535{,}984 \text{x} 21{,}75 - 254{,}563 \text{x} 254{,}563} = 0{,}1269$$

$$a_2 = \frac{S_{x_1}^2 COV(x_2, y) - COV(x_1, x_2)COV(x_1, y)}{S_{x_1}^2 S_{x_2}^2 - COV(x_1, x_2)^2}$$

$$= \frac{-5535{,}984 \text{x} 22{,}125 + 254{,}563 \text{x} 583{,}469}{5535{,}984 \text{x} 21{,}75 - 254{,}563 \text{x} 254{,}563} = 0{,}4684$$

$$b_0 = \bar{y} - a_1 \overline{x_1} - a_2 \overline{x_2}$$

$$= 113{,}25 - 0{,}1269 \text{x} 101{,}625 - 0{,}4684 \text{x} 10{,}5 = 95{,}4321$$

Le modèle linéaire de régression multiple est donc :

$y = 0{,}1269\ x_1 + 0{,}4684\ x_2 + 95{,}4321$

### 4.6.4. Qualité de l'ajustement.

#### 4.6.4.1. Coefficient de corrélation.

Dans le cas de la régression multiple, on parle de coefficient de corrélation multiple, il mesure la corrélation combinée de toutes les variables du modèle. Les valeurs du coefficient de corrélation s'interprètent de la même manière que pour la régression simple.

#### 4.6.4.2. Coefficient de détermination multiple.

De la même manière que pour la régression simple, le coefficient de détermination indique le pourcentage de la variation totale de y autour de sa moyenne qui est expliquée par la régression.

La variation totale $\sum (y_i - \bar{y})^2$ observée sur la variable expliquée y peut être décomposée en 2 parties :

$$\sum (y_i - Yy)^2 = \sum \left( \hat{y}_i - \bar{y} \right)^2 + \sum \left( y_i - \hat{y}_i \right)^2$$

Le premier terme $\sum\left(\hat{y}_i - \bar{y}\right)^2$ désigné par SCR mesure la variation autour du modèle de régression, on l'appelle Somme des Carrés due à la Régression. L'autre terme, $\sum\left(y_i - \hat{y}_i\right)^2$ désigné par SCE, mesure la variation résiduelle, on l'appelle la somme des carrés due à l'erreur.

La somme des carrées totale s'écrit :

SCT = SCR + SCE

Le coefficient de détermination multiple $R^2$ est définit par :

$$R^2 = \frac{SCR}{SCT} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(yi - \bar{y})^2}$$

On pourrait montrer par ailleurs que $R^2$ est égal au carré du coefficient de corrélation multiple.

Le coefficient de détermination multiple ne peut être inférieur au plus élevé des coefficients de détermination simple entre y et chacune des variables explicatives. Si les variables explicatives sont parfaitement indépendantes entre elles, le coefficient de détermination multiple sera égal à la somme des coefficients de détermination simple entre y et chacune des variables explicatives.

Le coefficient de détermination multiple tend à augmenter avec le nombre de variables explicatives. Pour pallier cet inconvénient, on calcule un coefficient de détermination ajusté $R^2_{aj}$ qui tient compte du nombre de variables explicatives (p) et de la taille de l'échantillon (n).

Le coefficient de détermination ajusté se calcule en terme de variances, il est définit par :

$$R^2_{aj} = 1 - \frac{S^2_e}{S^2_y} \text{ avec } S^2_e = \frac{SCE}{n-p-1} \text{ variance due à l'erreur}$$

$$S^2_y = \frac{SCT}{n-1} \text{ variance de y}$$

$$R^2_{aj} = 1 - \frac{SCE/n-p-1}{SCT/n-1} = 1 - \frac{SCE}{SCT} x \frac{n-1}{n-p-1}$$

$$R^2_{aj} = 1 - \frac{n-1}{n-p-1}(1-R^2)$$

Le $R^2$ ajusté est inférieur au $R^2$. Ce dernier est un estimateur biaisé, tandis que le premier est non biaisé.

Le $R^2$ ajusté est préférable à $R^2$ si la taille de l'échantillon est faible. Quand n sera supérieur à 30, il n'y aura habituellement pas beaucoup de différence entre les 2 indices.

Le $R^2$ ajusté est plus approprié pour comparer des modèles de régression d'une variable expliquée Y en fonction de différents sous-groupes de variables explicatives.

**Exemple 11 :** Reprenons les données de l'exemple 10 et calculons le coefficient de détermination.

Le modèle de régression linéaire multiple explique 29,83 % des variations de Y.

## 4.7. EXERCICES D'APPLICATION.

### 4.7.1. Exercice.

L'entreprise SATEX désire contrôler sa consommation d'énergie électrique, pour ce faire, elle dresse le tableau des statistiques de consommation et de production des 10 derniers mois et essaie, dans un premier temps de voir si la consommation dépend de la production.

Le tableau des statistiques est le suivant :

| **Productions $x_i$ (kg)** | **Consommation électrique $y_i$ (kwh)** |
|---|---|
| 125 | 4650 |
| 135 | 5010 |
| 154 | 5800 |
| 162 | 6000 |
| 175 | 6500 |
| 183 | 7000 |
| 195 | 7100 |
| 220 | 8000 |
| 235 | 8500 |
| 257 | 9500 |

a) Tracer le nuage de points ($x_i$ , $y_i$) et dire si cela inspire l'existence d'une liaison entre y et x. Donner une justification de cette liaison.

b) Déterminer s'il y a une corrélation entre consommation électrique et production et si oui établir la relation liant ces deux variables.
c) Interpréter la valeur de b, coordonnée à l'origine du modèle linéaire, c'est-à-dire au point d'abscisse $x_i = 0$.
d) Donner quelle serait la consommation énergétique pour une production de 300 kg.

**Solution** : a) Facile à faire ; b) R = 0,998 avec a = 35,6
et b = 245,3 ; c) sans aucune production, on consomme 245,3 kwh d'électricité ; d) 10925,3 kwh.

### 4.7.2. Exercice.

Une teinturerie consomme beaucoup d'eau, cette consommation est naturellement fonction du poids des tissus teints. Le tableau des consommations d'eau et des poids des tissus teints est donné ci-dessous.

| **Tissus teints $x_i$ (kg)** | **Consommation d'eau $y_i$ ($m^3$)** |
|---|---|
| 24 | 10 |
| 26 | 10,2 |
| 28 | 11 |
| 30 | 11,5 |
| 32 | 12 |
| 34 | 12,6 |
| 36 | 12,9 |
| 38 | 13 |
| 40 | 13,6 |
| 42 | 14,3 |

a) Tracer le nuage de points $(x_i , y_i)$ et dire si cela inspire l'existence d'une liaison entre y et x. Donner une justification de cette liaison.
b) Déterminer s'il y a une corrélation entre consommation d'eau et poids des tissus teints et si oui établir la relation liant ces deux variables.
c) Interpréter la valeur de la coordonnée à l'origine du modèle linéaire, c'est-à-dire au point d'abscisse $x_i = 0$.
d) Donner quelle serait la consommation d'eau pour une production de 50 kg de tissus teints.

**Solution** : a) Facile à faire ; b) R = 0,992 avec a = 0,2 et b = 4,4
c) sans teindre de tissu, on consomme 4,4 $m^3$ d'eau ; d) 14,4 $m^3$ d'eau.

### 4.7.3. Exercice.

Un commerçant désire savoir si son chiffre d'affaires d'une journée est fonction du nombre de

clients qu'il reçoit pendant cette journée. Il dresse le tableau statistique de ses chiffres d'affaires et du nombre de clients qu'il reçoit pendant les 10 derniers jours.

| **Nombre de clients** $x_i$ | **Chiffres d'affaires** $y_i$ **(DH)** |
|---|---|
| 12 | 190 |
| 13 | 230 |
| 15 | 280 |
| 18 | 300 |
| 22 | 310 |
| 23 | 400 |
| 26 | 420 |
| 31 | 480 |
| 32 | 540 |
| 37 | 620 |

a) Tracer le nuage de points ($x_i$ , $y_i$) et dire si cela inspire l'existence d'une liaison entre y et x. Donner une justification de cette liaison.
b) Déterminer s'il y a une corrélation entre nombre de clients et chiffres d'affaires et si oui établir la relation liant ces deux variables.
c) Interpréter la valeur de b, coordonnée à l'origine du modèle linéaire, c'est-à-dire au point d'abscisse $x_i = 0$
d) Donner quel serait le chiffre d'affaires pour 50 clients.

**Solution** : a) Facile à faire ; b) R = 0,983 avec a = 16,0
et b = 10,5 ; c) sans aucun client, on peut réaliser 10,5 DH de chiffre d'affaires, ce qui semble difficile à croire. Il s'agit d'un résultat aberrant ; d) 810,50 DH.

### 4.7.4. Exercice.

Le directeur d'une filature de nylon désire connaître la relation liant la consommation énergétique de son usine avec la production de fil total et de fil teint. Pour ce faire, il dresse le tableau de huit jours de production et classe ce tableau par ordre croissant. Etablir s'il y a :
a) une corrélation entre consommation d'électricité et production totale de fil et si oui, établir la relation liant ces deux variables.
b) une corrélation entre consommation d'électricité et production totale de fil teint et si oui, établir la relation liant ces deux variables.
c) une corrélation entre consommation d'électricité, production totale de fil et production de fil teint ; et, si oui, établir la relation liant ces deux variables
d) Interpréter la valeur de la coordonnée à l'origine du modèle linéaire, c'est-à-dire au point d'abscisses $x_{1i} = x_{2i} = 0$.
e) Compte tenu des résultats des questions a), b) et c) calculer de 3 façons différentes la consommation électrique pour une production globale de fil de 400kg et une production de fil teint de seulement 300 kg ? Interpréter chacun des 3 résultats et dire lequel choisir et pourquoi ?

On donne le tableau des relevés de la consommation d'électricité, de production totale de fil et de la production de fil teint

| $x_1$ (kg) | $x_2$ (kg) | $y_i$ (kwh) |
|---|---|---|
| 250 | 125 | 1510 |
| 275 | 130 | 1635 |
| 281 | 237 | 2400 |
| 292 | 246 | 2520 |
| 307 | 265 | 2730 |
| 314 | 268 | 2780 |
| 340 | 271 | 2800 |
| 355 | 272 | 2840 |

**Solution** : a) R = 0,853 avec a = 13,2 et b = -1568,6 ;
b) R = 0,996 avec a = 8,5 et b = 472,6 ; c) $a_1$ = 2,12 a2 = 7,56 et b = 48,79 et R = 0,999 ; d) Si l'usine ne produit pas de fil total et de fil teint, elle peut s'attendre à une consommation énergétique de 48,79 kwh ;
e) Selon le premier modèle : Y = 13,2 x 400 – 1568,6 = 3711,4 kwh avec R = 0,853
Selon le deuxième modèle : Y = 8,5 x 300 – 472,6 = 2077,4 kwh avec R = 0,996
Selon le troisième modèle : y = 2,12 x 400 + 7,56 x 300 + 48,79 = 3164,79 kwh avec R = 0,999
On choisit le troisième résultat puisque ce modèle a le plus grand coefficient de corrélation.

### 4.7.5. Exercice.

La production céréalière, en millions de quintaux, d'un pays évolue, avec le temps, comme le montre le tableau donné ci-dessous :

| Années xi | Productions yi |
|---|---|
| 1 | 6 |
| 2 | 6,5 |
| 3 | 7 |
| 4 | 9 |
| 5 | 11 |
| 6 | 15 |
| 7 | 19 |
| 8 | 22 |

a) Tracer le nuage de points ($x_i$ , $y_i$) et dire si cela inspire l'existence d'une liaison entre y et x. Donner une justification de cette liaison.
b) On opte pour un modèle d'ajustement exponentiel. Déterminer s'il y a une corrélation entre la production céréalière du pays et le temps et si oui d'établir la relation liant ces deux variables.
c) Interpréter la valeur de la coordonnée à l'origine du modèle linéaire, c'est-à-dire au point d'abscisse $x_i = 0$.
d) Donner quelle serait la production l'année 10.

**Solution** : a) Facile à faire ; b) R = 0,99 avec a = 4,31
et b = 1,22 ; c) L'ordonnée à l'origine 1,46 indique que si x = 0, c'est à dire l'année 0, ce qui n'a aucun sens, la production serait de 4,31 millions de quintaux ; d) 31,5 millions de quintaux.

### 4.7.6. Exercice.

Les relevés des consommations moyennes d'essence d'un véhicule, au 100 km, ainsi que ceux des vitesses auxquelles ces consommations ont été enregistrées sont donnés dans le tableau ci-dessous :

| **Vitesse en km/h** $v_i$ | **Consommation en l/100 km** $c_i$ |
|---|---|
| 95 | 7,05 |
| 100 | 7,21 |
| 105 | 7,41 |
| 110 | 7,81 |
| 115 | 8,12 |
| 120 | 8,65 |
| 125 | 9,41 |
| 130 | 10,13 |

a) Tracer le nuage de points ($v_i$ , $c_i$) et dire si cela inspire l'existence d'une liaison entre $v_i$ et $c_i$. Donner une justification de cette liaison.
b) On doit choisir entre un modèle d'ajustement exponentiel et un modèle d'ajustement parabolique, pour ce faire, il y a lieu d'abord de faire un changement de variables pour centrer le graphe. On pose donc : $V_i = v_i - 90$ et $C_i = c_i - 7$
Calculer le tableau des nouvelles variables.
c) Calculer les coefficients du modèle exponentiel ainsi que le coefficient de corrélation correspondant.
d) Calculer les coefficients du modèle parabolique ainsi que le coefficient de corrélation correspondant.
e) Choisir le modèle qui s'ajuste le mieux.
f) Donner quelles seraient les consommations pour des vitesses de 50 km/h, 70 km/h et 160 km/h.

**Solution** : a) Facile à faire ; b) Facile à faire ;
c) R = 0,96 avec a = 0,06 et b = 1,11 ; d) a = 0,0021 b = -0,0098 et c = 0,0716 avec R = 0,999 ;
e) Le modèle parabolique a le coefficient de corrélation le plus élevé, c'est donc le modèle qui s'ajuste le mieux ;
f) La consommation est donc $Ci = 0,0021\ Vi^2 - 0,0098\ Vi + 0,0716$
Pour une vitesse de 50 km/h, Ci = 10,91 l/100km ;
Pour une vitesse de 70 km/h, Ci = 8,13 l/100km ;
Pour une vitesse de 160 km/h, Ci = 16,94 l/100km.

### 4.7.7. Exercice.

L'entreprise SATEL désire connaître comment évolue son chiffre d'affaires mensuel en fonction de la publicité qu'elle passe dans les journaux et des prospectus qu'elle distribue dans les boites aux lettres des particuliers.
Les relevés de 10 mois des chiffres d'affaires, des dépenses publicitaires et des dépenses pour les prospectus sont résumés dans le tableau ci-dessous :

| **Dépenses en 1 000 DH** | | **CA en 1 000 DH** |
|---|---|---|
| $p_i$ | $f_i$ | $v_i$ |
| 100,00 | 1,20 | 195,25 |
| 125,00 | 2,10 | 235,65 |
| 130,00 | 3,20 | 241,15 |
| 132,00 | 3,30 | 242,85 |
| 140,00 | 4,20 | 250,55 |
| 152,00 | 4,80 | 265,25 |
| 155,00 | 5,50 | 270,15 |
| 157,00 | 5,70 | 274,55 |
| 159,00 | 5,90 | 275,95 |
| 163,00 | 6,50 | 281,45 |

a) Etablir s'il y a une corrélation entre le chiffre d'affaires mensuel et la publicité que passe l'entreprise dans les journaux.
b) Etablir s'il y a une corrélation entre le chiffre d'affaires mensuel et les prospectus que distribue l'entreprise dans les boites aux lettres.
c) Calculer les éléments du modèle d'ajustement linéaire du chiffre d'affaires mensuel en fonction de la publicité que passe l'entreprise dans les journaux.
d) Calculer les éléments du modèle d'ajustement linéaire du chiffre d'affaires mensuel en fonction de la dépense pour les prospectus que distribue l'entreprise dans les boites aux lettres.
e) Calculer les éléments du modèle d'ajustement linéaire du chiffre d'affaires mensuel en fonction de la dépense de la publicité que passe l'entreprise dans les journaux et de celle des prospectus que distribue l'entreprise dans les boites aux lettres.
f) Interpréter la valeur de la coordonnée aux origines (au point de coordonnées $p_i = 0$ et $f_i =$ 0,00 DH) du modèle linéaire.
g) Indiquer sur quelle variable pi ou fi doit agir le chef d'entreprise pour avoir la meilleure augmentation du chiffre d'affaires.
h) Compte tenu des résultats des questions c), d) et e) calculer, de 3 façons différentes, le chiffre d'affaires pour des dépenses de publicité de 185 735,32 DH et des dépenses de prospectus de 7 245,36 DH ? Indiquer lequel des 3 résultats choisir et dire pourquoi.

**Solution** : a) R = 0,99619 ; b) R = 0,9637 ; c) vi = 1,31 pi + 67,54 ; d) vi = 14,34 fi + 192,48 ;

e) Vi = 1,67 pi – 4,08 fi + 34,79 avec R = 0,998 ; f) Sans dépense de la publicité dans les journaux ni de dépenses dans des prospectus que distribue l'entreprise dans les boites aux lettres, on peut s'attendre en moyenne à un chiffre d'affaires mensuel de 34790 DH ; g) La dépense de la publicité dans les journaux est plus corrélée (0,99619) à la dépense dans des prospectus que distribue l'entreprise dans les boites aux lettres (0,9637). Le chef d'entreprise doit agir sur la dépense de la publicité dans les journaux pour avoir la meilleure augmentation du chiffre d'affaires ;
h) vi = 1,31 pi + 67,54 = 1,31 x 185,73532 + 67,54 = 310,853 soit 310853 DH
vi = 14,34 fi + 192,48 = 14,34 x 7,24536 +192,48 = 296,378 soit 296378 DH
vi = 1,67 pi – 4,08 fi + 34,79 = 1,67 x 185,73532 – 4,08 x 7,24536 + 34,79 = 315,407 soit 315407 DH.
On peut retenir le troisième résultat (315407 DH) puisque ce modèle possède le coefficient de corrélation le plus élevé.

### 4.7.8. Exercice.

La production intérieure brute d'un pays évolue, avec le temps, comme indiqué, dans le tableau, ci-dessous :

| **années** | $x_i$ | **PIB en milliards de DH** $p_i$ |
|---|---|---|
| 1997 | 1 | 2,79 |
| 1998 | 2 | 2,87 |
| 1999 | 3 | 2,95 |
| 2000 | 4 | 3,01 |
| 2001 | 5 | 3,15 |
| 2002 | 6 | 3,25 |
| 2003 | 7 | 3,27 |
| 2004 | 8 | 3,33 |
| 2005 | 9 | 3,45 |

a) Tracer le nuage de points ($x_i$ , $p_i$) et dire si cela inspire l'existence d'une liaison entre $p_i$ et $x_i$. Donner une justification de cette liaison.
b) On doit choisir entre un modèle d'ajustement exponentiel et un modèle d'ajustement parabolique, pour ce faire comparer les coefficients de corrélation des deux modèles.
c) Calculer les coefficients du modèle qui possède le meilleur coefficient de corrélation.
d) Donner quelles seraient les PIB pour les années 2006 et 2007.

**Solution** : a) Facile à faire ; b) $Pi = 2,73 \times 1,03^{xi}$ avec R = 0,992
c) $Pi = -0,001\ xi^2 + 0,093\ xi + 2,69$ avec R = 0,994 ;
d) Pour l'année 2006, xi = 10
$Pi = -0,001 \times 10^2 + 0,093 \times 10 + 2,69 = 3,52$ milliards de DH
Pour l'année 2007, xi = 11
$Pi = -0,001 \times 11^2 + 0,093 \times 11 + 2,69 = 3,592$ milliards de DH

### 4.7.9. Exercice.

Le nombre d'abonnés à un service téléphonique au cours des neuf premiers mois de son lancement sont comme suit :

| Mois | Période t | Nombre d'abonnés y(t) |
|---|---|---|
| Janvier | 1 | 1 |
| Février | 2 | 6 |
| Mars | 3 | 10 |
| Avril | 4 | 14 |
| Mai | 5 | 25 |
| Juin | 6 | 48 |
| Juillet | 7 | 63 |
| Août | 8 | 108 |
| septembre | 9 | 161 |

a) Tracer le nuage de points ($t_i$ , $y_{ti}$) et dire si cela inspire l'existence d'une liaison linéaire ou non linéaire. Donner une justification de cette liaison ;
b) On doit choisir entre un modèle d'ajustement exponentiel et un modèle d'ajustement linéaire, pour ce faire comparer les coefficients de corrélation des deux modèles ;
c) Calculer les coefficients du modèle qui possède le meilleur coefficient de corrélation ;
d) Donner quelles seraient le nombre de nouveaux abonnés pour les trois derniers mois de l'année.

**Solution** : a) Facile à faire ; b) Modèle linéaire R = 0,91 ; Modèle exponentiel R = 0,97 ; Modèle qui possède le meilleur coefficient de corrélation : $Y_{ti} = 1,29 \times 1,76^{ti}$
d) Pour le mois 10 : $Y_{ti} = 1,29 \times 1,76^{10} = 368$ abonnés ;
Pour le mois 11 : $Y_{ti} = 1,29 \times 1,76^{11} = 647$ abonnés ;
Pour le mois 12 : $Y_{ti} = 1,29 \times 1,76^{12} = 1140$ abonnés.

### 4.7.10. Exercice.

Une entreprise agricole dispose de données observées au cours de 10 années successives relatives aux variables suivantes :

y : Rendement d'une culture sous serre.
$x_1$ : Quantité d'eau d'irrigation en mm.
$x_2$ : Température moyenne.

Les données sont les suivantes :

| Année | $x_1$ | $x_2$ | y |
|---|---|---|---|
| 1 | 87,9 | 19,6 | 28,37 |

| | | | |
|---|---|---|---|
| 2 | 89,9 | 15,2 | 23,77 |
| 3 | 153,0 | 19,7 | 26,04 |
| 4 | 132,1 | 17,0 | 25,74 |
| 5 | 88,8 | 18,3 | 26,68 |
| 6 | 220,9 | 17,8 | 24,29 |
| 7 | 117,7 | 17,8 | 28,00 |
| 8 | 109,0 | 18,3 | 28,37 |
| 9 | 156,1 | 17,8 | 24,96 |
| 10 | 181,5 | 16,8 | 21,66 |

A partir de ces données, on cherche le modèle de régression linéaire qui permet d'expliquer au mieux le rendement en fonction des variables météorologiques.

a) Etablir s'il y a une corrélation entre y et $x_1$.
b) Etablir s'il y a une corrélation entre y et $x_2$.
c) Calculer les coefficients du modèle d'ajustement linéaire de y en fonction de $x_1$.
d) Calculer les coefficients du modèle d'ajustement linéaire de y en fonction de $x_2$.
e) Calculer les coefficients du modèle d'ajustement linéaire de y en fonction de $x_1$ et de $x_2$.
f) Calculer et interpréter le coefficient de détermination du modèle d'ajustement linéaire de y en fonction de $x_1$ et de $x_2$.
g) Indiquer sur quelle variable météorologique l'exploitant agricole doit agir pour avoir le meilleur rendement.

**Solution** : a) Corrélation entre y et $x_1$ : 0,52
b) Corrélation entre y et $x_2$ : - 0,30
c) Modèle d'ajustement linéaire de y en fonction de $x_1$ : $y = 0,31\ x1 + 212,12$
d) Modèle d'ajustement linéaire de y en fonction de $x_2$ : $y = - 5,86\ x2 + 357,74$
e) $Y = 0,30\ x1 – 5,60\ x2 + 312,60$ ; f) $R^2 = 0,354$
g) La variable $x_1$ est plus corrélée avec y (0,52) que la variable $x_2$ (-0,30), l'exploitant agricole doit donc agir sur la quantité d'eau d'irrigation pour avoir le meilleur rendement.

# CHAPITRE 5
# LES SERIES CHRONOLOGIQUES

## 5.1. DEFINITION.

Une série chronologique ou temporelle, est une suite d'observations numériques d'une grandeur effectuées à intervalles réguliers au cours du temps.

Les exemples dans le monde économique et social sont donc nombreux : inflation, cours boursiers, chômage, productions, exportations, natalité, immigration, scolarisation, logement, chiffre d'affaires, stocks, ventes, prix, vie d'un produit, clientèle, etc.

Si on note y la grandeur à laquelle se rapportent les observations, une série chronologique est donc une série statistique à deux variables (t , y) dont la seconde variable est le temps t.

La spécificité de l'analyse d'une série chronologique est l'importance accordée à l'ordre dans lequel sont effectuées les observations. En séries chronologiques la dépendance temporelle entre les variables constitue la source principale d'information.

L'échelle de mesure de la grandeur sera toujours représentée par une variable continue à valeurs réelles.

La fréquence des observations peut être journalière, hebdomadaire, mensuelle, trimestrielle, annuelle ou autre. Dans bien des situations économiques, un effet saisonnier lié à une période connue est pressenti. Une série journalière sera observée pendant plusieurs semaines avec une périodicité de 5, 6 ou 7 jours selon le cas; pour une série mensuelle observée sur plusieurs années, la période est égale à 12 ; pour une série trimestrielle observée sur plusieurs années, la période est égale à 4.

La variable mesurée peut être l'état d'une grandeur à l'instant de mesure, on parle de niveau d'un stock, du chiffre d'affaires, du bilan d'une activité au cours de la dernière période écoulée, etc.

Une série chronologique doit respecter les points suivants :

- **Régularité des observations** : ce n'est pas toujours vrai pour beaucoup de variables économiques ou financières puisque les mois ne comportent pas le même nombre de jours, en particulier de jours ouvrables.
- **Stabilité des structures conditionnant le phénomène étudié** : La plupart des séries étudiées concernent des grandeurs économiques et les techniques d'analyse cherchent à déterminer l'évolution lente du phénomène ainsi que ses variations saisonnières (pour une meilleure compréhension ou à des fins de prévision). Cela suppose une certaine stabilité qui, lorsqu'elle n'est pas vérifiée, peut être obtenue en décomposant la série observée en plusieurs séries successives.
- **Permanence de la définition de la grandeur étudiée** : Cette condition, qui parait évidente, n'est parfois pas respectée. C'est en particulier le cas de certains indices économiques (changement de base ou carrément du mode de calcul de l'indice).
- **Aspect périodique d'une partie de la grandeur observée** : Cette condition est indispensable dans l'usage des techniques cherchant à déterminer des variations saisonnières. Elle suppose comparable deux observations relatives au même mois de deux années différentes. Elle n'exclut pas l'existence d'une évolution lente. Elle indique qu'une part du phénomène (la composante saisonnière) se répète de façon plus ou moins identique d'une année à l'autre. Dans ce cas il est souvent commode de présenter les données dans une table à double entrée.

**Exemple 1 :** les ventes trimestrielles en milliers de DH réalisées par une entreprise au cours des quatre dernières années sont regroupées dans le tableau suivant :

| Années | Trimestre 1 | Trimestre 2 | Trimestre 3 | Trimestre 4 |
|---|---|---|---|---|
| 2002 | 190 | 160 | 251 | 200 |
| 2003 | 320 | 290 | 359 | 317 |
| 2004 | 426 | 405 | 483 | 433 |
| 2005 | 558 | 525 | 607 | 550 |

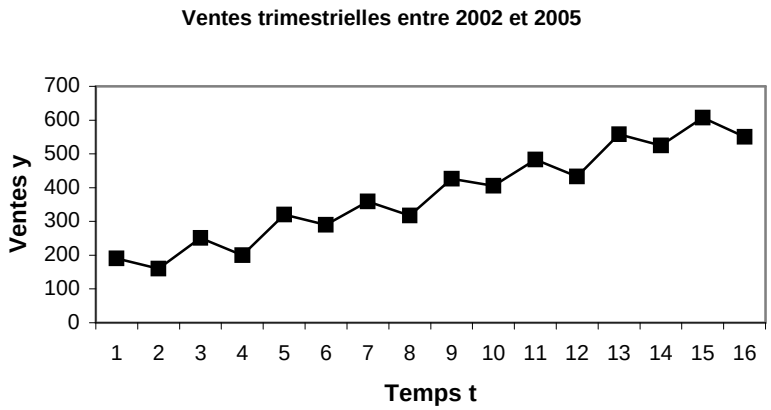
## 5.2. REPRESENTATION GRAPHIQUE.

La représentation graphique des observations est une étape indispensable avant d'entreprendre une analyse plus technique d'une série chronologique. Les points (t , y), avec t = 1, 2, 3, etc. sont représentés dans un système d'axes orthogonaux. Ils sont joints chronologiquement par des segments de droite pour faciliter la visualisation. Cette représentation permet d'apprécier l'évolution lente du phénomène, de dégager les périodes de stabilité. Elle suggère parfois d'opérer une transformation de la grandeur. Cette représentation graphique est également utile pour le choix d'un modèle.

**Exemple 2 :** Reprenons les données de l'exemple 1 et représentons graphiquement la série des ventes. Pour ce faire, classons les données par ordre chronologique en affectant à chaque trimestre son numéro d'ordre.

Pour cette présentation, les données doivent être transformées en une série statistique à deux variables, la variable y désignant les ventes et la variable t représentant le temps.

| Temps t | Vente y |
|---|---|
| 1 | 190 |
| 2 | 160 |
| 3 | 251 |
| 4 | 200 |
| 5 | 320 |
| 6 | 290 |
| 7 | 359 |
| 8 | 317 |
| 9 | 426 |
| 10 | 405 |
| 11 | 483 |
| 12 | 433 |
| 13 | 558 |
| 14 | 525 |
| 15 | 607 |
| 16 | 550 |



## 5.3. LES PRINCIPAUX MOUVEMENTS DES SERIES CHRONOLOGIQUES.

La succession des données observées ou série brute, résulte de quatre composantes ou mouvements :

### 5.3.1. Tendance.

La composante fondamentale ou tendance (trend, en Anglais) traduit l'évolution à moyen terme du phénomène. On parle aussi de mouvement conjoncturel ou mouvement extra-saisonnier. La série chronologique peut être globalement croissante, décroissante ou stable. La connaissance du trend permet la comparaison des séries chronologiques. De plus c'est à partir de la tendance que seront étudiées les autres composantes de la série. En effet, la grandeur étudiée ne suit pas généralement un mouvement régulier, mais fluctue au cours du temps. Ces fluctuations sont de natures différentes selon leur périodicité.

Le trend est une fonction à variation lente, elle sera estimée sous forme paramétrique ou comme le résultat d'une opération de lissage.

### 5.3.2. La composante saisonnière.

La composante saisonnière ou mouvement saisonnier représente des effets périodiques de période connue p qui se reproduisent de façon plus ou moins identique d'une période sur l'autre. La composante saisonnière permet simplement de distinguer à l'intérieur d'une même période une répartition stable dans le temps d'effets positifs ou négatifs qui se compensent sur l'ensemble de la période, c'est-à-dire, au-dessus ou au-dessous du trend. L'étude de ces fluctuations est indispensable pour la prévision à court terme. L'élimination du mouvement saisonnier est nécessaire à la poursuite de l'étude de la série.

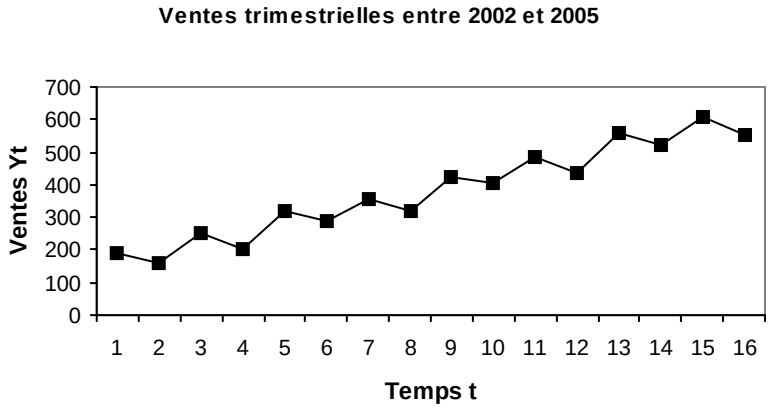
### 5.3.3. La composante cyclique.

La composante cyclique rend compte des fluctuations longues que la variable peut parfois présenter autour de la tendance. Les fluctuations cycliques qui traduisent la vie économique peuvent avoir une amplitude de plusieurs années qui est souvent mal définie. Cette composante est prise en compte dans la tendance sur les séries de taille moyenne et ne sera pas étudiée en tant que telle ici.

### 5.3.4. La composante résiduelle.

La composante résiduelle ou variations accidentelles est la partie non structurée du phénomène. Ce sont des variations à caractère souvent imprévisible et qui modifient ponctuellement la série chronologique : grève, guerre, mesures fiscales, sécheresse pour les productions agricoles. On parle de bruit blanc.

**Exemple 3 :** Reprenons le graphique de l'exemple 2.




On remarque sur la représentation graphique ci-dessus, ce qui suit :

- l'évolution à moyen terme des ventes se traduit par une tendance croissante ;

- la représentation graphique n'indique aucune fluctuation non structurée qui modifie ponctuellement la série des ventes. Il y a donc une faible présence de la composante résiduelle.

- la série des ventes ne suit pas un mouvement régulier, mais fluctue au cours du temps, autour de sa tendance. Ces fluctuations sont de natures différentes selon leur périodicité.

La représentation graphique indique des fluctuations périodiques de période 4 qui se reproduisent de façon plus ou moins identique d'un trimestre sur l'autre. En effet :

- du premier au deuxième trimestre de chaque année on constate une baisse des ventes ;

- au troisième trimestre de chaque année, il y a une hausse des ventes ;

- au quatrième trimestre de chaque année, on note de nouveau une baisse des ventes.

On peut donc parler d'un effet saisonnier.

## 5.4. LES SCHEMAS DE COMPOSITION.

La donnée observée à la date t ou donnée brute d'une série chronologique, désignée par y(t), peut donc s'interpréter comme résultant de la superposition des quatre composantes, le Trend désigné par Tt ; la composante saisonnière St, la composante cyclique Ct et la composante résiduelle Rt.

Pour pouvoir séparer les quatre composantes servant à décrire la série observée, il est nécessaire de préciser leur mode d'interaction. La plupart des séries chronologiques entrent dans l'un des schémas suivants :

### 5.4.1. Schéma additif.

Selon ce schéma, la série brute résulte de la somme du mouvement de longue durée Tt, du mouvement saisonnier St, du mouvement cyclique Ct et du mouvement accidentel ou résiduel Rt :

$$y(t) = Tt + St + Ct + Rt$$

St, Ct, et Rt sont alors les éléments que l'on doit ajouter à la valeur Tt de la tendance à la date t pour obtenir la donnée observée y(t). Ce modèle considère que les mouvements saisonnier et cyclique sont indépendants du niveau de y atteint sur le trend.

### 5.4.2. Schéma multiplicatif.

On peut au contraire penser que les variations cycliques et saisonnières suivent l'évolution générale de la grandeur. On adopte alors un modèle multiplicatif :

$$Y(t) = Tt \times St \times Ct \times Rt$$

Où St, Ct et Rt sont les coefficients par lesquels on doit multiplier Tt, position sur le Trend à la date t, pour obtenir la donnée observée y.

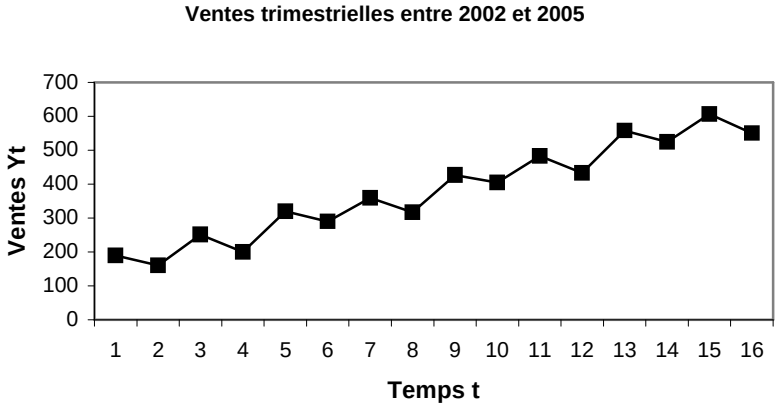
### 5.4.3. Schéma mixte.

On peut aussi noter que ces deux hypothèses ne sont pas incompatibles. Le schéma additif et le schéma multiplicatif peuvent être combinés pour donner un schéma dit « mixte ».

$$Yt = Tt \times St + Ct + Rt$$

Les modèles sus-indiqués sont tous acceptables. Cependant, il est fréquemment fait usage du modèle multiplicatif pour étudier les techniques associées à l'analyse des séries chronologiques.

**Exemple 4 :** Reprenons le graphique de l'exemple 2.




On peut remarquer sur le graphique que les variations saisonnières suivent l'évolution générale de la série, on adopte alors un modèle multiplicatif :

$$Y(t) = Tt \times St \times Rt$$

Où St et Rt sont les coefficients par lesquels on doit multiplier Tt, position sur le Trend à la date t, pour obtenir la donnée observée y(t).

## 5.5. LES METHODES DE LISSAGE.

Les méthodes de lissage sont des méthodes de réduction ou d'élimination des fluctuations aléatoires dans le but de découvrir l'existence d'autres composantes.

### 5.5.1. La méthode des moyennes mobiles.

Les opérations de lissage sont réalisées par le biais de moyennes mobiles. Celles-ci sont très utilisées car elles sont à la fois de conceptions simples, faciles à mettre en œuvre et suffisantes dans bien des situations.

Une série chronologique est lissée en remplaçant chaque valeur y(t) par une moyenne arithmétique des valeurs qui l'entourent. Une moyenne mobile pour une période de temps est une moyenne arithmétique simple des valeurs de cette période et de celles avoisinantes.

Le lissage d'une série chronologique y(t), par une moyenne mobile d'ordre impair n = 2k + 1 est défini pour t = k + 1, . . . , T - k, par :

$$MM(y(t)) = \frac{1}{n}(Y_{t-k} + \ldots + Y_t + \ldots + Y_{t+k})$$

Par exemple, pour calculer les moyennes mobiles de longueur 3 pour une période quelconque, nous sommons 3 valeurs de la série chronologique : la valeur de la série de la période en question, la valeur de celle qui précède et la valeur de celle qui suit et nous divisons par 3. Nous calculons les moyennes mobiles pour toutes les périodes exceptés la première et la dernière.

Il est difficile de discerner les composantes de la série chronologique si l'on se réfère uniquement au graphe représentatif de la série brute et ce en raison du large volume ou effet de la variation aléatoire présente. Pour essayer de voir comment la méthode des moyennes mobiles réduit les fluctuations aléatoires, on se réfère à la représentation graphique de la série des moyennes mobiles.

Il est à noter aussi que les moyennes mobiles de longueur 5 «lissent» la série brute plus que lorsqu'on utilise les moyennes mobiles de longueur 3. En général, plus la période sur laquelle nous faisons les moyennes est longue, plus la série brute devient lisse.

La série lissée est plus courte que l'originale puisque des valeurs sont manquantes à chaque extrémité de la période d'observation.

**Exemple 5 :** Reprenons les données de l'exemple 1 et calculons les moyennes mobiles d'ordre 3 et les moyennes mobiles d'ordre 5.

Utilisons la présentation des données sous forme d'une série statistique à deux variables, la variable y(t) désignant les ventes et la variable t représentant le temps.

Pour calculer les moyennes mobiles de longueur 3 pour une période quelconque, nous sommons la valeur de la série chronologique de la période en question aux valeurs de celle qui précède et de celle qui suit et nous divisons par 3. Nous calculons les moyennes mobiles pour toutes les périodes exceptés la première et la dernière.

$$MM3(y(t)) = \frac{1}{3}(y_{t-1} + y_t + y_{t+1})$$

| **Temps t** | **Ventes y** |
|---|---|
| 1 | 190 |
| 2 | 160 |
| 3 | 251 |
| 4 | 200 |
| 5 | 320 |
| 6 | 290 |
| 7 | 359 |
| 8 | 317 |
| 9 | 426 |
| 10 | 405 |
| 11 | 483 |

| | |
|---|---|
| 12 | 433 |
| 13 | 558 |
| 14 | 525 |
| 15 | 607 |
| 16 | 550 |

Faisons, par exemple, les calculs pour MM3($Y_2$) et MM3($Y_3$) :

$$MM3(y_2) = \frac{1}{3}(190 + 160 + 251) = 200,33$$

$$MM3(y_3) = \frac{1}{3}(160 + 251 + 200) = 203,67$$

Pour calculer les moyennes mobiles de longueur 5, pour une période quelconque, nous sommons la valeur de la série chronologique de la période en question aux 2 valeurs précédentes et aux 2 valeurs suivantes et nous divisons par 5. Nous calculons les moyennes mobiles pour toutes les périodes exceptés les 2 premières et les 2 dernières.

$$MM5(y(t)) = \frac{1}{5}(y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2})$$

Faisons, par exemple, les calculs pour MM5($Y_3$) et MM5($Y_4$) :

$$MM5(Y_3) = \frac{1}{5}(190 + 160 + 251 + 200 + 320) = 224,20$$

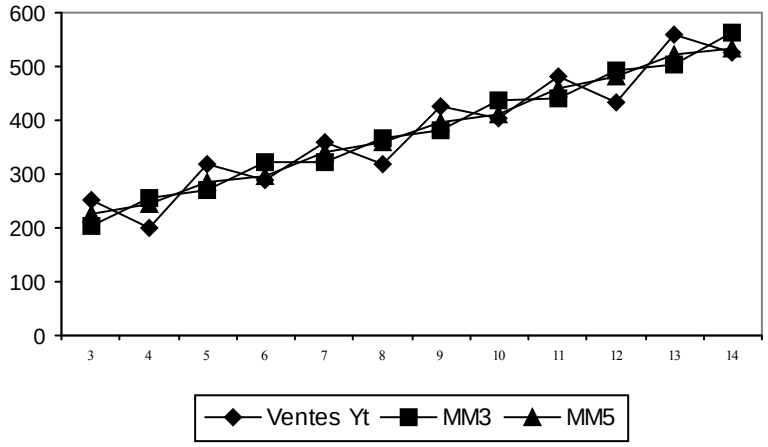
$$MM5(Y_4) = \frac{1}{5}(160 + 251 + 200 + 320 + 290) = 244,20$$

Le tableau ci-dessous donne les résultats pour les moyennes mobiles de longueur 3, MM3 et de longueur 5, MM5 :

| **Temps t** | **Vente y** | **Moyennes mobiles d'ordre 3 (MM3)** | **Moyennes mobiles d'ordre 5 (MM5)** |
|---|---|---|---|
| 1 | 190 | - | - |
| 2 | 160 | 200,33 | - |
| 3 | 251 | 203,67 | 224,20 |
| 4 | 200 | 257,00 | 244,20 |
| 5 | 320 | 270,00 | 284,00 |
| 6 | 290 | 323,00 | 297,20 |
| 7 | 359 | 322,00 | 342,40 |
| 8 | 317 | 367,33 | 359,40 |
| 9 | 426 | 382,67 | 398,00 |
| 10 | 405 | 438,00 | 412,80 |
| 11 | 483 | 440,33 | 461,00 |
| 12 | 433 | 491,33 | 480,80 |
| 13 | 558 | 505,33 | 521,20 |

| | | | |
|---|---|---|---|
| 14 | 525 | 563,33 | 534,60 |
| 15 | 607 | 560,67 | - |
| 16 | 550 | - | - |

Pour essayer de voir comment la méthode des moyennes mobiles réduit les fluctuations aléatoires, examinons les représentations graphiques de la série brute, de la série des moyennes mobiles MM3 et de la série des moyennes mobiles MM5.




On remarque bien, sur le graphique, que les moyennes mobiles de longueur 5 «lissent» la série brute plus que les moyennes mobiles de longueur 3. En général, plus la période sur laquelle nous faisons les moyennes est longue, plus la série brute devient lisse.

### 5.5.2. La méthode des moyennes mobiles centrées.

Si l'on décide d'adopter un nombre pair de périodes pour calculer les moyennes mobiles, nous serons confrontés au problème de la place ou position des moyennes mobiles calculées. Obtenir des moyennes mobiles qui se situent entre deux périodes cause des problèmes notamment d'interprétation. La méthode des moyennes mobiles centrées corrige ce problème. Cette méthode consiste à calculer des moyennes mobiles d'ordre 2 aux moyennes mobiles déjà obtenues.

**Exemple 6 :** Reprenons les données de l'exemple 5 et calculons les moyennes mobiles d'ordre 4.

Pour calculer les moyennes mobiles de longueur 4 nous sommons les valeurs de la série chronologique de 4 périodes successives et nous divisons par 4. Les moyennes mobiles ainsi calculées se positionnent entre 2 périodes. La méthode des moyennes mobiles centrées

corrige ce problème. Cette méthode consiste à calculer des moyennes mobiles d'ordre 2 aux moyennes mobiles déjà obtenues.

$$MM4(y_{(t-1\,;\,t)}) = \frac{1}{4}(y_{t-2} + y_{t-1} + y_t + y_{t+1})$$

$$MM4(y_{(t\,;\,t+1)}) = \frac{1}{4}(y_{t-1} + y_t + y_{t+1} + y_{t+2})$$

La moyenne mobile centrée pour la période t est :

$$MMC4(y_t) = \frac{1}{2}[MM4(y_{(t-1;\,t)}) + MM4(y_{(t\,;\,t+1)})]$$

Des trois égalités précédentes, nous pouvons, sans calculer les moyennes mobiles d'ordre 4, donner directement l'expression de la moyenne mobile centrée pour la période t :

$$MMC4(y_t) = \frac{1}{4}(0{,}5\, y_{t-2} + y_{t-1} + y_t + y_{t+1} + 0{,}5\, y_{t+2})$$

Faisons, par exemple, les calculs pour $MM4(Y_{(2;\,3)})$ et $MM4(Y_{(3;\,4)})$ :

$$MM4(y_{(2\,;\,3)}) = \frac{1}{4}(190 + 160 + 251 + 200) = 200{,}25$$

$$MM4(y_{(3\,;\,4)}) = \frac{1}{4}(160 + 251 + 200 + 320) = 232{,}75$$

La moyenne mobile centrée pour la période 3 est :

$$MMC4(y_3) = \frac{1}{2}(200{,}25 + 232{,}75) = 216{,}5$$

La moyenne mobile centrée pour la période 3 peut être directement calculée par :

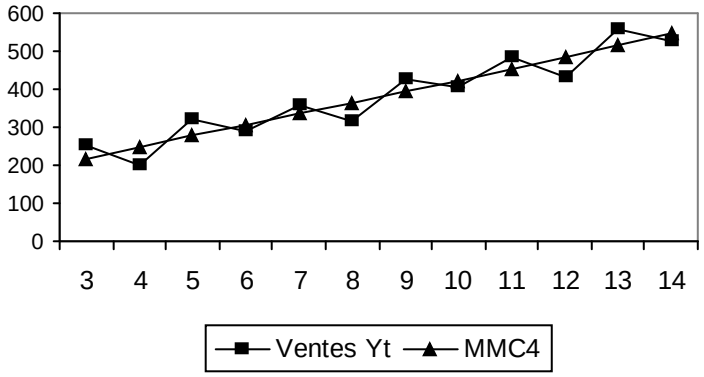
$$MMC4(y_3) = \frac{1}{4}(0{,}5 \times 190 + 160 + 251 + 200 + 0{,}5 \times 320) = 216{,}5$$

Le tableau ci-dessous donne les résultats pour les moyennes mobiles de longueur 4 MM4 et les moyennes mobiles centrées MMC4 :

| **Temps t** | **Vente Y(t)** | **Moyennes mobiles d'ordre 4 (MM4)** | **Moyennes mobiles centrées d'ordre 4 (MMC4)** |
|---|---|---|---|
| 1 | 190 | | - |
| 2 | 160 | | - |
| | | 200,25 | |

| | | | |
|---|---|---|---|
| 3 | 251 | | 216,50 |
| | | 232,75 | |
| 4 | 200 | | 249,00 |
| | | 265,25 | |
| 5 | 320 | | 278,75 |
| | | 292,25 | |
| 6 | 290 | | 306,88 |
| | | 321,50 | |
| 7 | 359 | | 334,75 |
| | | 348,00 | |
| 8 | 317 | | 362,38 |
| | | 376,75 | |
| 9 | 426 | | 392,25 |
| | | 407,75 | |
| 10 | 405 | | 422,25 |
| | | 436,75 | |
| 11 | 483 | | 453,25 |
| | | 469,75 | |
| 12 | 433 | | 484,75 |
| | | 499,75 | |
| 13 | 558 | | 515,25 |
| | | 530,75 | |
| 14 | 525 | | 545,38 |
| | | 560,00 | |
| 15 | 607 | | - |
| 16 | 550 | | - |

Examinons la représentation graphique de la série brute et de la série des moyennes mobiles centrées d'ordre 4 MMC4.




On remarque bien, sur le graphique, que les moyennes mobiles centrées d'ordre 4 ont lissé la série brute.

### 5.5.3. La méthode exponentielle

Deux inconvénients sont associés à la méthode des moyennes mobiles pour le lissage d'une série chronologique :

- Premièrement, nous n'avons pas de moyennes mobiles pour le premier et le dernier groupes de périodes de la série. Au cas où la série chronologique serait composée d'un nombre limité d'observations, les valeurs omises peuvent représenter une importante perte d'information ;

- Deuxièmement, les moyennes mobiles «négligent» la plupart des valeurs précédentes de la série chronologique, la moyenne mobile reflète des périodes avoisinantes mais n'est pas affectée par tout le passé.

Ces deux inconvénients sont corrigés par la méthode exponentielle d'une série qui est définie de la façon suivante :

$$St = w\, y_t + (1-w)\, S_{t-1} \qquad \text{pour } t \geq 2$$

Avec :

- St : valeur de la série chronologique lissée exponentiellement à la date t.
* $y(t) = y_t$ : valeur de la série chronologique à la date t .
* $S_{t-1}$ : valeur de la série chronologique lissée exponentiellement à la date t 1.
* w : constante ou coefficient de lissage, avec $0 \leq w \leq 1$.
* (1-w), appelé facteur d'oubli, représente le poids accordé à la nouvelle acquisition.

On commence par poser : $S_1 = y_{1,}$ ce qui donne :

$S_2 = w\, y_2 + (1\text{-}w)\, S_1 = wy_2 + (1\text{-}w)\, y_1$
$S_3 = w\, y_3 + (1\text{-}w)\, S_2 = w\, y_3 + (1\text{-}w)\, [w\, y_2 + (1\text{-}w)\, y_1]$
$S_3 = w\, y_3 + w\,(1\text{-}w)\, y_2 + (1\text{-}w)^2\, y_1$

En règle générale, on obtient :

$$St = w\, y_t + w\,(1\text{-}w)\, y_{t-1} + w\,(1\text{-}w)^2\, y_{t-2} + \ldots + (1\text{-}w)^{t-1}\, y_1$$

Cette dernière formule indique que la série « lissée » à la date t, dépend de toutes les observations antérieures de la série chronologique. L'intérêt de la méthode réside dans la facilité de mise à jour lors de l'acquisition d'une nouvelle donnée.

Le choix de la constante de lissage est important. Les valeurs proches de 0 produisent un degré de lissage assez important et correspondent à un lissage rigide, car le passé intervient peu, alors que les valeurs proches de 1 résultent dans un lissage assez limité de la série et donnent un lissage souple où le passé conserve, assez longtemps, son influence.

La particularité consiste à accorder aux valeurs passées une importance qui décroît de manière exponentielle avec le temps, on parle de facteur d'oubli. L'autre point important est que la mise à jour, lors de l'acquisition d'une nouvelle observation $y_{T+1}$, est réalisée de façon simple.

Le lissage exponentiel n'est pas adapté à une série chronologique présentant une tendance variant fortement ou un effet saisonnier très marqué.

**Exemple 7 :** Reprenons les données de l'exemple 5 et appliquons la méthode exponentielle de lissage avec w = 0,2 et w = 0,7 et représentons graphiquement les résultats.

Les valeurs lissées exponentiellement sont obtenues à partir de la formule suivante :

$$St = w\, y_t + (1\text{-}w)\, S_{t-1} \qquad \text{Pour } t \geq 2$$

On commence par poser : $S_1 = y_1 = 190$

Pour w = 0,2 on a :

$S_2 = 0{,}2 \times 160 + 0{,}8 \times 190 = 184$
$S_3 = 0{,}2 \times 251 + 0{,}8 \times 184 = 197{,}40$

Pour w = 0,7 on a :

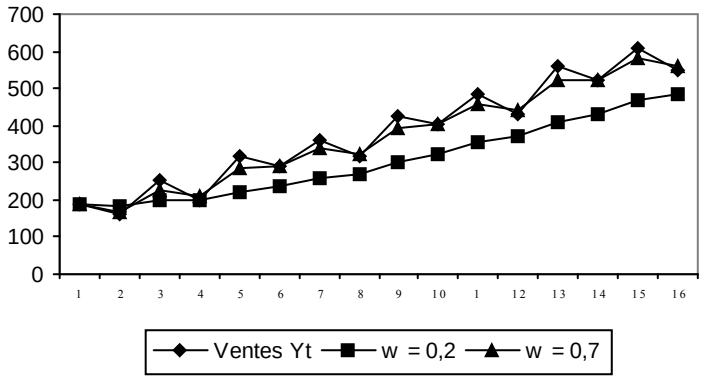
$$S_2 = 0{,}7 \times 160 + 0{,}3 \times 190 = 169$$

$$S_3 = 0,7 \times 251 + 0,3 \times 169 = 226,40$$

Le tableau, ci-dessous, donne les résultats de calculs pour le lissage exponentiel, pour w = 0,2 et le lissage exponentiel, pour w = 0,7 :

| Temps t | Ventes yt | Lissage exponentiel w = 0,2 | Lissage exponentiel w = 0,7 |
|---|---|---|---|
| 1 | 190 | 190,00 | 190,00 |
| 2 | 160 | 184,00 | 169,00 |
| 3 | 251 | 197,40 | 226,40 |
| 4 | 200 | 197,92 | 207,92 |
| 5 | 320 | 222,34 | 286,38 |
| 6 | 290 | 235,87 | 288,91 |
| 7 | 359 | 260,50 | 337,97 |
| 8 | 317 | 271,80 | 323,29 |
| 9 | 426 | 302,64 | 395,19 |
| 10 | 405 | 323,11 | 402,06 |
| 11 | 483 | 355,09 | 458,72 |
| 12 | 433 | 370,67 | 440,72 |
| 13 | 558 | 408,14 | 522,81 |
| 14 | 525 | 431,51 | 524,34 |
| 15 | 607 | 466,61 | 582,20 |
| 16 | 550 | 483,29 | 559,66 |

Pour essayer de voir comment la méthode exponentielle réduit les fluctuations aléatoires, examinons les représentations graphiques de la série brute, de la série lissée exponentiellement à 0,2 et de la série lissée exponentiellement à 0,7.



On voit bien, sur le graphique, que le lissage exponentiel w = 0,2 «lissent» la série brute plus que le lissage exponentiel w = 0,5. En général, plus le coefficient de lissage est faible, plus la série brute devient lisse.

## 5.6. ETUDE DU TREND.

La régression linéaire est la méthode la plus simple pour analyser la tendance générale d'une série chronologique où la variable indépendante est le temps t.

Le trend peut être soit linéaire ou non linéaire et par conséquent peut prendre des formes fonctionnelles assez diverses.

### 5.6.1. Modèle linéaire.

Si nous estimons que la tendance de longue période est essentiellement linéaire, on utilisera le modèle suivant :

$$\hat{y}_t = a\,t + b$$

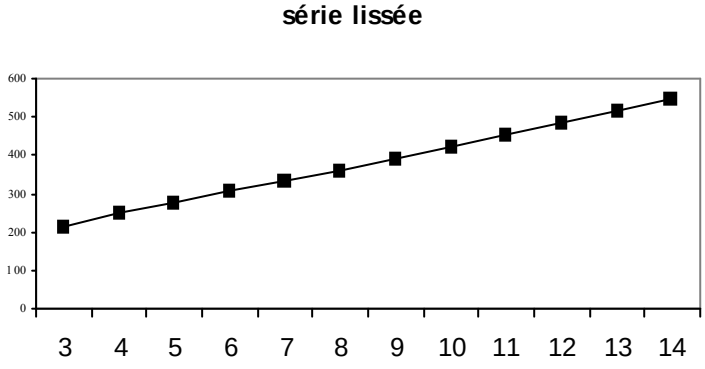
L'estimation de a et de b par la méthode des moindres carrés se fait par les formules développées dans le chapitre précédent :

$$a = \frac{\sum t_i y_i - n\,\bar{t}\bar{y}}{\sum t_i^2 - n\,\bar{t}^2} = \frac{COV(t,y)}{S_t^2} \qquad \text{et} \qquad b = \bar{y} - a\,\bar{t}$$

**Exemple 8 :** Reprenons les données de l'exemple 1 et déterminons l'équation du trend.

L'équation du trend sera déterminée à partir de la série lissée par la méthode des moyennes mobiles centrées d'ordre 4 calculée à l'exemple 6.

Représentons graphiquement la série lissée :



D'après le graphique, on voit bien que le trend est linéaire.

Calculons alors l'équation du trend : $\hat{y}_t = a\,t + b$

| | **Temps t** | **MMC4** | **t²** | **t x MMC4** |
|---|---|---|---|---|
| | 3 | 216,5 | 9 | 649,5 |
| | 4 | 249 | 16 | 996 |
| | 5 | 278,75 | 25 | 1393,75 |
| | 6 | 306,875 | 36 | 1841,25 |
| | 7 | 334,75 | 49 | 2343,25 |
| | 8 | 362,375 | 64 | 2899 |
| | 9 | 392,25 | 81 | 3530,25 |
| | 10 | 422,25 | 100 | 4222,5 |
| | 11 | 453,25 | 121 | 4985,75 |
| | 12 | 484,75 | 144 | 5817 |
| | 13 | 515,25 | 169 | 6698,25 |
| | 14 | 545,375 | 196 | 7635,25 |
| **Total** | **102** | **4561,375** | **1010** | **43011,75** |

$$\bar{t} = \frac{102}{12} = 8{,}5$$

$$\overline{MMC4} = \frac{4561{,}375}{12} = 380{,}11$$

$$S_t^{\,2} = \frac{1010}{12} - 8{,}5^2 = 11{,}92$$

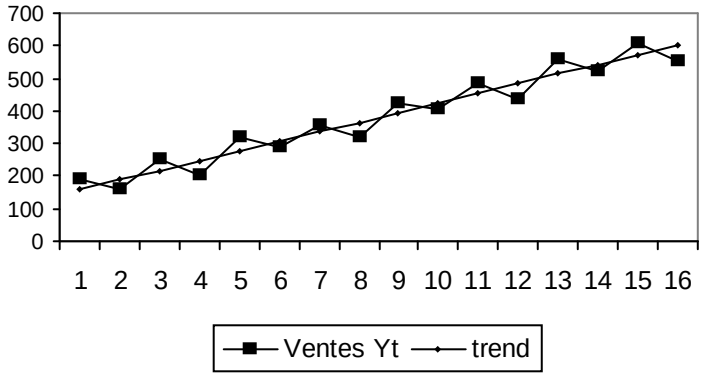
$$COV(t\,;\,MMC4) = \frac{43011{,}75}{12} - 8{,}5 \times 380{,}11 = 353{,}3775$$

$$a = \frac{353{,}3775}{11{,}92} = 29{,}65$$

b = 380,11 – 29,65 x 8,5 = 128,08

L'équation du trend est : $\hat{y}_t = 29{,}65\,t + 128{,}08$

Reportons la droite d'équation y = 29,65 t + 128,08 sur le graphe de la série tel que nous l'avons représenté pour l'exemple 1.




La droite de régression s'ajuste bien au nuage de points de la série chronologique, elle montre clairement une tendance linéaire croissante de la série.

### 5.6.2. Modèle exponentiel.

$$\hat{y}_t = a \times b^t$$

Le modèle logarithmique peut être traduit en termes de log de la façon suivante :

$$\text{Log}(\hat{y}_t) = \log(a) + (\log b)\, t$$

Si l'on pose les changements de variables suivants :

**Y' = log y** **a' = log (a)** **et** **b' = log (b)**

Le modèle devient linéaire :

$$Y' = a' + b' t$$

On calcule a' et b' à l'aide de la méthode des moindres carrés, comme développée, dans le chapitre précédent :

$$b' = \frac{\text{COV}(t,y')}{S_t^2} \quad \text{et} \quad a' = \bar{y}' - b'\bar{t}$$

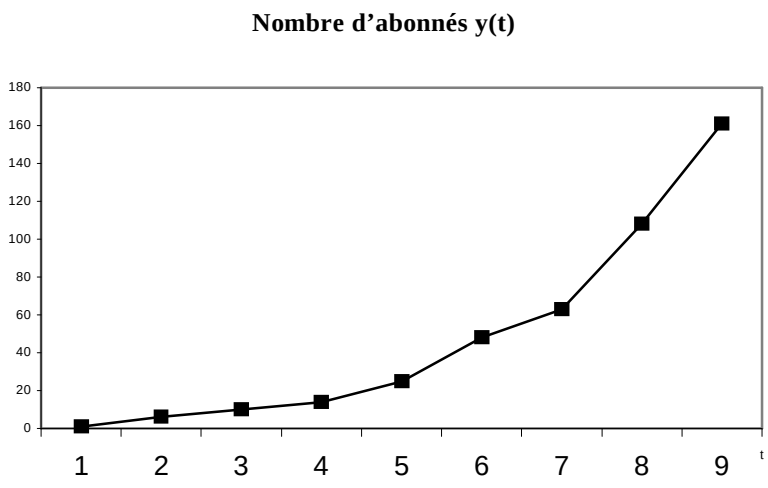
Les constantes a et b sont alors :

$$b = e^{b'} \quad \text{et} \quad a = e^{a'}$$

**Exemple 9 :** Le nombre d'abonnés à un service téléphonique au cours des neuf premiers mois de son lancement est comme suit :

| Mois | Période | Nombre d'abonnés y(t) |
|---|---|---|
| Janvier | 1 | 1 |
| Février | 2 | 6 |
| Mars | 3 | 10 |
| Avril | 4 | 14 |
| Mai | 5 | 25 |
| Juin | 6 | 48 |
| Juillet | 7 | 63 |
| Août | 8 | 108 |
| septembre | 9 | 161 |

Représentons graphiquement cette série :




D'après le graphique, on voit bien que la série présente une tendance exponentielle de la forme :

$$\hat{y}_t = a \times b^t$$

Le modèle logarithmique peut être traduit en termes de log de la façon suivante :

$$\text{Log}(\hat{y}_t) = \log(a) + (\log b)\, t$$

Si l'on pose les changements de variables suivants :

**y' = log y** **a' = log (a)** **et** **b' = log (b)**

Le modèle devient linéaire :

$$y' = a' + b't$$

On calcule a' et b' à l'aide de la méthode des moindres carrés :

| | **t** | **yt** | **y'** | **t²** | **t y'** |
|---|---|---|---|---|---|
| | 1 | 1 | 0,000 | 1 | 0,000 |
| | 2 | 6 | 1,792 | 4 | 3,584 |
| | 3 | 10 | 2,303 | 9 | 6,908 |
| | 4 | 14 | 2,639 | 16 | 10,556 |
| | 5 | 25 | 3,219 | 25 | 16,094 |
| | 6 | 48 | 3,871 | 36 | 23,227 |
| | 7 | 63 | 4,143 | 49 | 29,002 |
| | 8 | 108 | 4,682 | 64 | 37,457 |
| | 9 | 161 | 5,081 | 81 | 45,733 |
| **Total** | **45** | **- - -** | **27,730** | **285** | **172,561** |

$$\bar{t} = \frac{45}{9} = 5$$

$$\bar{Y'} = \frac{27,73}{9} = 3,08$$

$$S_t^2 = \frac{285}{9} - 5^2 = 6,67$$

$$COV(t ; Y') = \frac{172,561}{9} - 5 \times 3,08 = 3,773$$

$$b' = \frac{3,773}{6,67} = 0,566$$

a' = 3,08 – 0,566 x 5 = 0,25

Les constantes a et b sont alors :

$b = e^{b'} = e^{0,566} = 1,76$
$a = e^{a'} = e^{0,25} = 1,28$

L'équation du trend est donc : $\hat{y}_t = 1,28 \times b^{1,76}$

## 5.7. ETUDE DE LA COMPOSANTE SAISONNIERE.

### 5.7.1. Calcul des coefficients saisonniers.

Dans le but de mesurer l'effet saisonnier, on calcule des coefficients saisonniers, qui ont pour objet de mesurer le degré de différence entre les saisons.

Le calcul des coefficients saisonniers repose sur une démarche générale qui peut être décomposée en trois étapes essentielles :

- **La première étape** consiste à estimer les valeurs de la tendance $\hat{y}_t$ . La détermination de la tendance «trend» consiste à réduire les fluctuations et à dégager une évolution à long terme ;

- **La deuxième étape** consiste, par une confrontation entre les valeurs de la série brute et celles de la tendance, à calculer les valeurs du mouvement saisonnier. Cette confrontation peut se faire, de deux façons :

* Par un modèle multiplicatif, sous forme de rapports entre les valeurs de la série brute et celles de la tendance, ces rapports sont appelés rapports aux trends $r_t$ :

$$r_t = \frac{Yt}{\hat{Y}t}$$

* Par un modèle additif, on calcule les différences entre les valeurs de la série brute et celles de la tendance, on parle de différences aux trends.

- **La troisième étape** consiste à mesurer l'effet saisonnier à l'aide des coefficients saisonniers. Ces derniers sont obtenus en calculant, pour chaque saison, la moyenne des rapports ou des différences aux trends.

$$\bar{r}_k = \frac{\sum r_t}{n}$$

Avec k = 1 jusqu'au nombre de saisons et n est le nombre d'observations pour chaque saison.

Les coefficients saisonniers correspondent aux rapports moyens ajustés :

$$cs_k = \frac{\bar{r}_k}{\bar{r}}$$

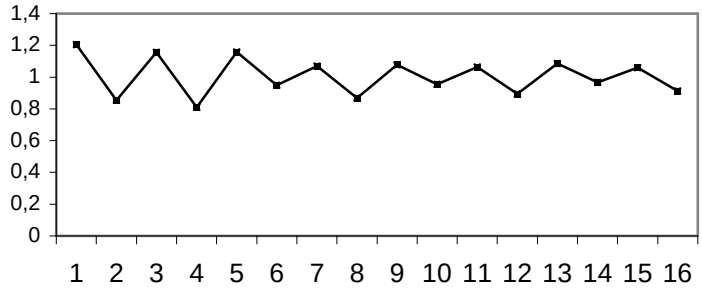
**Exemple 10 :** Reprenons les données de l'exercice 1 et calculons les coefficients saisonniers.

Calculons les valeurs du trend $\hat{y}_t$ à l'aide de l'équation du trend déjà calculée dans l'exemple 8 à savoir :

$$\hat{y}_t = 29{,}65\ t + 128{,}08$$

| | **Yt** | $\hat{y}_t$ | $r_t$ |
|---|---|---|---|
| 3 | 251 | 217,03 | 1,1565 |
| 4 | 200 | 246,68 | 0,8108 |
| 5 | 320 | 276,33 | 1,1580 |
| 6 | 290 | 305,98 | 0,9478 |
| 7 | 359 | 335,63 | 1,0696 |
| 8 | 317 | 365,28 | 0,8678 |
| 9 | 426 | 394,93 | 1,0787 |
| 10 | 405 | 424,58 | 0,9539 |
| 11 | 483 | 454,23 | 1,0633 |
| 12 | 433 | 483,88 | 0,8948 |
| 13 | 558 | 513,53 | 1,0866 |
| 14 | 525 | 543,18 | 0,9665 |

**rapports aux trend**




Le graphique des rapports aux valeurs du trend fait apparaître des fluctuations saisonnières. Les périodes 2 ; 4 ; 6 ; 8 ; 10 ; 12 ; 14 et 16 qui correspondent aux deuxième et quatrième trimestres sont des basses saisons, alors que les périodes 1 ; 3 ; 5 ; 7 ; 9 ; 11 ; 13 et 15 qui correspondent aux premier et troisième trimestres sont des hautes saisons.

Calculons les coefficients saisonniers dans le cas d'un modèle multiplicatif :

| **Années** | **Trimestre 1** | **Trimestre 2** | **Trimestre 3** | **Trimestre 4** |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| 2002 | - | - | 1,1565 | 0,8108 |
| 2003 | 1,1580 | 0,9478 | 1,0696 | 0,8678 |
| 2004 | 1,0787 | 0,9539 | 1,0633 | 0,8948 |
| 2005 | 1,0866 | 0,9665 | - | - |
| $\bar{n}_k$ | 1,1078 | 0,9561 | 1,0965 | 0,8578 |
| $\bar{r}$ | 1,00455 | | | |
| **Cs** | **1,1028** | **0,9518** | **1,0915** | **0,8539** |

Les coefficients saisonniers du premier et troisième trimestre sont supérieurs à 1 alors que ceux du deuxième et quatrième trimestre sont inférieurs à 1. Le deuxième et le quatrième trimestre sont donc des basses saisons, alors que le premier et le troisième trimestre sont des hautes saisons.

### 5.7.2. Désaisonnalisation d'une série chronologique.

Les techniques de désaisonnalisation consistent à éliminer d'une série chronologique l'effet de la composante saisonnière.

La série désaisonnalisée est obtenue :

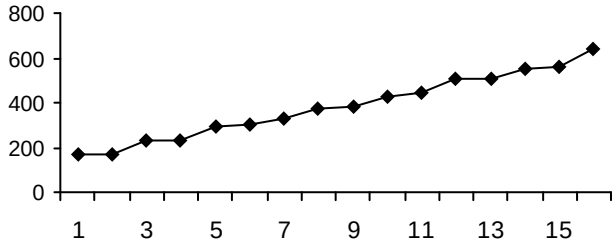
* dans le cas du modèle multiplicatif, en divisant les valeurs de la série brute par les coefficients saisonniers moyens correspondants ;
* dans le cas du modèle additif, en soustrayant des valeurs de la série brute les coefficients saisonniers moyens correspondants.

**Exemple 11 :** Reprenons les données de l'exemple 10 et calculons la série désaisonnalisée.

Rappelons que nous avons opté, dans l'exemple 10, pour un modèle multiplicatif, de ce fait et pour désaisonnaliser notre série, on divise les valeurs de la série brute par les coefficients saisonniers moyens correspondants, on obtient, après calculs :

| Années | Trimestre 1 | Trimestre 2 | Trimestre 3 | Trimestre 4 |
|---|---|---|---|---|
| 2002 | 172,29 | 168,10 | 229,96 | 234,22 |
| 2003 | 290,17 | 304,69 | 328,91 | 371,24 |
| 2004 | 386,29 | 425,51 | 442,51 | 507,09 |
| 2005 | 505,98 | 551,59 | 556,12 | 644,10 |

**Série désaisonnalisée**




La représentation de la série désaisonnalisée montre clairement l'élimination de l'effet saisonnier. La série désaisonnalisée conserve les irrégularités dues à la composante

résiduelle. En effet, au deuxième trimestre 2002 et premier trimestre 2005 on note une petite baisse accidentelle.

### 5.7.3. Calcul des prévisions.

A partir de l'équation du trend et des coefficients saisonniers, on peut prévoir les valeurs de la série pour les périodes à venir.

La prévision de la valeur de la série à la période t+k est, pour un modèle multiplicatif, la valeur estimée du trend multipliée par le coefficient saisonnier moyen de la saison correspondante.

**Exemple 12 :** Reprenons les données de l'exemple 10 et calculons les prévisions des ventes pour les quatre trimestres de l'année 2006.

L'équation du trend déjà calculée à l'exemple 8 est :

$$\hat{y}_t = 29{,}65\ t + 128{,}08$$

$$y_{t+k} = \hat{y}_{t+k} \times CS$$

| **Trimestres année 2006** | **Périodes** | **Valeurs du trend : $\hat{y}_t$** | **Coefficients saisonniers** | **Prévisions $\hat{Y}_{t+k} \times CS$** |
|---|---|---|---|---|
| 1er trimestre | t = 17 | 615,13 | 1,1028 | 678 |
| 2ème trimestre | t = 18 | 644,78 | 0,9518 | 614 |
| 3ème trimestre | t = 19 | 674,43 | 1,0915 | 736 |
| 4ème trimestre | t = 20 | 704,08 | 0,8539 | 601 |

**Exemple 13 :** Reprenons les données de l'exemple 1, relatives aux ventes trimestrielles réalisées par une entreprise au cours des quatre dernières années, et utilisons un modèle additif pour le calcul des coefficients saisonniers.

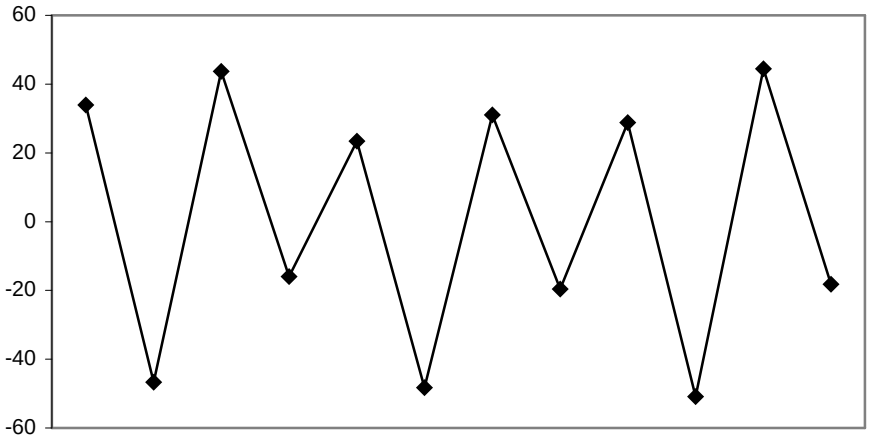
| **Années** | **Trimestre 1** | **Trimestre 2** | **Trimestre 3** | **Trimestre 4** |
|---|---|---|---|---|
| 2002 | 190 | 160 | 251 | 200 |
| 2003 | 320 | 290 | 359 | 317 |
| 2004 | 426 | 405 | 483 | 433 |
| 2005 | 558 | 525 | 607 | 550 |

Le modèle additif s'écrit : Yt = Tt + St + Ct + Rt

Le tableau suivant regroupe la série brute, les valeurs du trend et les différences au trend :

| **t** | $y_t$ | $\hat{y}_t$ | $d_t$ |
|---|---|---|---|
| 3 | 251 | 217,03 | 33,97 |
| 4 | 200 | 246,68 | -46,68 |
| 5 | 320 | 276,33 | 43,67 |
| 6 | 290 | 305,98 | -15,98 |
| 7 | 359 | 335,63 | 23,37 |
| 8 | 317 | 365,28 | -48,28 |
| 9 | 426 | 394,93 | 31,07 |
| 10 | 405 | 424,58 | -19,58 |
| 11 | 483 | 454,23 | 28,77 |
| 12 | 433 | 483,88 | -50,88 |
| 13 | 558 | 513,53 | 44,47 |
| 14 | 525 | 543,18 | -18,18 |

**différences au trend**




Le graphique des différences aux valeurs du trend fait apparaître des fluctuations saisonnières. Les périodes 2 ; 4 ; 6 ; 8 ; 10 ; 12 ; 14 et 16 qui correspondent aux deuxième et quatrième trimestres sont des basses saisons, alors que les périodes 1 ; 3 ; 5 ; 7 ; 9 ; 11 ; 13 et 15 qui correspondent aux premier et troisième trimestres sont des hautes saisons.

Calculons les coefficients saisonniers dans le cas du modèle additif :

| Années | Trimestre 1 | Trimestre 2 | Trimestre 3 | Trimestre 4 |
|---|---|---|---|---|
| 2002 | - | - | 33,97 | - 46,68 |
| 2003 | 43,67 | - 15,98 | 23,37 | - 48,28 |
| 2004 | 31,07 | - 19,58 | 28,77 | - 50,88 |
| 2005 | 44,47 | - 18,18 | - | - |
| $cs = d_k^J$ | **39,74** | **- 17,91** | **28,70** | **- 48,61** |

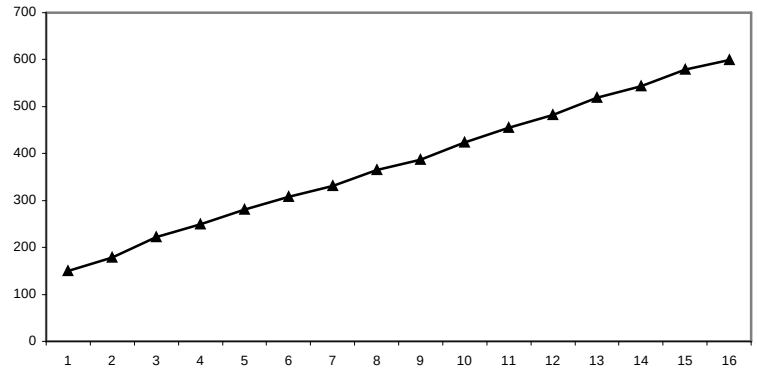
Les coefficients saisonniers du premier et troisième trimestre sont supérieurs à 0 alors que ceux du deuxième et quatrième trimestre sont inférieurs à 0. Le deuxième et le quatrième trimestre sont donc des basses saisons, alors que le premier et le troisième trimestre sont des hautes saisons.

**Désaisonnalisation de la série brute :**

La série désaisonnalisée est obtenue en soustrayant le coefficient saisonnier moyen de la valeur de la série brute.

| Années | Trimestre 1 | Trimestre 2 | Trimestre 3 | Trimestre 4 |
|---|---|---|---|---|
| 2002 | 150 | 178 | 222 | 249 |
| 2003 | 280 | 308 | 330 | 366 |
| 2004 | 386 | 423 | 454 | 482 |
| 2005 | 518 | 543 | 578 | 599 |

**Série désaisonnalisée**



La représentation de la série désaisonnalisée montre clairement l'élimination de l'effet saisonnier.

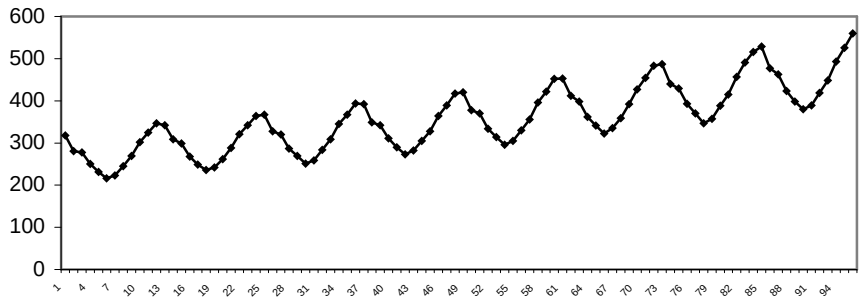
**Exemple 14** : Le tableau suivant donne la consommation mensuelle en électricité de l'entreprise Matex, pendant 8 ans.

| années | jan | fév | mar | avr | mai | juin | juil | aoû | sep | oct | nov | Déc |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1998 | 318 | 281 | 278 | 250 | 231 | 216 | 223 | 245 | 269 | 302 | 325 | 347 |
| 1999 | 342 | 309 | 299 | 268 | 249 | 236 | 242 | 262 | 288 | 321 | 342 | 364 |
| 2000 | 367 | 328 | 320 | 287 | 269 | 251 | 259 | 284 | 309 | 345 | 367 | 394 |
| 2001 | 392 | 349 | 342 | 311 | 290 | 273 | 282 | 305 | 328 | 364 | 389 | 417 |
| 2002 | 420 | 378 | 370 | 334 | 314 | 296 | 305 | 330 | 356 | 396 | 422 | 452 |
| 2003 | 453 | 412 | 398 | 362 | 341 | 322 | 335 | 359 | 392 | 427 | 454 | 483 |
| 2004 | 487 | 440 | 429 | 393 | 370 | 347 | 357 | 388 | 415 | 457 | 491 | 516 |
| 2005 | 529 | 477 | 463 | 423 | 398 | 380 | 389 | 419 | 448 | 493 | 526 | 560 |

**Représentation graphique :**

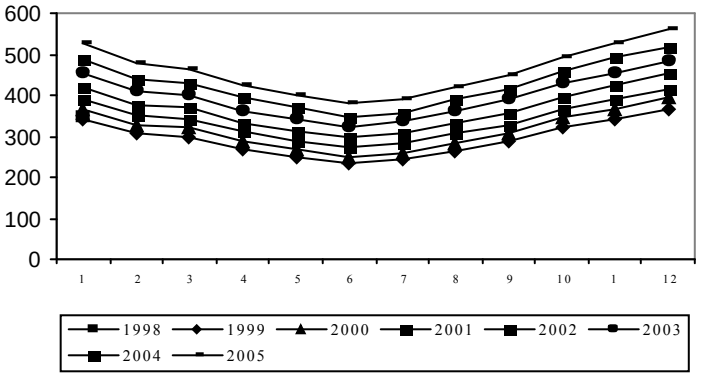
La présentation des données doit être transformée en une série statistique à deux variables, la variable y(t) désignant les ventes et la variable t représentant le temps.

**y(t)**




D'après ce graphe, la consommation mensuelle en électricité présente une tendance croissante. La série fluctue au cours du temps.

On pourrait donner une meilleure appréciation de ces fluctuations en représentant les graphes des nuages de points pour les 12 mois de chaque année. On obtient ainsi 8 courbes qui ont la même allure :




Ce deuxième graphique indique des fluctuations périodiques de période 12 qui se reproduisent de façon plus ou moins identique d'un mois à l'autre. On peut donc parler d'un effet saisonnier mensuel très net.

La représentation graphique n'indique aucune fluctuation accidentelle qui modifie ponctuellement la série. Il y a donc une faible présence de la composante résiduelle. La série brute peut être lissée à un ordre faible.

**Lissage de la série brute par la méthode des moyennes mobiles d'ordre 3 :**

Pour calculer les moyennes mobiles de longueur 3 pour une période quelconque, nous sommons la valeur de la série chronologique de la période en question aux valeurs de celle qui précède et de celle qui suit et nous divisons par 3. Nous calculons les moyennes mobiles pour toutes les périodes exceptés la première et la dernière. La formule de calcul est donc :

$$MM3(yt) = \frac{1}{3}(y_{t-1} + y_t + y_{t+1})$$

Faisons, à titre d'exemple, le calcul de $MM3(y_2)$ et de $MM3(y_3)$ :

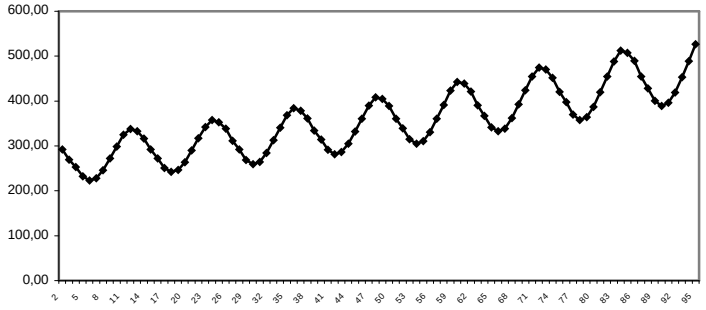
$$MM3(y_2) = \frac{1}{3}(318 + 281 + 278) = 292{,}33$$

$$MM3(y_3) = \frac{1}{3}(281 + 278 + 250) = 269{,}67$$

| jan | fév | mars | avr | mai | juin | juil | août | sept | oct | nov | déc |
|---|---|---|---|---|---|---|---|---|---|---|---|
| - | 292,33 | 269,67 | 253 | 232,33 | 223,33 | 228 | 245,67 | 272 | 298,67 | 324,67 | 338 |
| 332,67 | 316,67 | 292 | 272 | 251 | 242,33 | 246,67 | 264 | 290,33 | 317 | 342,33 | 357,67 |
| 353 | 338,33 | 311,67 | 292 | 269 | 259,67 | 264,67 | 284 | 312,67 | 340,33 | 368,67 | 384,33 |
| 378,33 | 361 | 334 | 314,33 | 291,33 | 281,67 | 286,67 | 305 | 332,33 | 360,33 | 390 | 408,67 |
| 405 | 389,33 | 360,67 | 339,33 | 314,67 | 305 | 310,33 | 330,33 | 360,67 | 391,33 | 423,33 | 442,33 |
| 439 | 421 | 390,67 | 367 | 341,67 | 332,67 | 338,67 | 362 | 392,67 | 424,33 | 454,67 | 474,67 |
| 470 | 452 | 420,67 | 397,33 | 370 | 358 | 364 | 386,67 | 420 | 454,33 | 488 | 512 |
| 507,33 | 489,67 | 454,33 | 428 | 400,33 | 389 | 396 | 418,67 | 453,33 | 489 | 526,33 | - |

Pour essayer de voir comment la méthode des moyennes mobiles réduit les fluctuations aléatoires, examinons la représentation graphique de la série des moyennes mobiles MM3.

**MM3**




On voit bien, sur le graphique, que la série des moyennes mobiles de longueur 3 est plus lisse que la série brute.

**Détermination du trend :**

D'après le graphique de la série brute, on peut affirmer que la tendance de longue période est linéaire, on utilisera le modèle suivant :

$$\hat{y}_t = a\,t + b$$

L'estimation de a et de b par la méthode des moindres carrés se fait par les formules :

$$a = \frac{\sum t_i y_i - n\,\bar{t}\bar{y}}{\sum t_i^2 - n\,\bar{t}^2} = \frac{COV(t,y)}{S_t^2} \qquad \text{et} \qquad b = \bar{y} - a\,\bar{t}$$

Les résultats des calculs sont regroupés dans le tableau suivant :

| | **Temps t** | **MM3** | **$t^2$** | **t x MM3** |
|---|---|---|---|---|
| Total | 4559,0 | 33480,66 | 290319 | 1775869,35 |
| moyenne | 48,5 | 356,2 | 3088,5 | 18892,2 |
| variance | 736,3 | | | |
| covariance | 1617,6 | | | |
| a | 2,2 | | | |
| b | 249,6 | | | |

L'équation du trend est donc : $\hat{y}_t = 2{,}2\,t + 249{,}6$

**Calcul des valeurs du trend** :

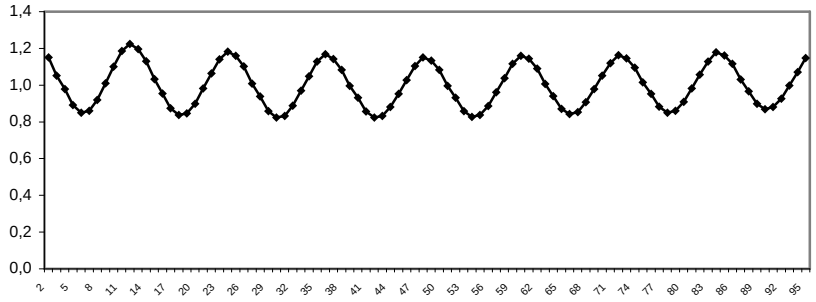
| **jan** | **fév** | **mars** | **avr** | **mai** | **juin** | **juil** | **août** | **sept** | **oct** | **nov** | **déc** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 251,8 | 254,0 | 256,2 | 258,4 | 260,6 | 262,8 | 265,0 | 267,2 | 269,4 | 271,6 | 273,8 | 276,0 |
| 278,2 | 280,4 | 282,6 | 284,8 | 287 | 289,2 | 291,4 | 293,6 | 295,8 | 298 | 300,2 | 302,4 |
| 304,6 | 306,8 | 309 | 311,2 | 313,4 | 315,6 | 317,8 | 320 | 322,2 | 324,4 | 326,6 | 328,8 |
| 331 | 333,2 | 335,4 | 337,6 | 339,8 | 342 | 344,2 | 346,4 | 348,6 | 350,8 | 353 | 355,2 |
| 357,4 | 359,6 | 361,8 | 364 | 366,2 | 368,4 | 370,6 | 372,8 | 375 | 377,2 | 379,4 | 381,6 |
| 383,8 | 386 | 388,2 | 390,4 | 392,6 | 394,8 | 397 | 399,2 | 401,4 | 403,6 | 405,8 | 408 |
| 410,2 | 412,4 | 414,6 | 416,8 | 419 | 421,2 | 423,4 | 425,6 | 427,8 | 430 | 432,2 | 434,4 |
| 436,6 | 438,8 | 441 | 443,2 | 445,4 | 447,6 | 449,8 | 452 | 454,2 | 456,4 | 458,6 | 460,8 |

**Détermination des coefficients saisonniers** :

| **jan** | **fév** | **mars** | **avr** | **mai** | **juin** | **juil** | **août** | **sept** | **oct** | **nov** | **déc** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| - | 1,151 | 1,053 | 0,979 | 0,892 | 0,850 | 0,860 | 0,919 | 1,010 | 1,100 | 1,186 | 1,225 |
| 1,196 | 1,129 | 1,033 | 0,955 | 0,875 | 0,838 | 0,846 | 0,899 | 0,982 | 1,064 | 1,140 | 1,183 |
| 1,159 | 1,103 | 1,009 | 0,938 | 0,858 | 0,823 | 0,833 | 0,888 | 0,970 | 1,049 | 1,129 | 1,169 |
| 1,143 | 1,083 | 0,996 | 0,931 | 0,857 | 0,824 | 0,833 | 0,880 | 0,953 | 1,027 | 1,105 | 1,151 |
| 1,133 | 1,083 | 0,997 | 0,932 | 0,859 | 0,828 | 0,837 | 0,886 | 0,962 | 1,037 | 1,116 | 1,159 |
| 1,144 | 1,091 | 1,006 | 0,940 | 0,870 | 0,843 | 0,853 | 0,907 | 0,978 | 1,051 | 1,120 | 1,163 |
| 1,146 | 1,096 | 1,015 | 0,953 | 0,883 | 0,850 | 0,860 | 0,909 | 0,982 | 1,057 | 1,129 | 1,179 |
| 1,162 | 1,116 | 1,030 | 0,966 | 0,899 | 0,869 | 0,880 | 0,926 | 0,998 | 1,071 | 1,148 | - |
| **Calcul des rapports moyens par mois** | | | | | | | | | | | |
| 1,155 | 1,106 | 1,017 | 0,949 | 0,874 | 0,840 | 0,850 | 0,902 | 0,979 | 1,057 | 1,134 | 1,552 |
| **Calcul de la moyenne des rapports moyens** | | | | | | | | | | | |

| 0,964 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Calcul des coefficients saisonniers moyens** | | | | | | | | | | | |
| 1,198 | 1,148 | 1,055 | 0,985 | 0,907 | 0,872 | 0,882 | 0,936 | 1,016 | 1,097 | 1,177 | 1,610 |

**rapports au trend**

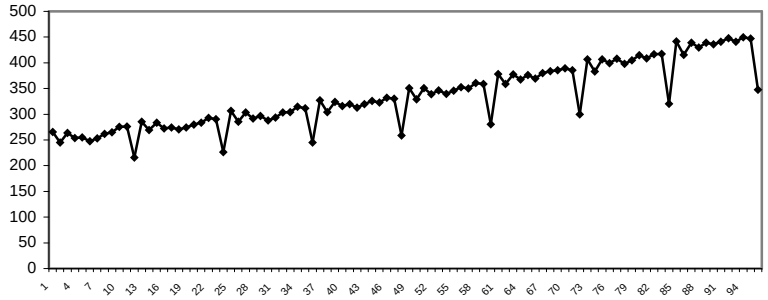



Le graphique des rapports aux valeurs du trend fait apparaître des fluctuations saisonnières. Les mois 4 ; 5 ; 6 ; 7 et 8 correspondent à une basse saison, alors que les mois 1 ; 2 ; 3 ; 9 ; 10 ; 11 et 12 correspondent à une haute saison.

**Détermination de la série désaisonnalisée** :

| | **jan** | **fév** | **mars** | **avr** | **mai** | **juin** | **juil** | **août** | **sept** | **oct** | **nov** | **Déc** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1998 | 265 | 245 | 264 | 254 | 255 | 248 | 253 | 262 | 265 | 275 | 276 | 216 |
| 1999 | 285 | 269 | 283 | 272 | 275 | 271 | 274 | 280 | 283 | 293 | 291 | 226 |
| 2000 | 306 | 286 | 303 | 291 | 297 | 288 | 294 | 303 | 304 | 314 | 312 | 245 |
| 2001 | 327 | 304 | 324 | 316 | 320 | 313 | 320 | 326 | 323 | 332 | 331 | 259 |
| 2002 | 351 | 329 | 351 | 339 | 346 | 339 | 346 | 353 | 350 | 361 | 359 | 281 |
| 2003 | 378 | 359 | 377 | 368 | 376 | 369 | 380 | 384 | 386 | 389 | 386 | 300 |
| 2004 | 407 | 383 | 407 | 399 | 408 | 398 | 405 | 415 | 408 | 417 | 417 | 320 |
| 2005 | 442 | 416 | 439 | 429 | 439 | 436 | 441 | 448 | 441 | 449 | 447 | 348 |

**série désaisonnalisée**




La représentation de la série désaisonnalisée montre clairement l'élimination de l'effet saisonnier. La série désaisonnalisée fait apparaître quelques irrégularités dues à la composante résiduelle.

## 5.8. EXERCICES D'APPLICATION.

### 5.8.1. Exercice.

Le tableau suivant indique, pour les années 1995 à 2005, la production annuelle de céréales en millions de quintaux.

| **Année** | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Production** | 50 | 36,5 | 43 | 44,5 | 38,9 | 38,1 | 32,6 | 38,7 | 41,7 | 41,1 | 33,8 |

a) Représenter graphiquement la série brute. Quelle est la nature du trend ?
b) Lisser la série brute par la méthode des moyennes mobiles d'ordre 3 puis d'ordre 5. Représenter graphiquement les deux séries des moyennes mobiles et interpréter les graphiques obtenus.
c) Lisser la série brute, dans la cas d'un modèle multiplicatif, par la méthode des moyennes mobiles d'ordre 4. Représenter graphiquement la série des moyennes mobiles et interpréter.
d) Lisser la série brute par la méthode exponentielle avec un coefficient de lissage de 0,7 puis 0,1. Représenter graphiquement les deux séries lissées et interpréter.
e) Déterminer l'équation du trend.

**Solution** : On ne donnera que la réponse à la question e

L'équation du trend est : $\hat{y}_t = - 0,54\ t + 42,95$.

### 5.8.2. Exercice.

Au cours des deux exercices 2004, 2005, les chiffres d'affaires mensuels d'une entreprise de transports ont été les suivants :

| ans | jan | fév | mar | avril | mai | juin | juil | août | Sept | oct | nov | déc |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2004 | 50 | 46 | 64 | 65 | 63 | 70 | 85 | 63 | 59 | 56 | 49 | 56 |
| 2005 | 54 | 51 | 69 | 71 | 70 | 78 | 93 | 70 | 65 | 62 | 54 | 61 |

a) Lisser la série, selon le modèle multiplicatif, par la méthode des moyennes mobiles d'ordre 12.
b) Représenter graphiquement la série lissée. Interpréter.
c) Déterminer l'équation du trend.

**Solution** : On ne donnera que la réponse à la question c.

L'équation du trend est : $\hat{y}_t = 0{,}5\ t + 51{,}6$.

### 5.8.3. Exercice.

Considérons la production trimestrielle, en tonnes, durant 5 années de l'entreprise SATAM.

| Années | Trimestre 1 | Trimestre 2 | Trimestre 3 | Trimestre 4 |
|---|---|---|---|---|
| 1 | 920 | 1114 | 1310 | 1047 |
| 2 | 953 | 1241 | 1468 | 1183 |
| 3 | 1002 | 1343 | 1571 | 1314 |
| 4 | 1128 | 1544 | 1747 | 1446 |
| 5 | 1257 | 1589 | 1911 | 1465 |

a) Représenter graphiquement la série brute. Quelle est la nature du trend ? Juger les fluctuations aléatoires et l'effet saisonnier.
b) Lisser la série brute, selon un modèle multiplicatif, par la méthode des moyennes mobiles. Représenter graphiquement la série des moyennes mobiles et interpréter.
c) Déterminer l'équation du trend.
d) Calculer les coefficients saisonniers.
e) Désaisonnaliser la série chronologique. Représenter graphiquement la série désaisonnalisée et interpréter.
f) Calculer les prévisions des ventes trimestrielles pour l'année 6.

**Solution** : On ne donnera que la réponse à la question c, d, et f.

c) L'équation du trend est : $\hat{y}_t = 31{,}74\ t + 994{,}36$ ;

d)

| cs | 0,8254 | 1,0424 | 1,1925 | 0,9397 |
|---|---|---|---|---|

f)

| Trimestres année 6 | Prévisions |
|---|---|

| | |
|---|---|
| 1er trimestre | 1371 |
| 2ème trimestre | 1765 |
| 3ème trimestre | 2056 |
| 4ème trimestre | 1650 |

### 5.8.4. Exercice.

Le tableau ci-dessous indique les ventes mensuelles, en millions de dirhams, pendant les années 1998 à 2005, de l'entreprise MOTEL :

| ans | jan | fév | mars | avr | mai | juin | juil | août | sept | oct | nov | Déc |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1998 | 12,63 | 11,72 | 13,43 | 12,53 | 13,29 | 13,27 | 12,36 | 13,27 | 13,10 | 13,86 | 13,39 | 15,38 |
| 1999 | 11,84 | 11,74 | 12,74 | 13,40 | 14,85 | 13,81 | 13,40 | 13,45 | 13,62 | 14,82 | 14,01 | 16,91 |
| 2000 | 13,05 | 12,33 | 13,96 | 14,17 | 14,66 | 14,58 | 14,38 | 14,18 | 14,08 | 14,95 | 13,96 | 16,44 |
| 2001 | 12,34 | 12,06 | 13,54 | 14,32 | 14,25 | 14,66 | 14,39 | 13,90 | 14,14 | 14,66 | 14,53 | 17,87 |
| 2002 | 13,15 | 12,64 | 14,57 | 15,49 | 15,33 | 15,60 | 15,26 | 15,48 | 15,76 | 15,68 | 15,75 | 19,12 |
| 2003 | 13,73 | 13,55 | 15,72 | 14,89 | 16,11 | 16,58 | 15,38 | 16,19 | 15,58 | 16,13 | 16,49 | 19,38 |
| 2004 | 14,74 | 14,06 | 15,79 | 16,44 | 17,20 | 17,11 | 16,86 | 17,49 | 16,37 | 16,95 | 17,13 | 19,84 |
| 2005 | 15,29 | 13,78 | 15,55 | 16,27 | 17,36 | 16,60 | 16,60 | 17,00 | 16,33 | 17,36 | 17,04 | 21,17 |

a) Représenter graphiquement la série brute. Quelle est la nature du trend ? Juger les fluctuations aléatoires et l'effet saisonnier.
b) Lisser la série brute, selon le modèle multiplicatif, par la méthode des moyennes mobiles. Représenter graphiquement la série des moyennes mobiles et interpréter.
c) Déterminer l'équation du trend.
d) Calculer les coefficients saisonniers.
e) Désaisonnaliser la série chronologique. Représenter graphiquement la série désaisonnalisée et interpréter.
f) Calculer les prévisions des ventes mensuelles pour l'année 2007.

**Solution** : On ne donnera que la réponse à la question c, d, et f.

c) L'équation du trend est : $\hat{y}_t = 0{,}04\ t + 11{,}62$

d)

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **cs** | 0,91 | 0,86 | 0,97 | 0,99 | 1,03 | 1,02 | 0,99 | 1,00 | 0,99 | 1,03 | 1,01 | 1,20 |

f)

| Année 2007 | Prévisions |
|---|---|
| Janvier | 14,31 |

| | |
|---|---|
| Février | 13,66 |
| Mars | 15,44 |
| Avril | 15,71 |
| Mai | 16,45 |
| Juin | 16,32 |
| Juillet | 15,82 |
| Août | 16,12 |
| Septembre | 15,88 |
| Octobre | 16,61 |
| Novembre | 16,29 |
| Décembre | 19,44 |

### 5.8.5. Exercice.

L'évolution du chiffre d'affaires trimestriel (en milliers de dirhams) d'une entreprise commerciale a été la suivante, au cours de trois années consécutives :

| **Trimestres** | **2003** | **2004** | **2005** |
|---|---|---|---|
| $1^{er}$ trimestre | 880 | 810 | 740 |
| $2^{ème}$ trimestre | 960 | 880 | 800 |
| $3^{ème}$ trimestre | 1030 | 950 | 960 |
| $4^{ème}$ trimestre | 920 | 840 | 760 |

a) Représenter graphiquement la série brute. Quelle est la nature du trend ? Juger les fluctuations aléatoires et l'effet saisonnier.
b) Lisser la série brute, selon le modèle multiplicatif par la méthode des moyennes mobiles d'ordre 4. Représenter graphiquement la série des moyennes mobiles et interpréter.
c) Déterminer l'équation du trend.
d) Calculer les coefficients saisonniers.
e) Désaisonnaliser la série chronologique. Représenter graphiquement la série désaisonnalisée et interpréter.
f) Calculer les prévisions des chiffres d'affaires trimestriels pour l'année 2008.

**Solution** : On ne donnera que la réponse à la question c, d, et f.

c) L'équation du trend est : $\hat{y}_t = -12{,}83\, t + 960{,}91$

d)

| **Trimestres** | **Cs** |
|---|---|
| $1^{er}$ trimestre | 0,9023 |
| $2^{ème}$ trimestre | 0,9943 |

| | |
|---|---|
| 3ème trimestre | 1,1261 |
| 4ème trimestre | 0,9773 |

f)

| **Trimestres année 2008** | **Prévisions** |
|---|---|
| 1er trimestre | 624 |
| 2ème trimestre | 675 |
| 3ème trimestre | 750 |
| 4ème trimestre | 638 |

### 5.8.6. Exercice.

Les ventes quotidiennes d'une société commerciale sont consignées dans le tableau ci-dessous :

| **Jours** | **Semaine 1** | **Semaine 2** | **Semaine 3** | **Semaine 4** |
|---|---|---|---|---|
| Lundi | 43 | 51 | 40 | 64 |
| Mardi | 45 | 41 | 57 | 58 |
| Mercredi | 22 | 37 | 30 | 33 |
| Jeudi | 25 | 22 | 33 | 38 |
| Vendredi | 31 | 25 | 37 | 25 |

a) Représenter graphiquement la série brute. Quelle est la nature du trend ? Juger les fluctuations aléatoires et l'effet saisonnier.
b) Lisser la série brute par la méthode des moyennes mobiles. Représenter graphiquement la série des moyennes mobiles et interpréter.
c) Déterminer l'équation du trend.
d) Calculer les coefficients saisonniers.
e) Désaisonnaliser la série chronologique. Représenter graphiquement la série désaisonnalisée et interpréter.
f) Calculer les prévisions des ventes pour la cinquième semaine et pour la sixième semaine.

**Solution** : On ne donnera que la réponse à la question c, d, et f.

c) L'équation du trend est : $\hat{y}_t = 0{,}85\ t + 29{,}50$.

d)

| **Jours** | **Cs** |
|---|---|
| Lundi | 1,3395 |
| Mardi | 1,3139 |
| Mercredi | 0,8015 |
| Jeudi | 0,7217 |

| | |
|---|---|
| Vendredi | 0,8235 |

f)

| **Semaines 5 et 6** | **Prévisions** |
|---|---|
| Lundi | 63 |
| Mardi | 63 |
| Mercredi | 39 |
| Jeudi | 36 |
| Vendredi | 42 |
| Lundi | 69 |
| Mardi | 69 |
| Mercredi | 43 |
| Jeudi | 39 |
| Vendredi | 45 |

### 5.8.7. Exercice.

Le tableau suivant donne l'évolution trimestrielle des exportations par tonne denrées pour une entreprise donnée.

| **Année** | **Trimestre 1** | **Trimestre 2** | **Trimestre 3** | **Trimestre 4** |
|---|---|---|---|---|
| 2002 | 185 | 155 | 246 | 195 |
| 2003 | 315 | 285 | 354 | 312 |
| 2004 | 421 | 400 | 478 | 428 |
| 2005 | 553 | 520 | 602 | 545 |

a) Déterminer les coefficients saisonniers ;
b) Donner des prévisions des exportations pour les quatre trimestres de l'année 2006.

**Solution** :
a)

| **Année** | **Cs** |
|---|---|
| **Trimestre 1** | 1,1202 |
| **Trimestre 2** | 0,9236 |
| **Trimestre 3** | 1,0854 |
| **Trimestre 4** | 0,8708 |

b)

| **Année 2006** | **Prévisions** |
|---|---|
| **Trimestre 1** | 692 |
| **Trimestre 2** | 597 |
| **Trimestre 3** | 733 |
| **Trimestre 4** | 613 |

### 5.8.8. Exercice.

Le tableau ci-dessous indique la quantité mensuelle de marchandises transportées, en tonnes, pendant les années 1998 à 2005.

| ans | jan | fév | mars | avr | mai | juin | juil | août | sept | oct | nov | Déc |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1998 | 3661 | 2834 | 2999 | 3152 | 3977 | 3295 | 3807 | 3307 | 3312 | 4317 | 3139 | 2700 |
| 1999 | 3562 | 2911 | 2868 | 2912 | 3678 | 2606 | 2969 | 3149 | 3364 | 4156 | 3139 | 2672 |
| 2000 | 3351 | 2730 | 2801 | 2957 | 3883 | 3204 | 3758 | 3229 | 3153 | 4024 | 2797 | 2413 |
| 2001 | 2967 | 2462 | 2412 | 2445 | 3345 | 2730 | 3251 | 2708 | 2711 | 3629 | 2685 | 2518 |
| 2002 | 2505 | 2556 | 3256 | 2757 | 3754 | 3052 | 3015 | 3883 | 3148 | 3282 | 3758 | 2669 |
| 2003 | 2713 | 2751 | 3517 | 2971 | 3835 | 3143 | 2397 | 3700 | 3155 | 3284 | 3740 | 2641 |
| 2004 | 2565 | 2616 | 3446 | 2696 | 3558 | 2959 | 2708 | 3737 | 2849 | 2920 | 3223 | 2221 |
| 2005 | 2164 | 2108 | 2702 | 2105 | 2729 | 2489 | 2138 | 3146 | 2570 | 2733 | 2462 | 2188 |

a) Représenter graphiquement la série brute. Quelle est la nature du trend ? Juger les fluctuations aléatoires et l'effet saisonnier.
b) Lisser la série brute, selon un modèle multiplicatif, par la méthode des moyennes mobiles. Représenter graphiquement la série des moyennes mobiles et interpréter.
c) Déterminer l'équation du trend.
d) Calculer les coefficients saisonniers.
e) Désaisonnaliser la série chronologique. Représenter graphiquement la série désaisonnalisée et interpréter.
f) Calculer les prévisions pour les années 2006 et 2007.

**Solution** : On ne répondra qu'aux questions c, d et f.

c) L'équation du trend est : $\hat{y}_t = -4{,}92\ t + 3035{,}97$

d)

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **cs** | 0,9540 | 0,8572 | 0,9877 | 0,9018 | 1,1817 | 0,9681 | 0,9872 | 1,1152 | 1,0042 | 1,1703 | 1,0391 | 0,8336 |

f)

| Années 2006 | Prévisions | | Années 2007 | Prévisions |
|---|---|---|---|---|
| Janvier 2006 | 2441 | | Janvier | 2385 |
| Février | 2189 | | Février | 2139 |
| Mars | 2518 | | Mars | 2459 |
| Avril | 2294 | | Avril | 2241 |

| | | | | |
|---|---|---|---|---|
| Mai | 3000 | | Mai | 2931 |
| Juin | 2453 | | Juin | 2396 |
| Juillet | 2497 | | Juillet | 2439 |
| Août | 2815 | | Août | 2749 |
| Septembre | 2530 | | Septembre | 2471 |
| Octobre | 2943 | | Octobre | 2874 |
| Novembre | 2608 | | Novembre | 2546 |
| Décembre | 2088 | | Décembre | 2039 |

### 5.8.9. Exercice.

La série chronologique définie par le tableau ci-après représente l'évolution, de 2002 à 2005, du nombre trimestriel de mariages enregistrés dans un pays donné (données brutes en milliers).

| **Années** | **Trimestres** | | | |
|---|---|---|---|---|
| | **1er trimestre** | **2ème trimestre** | **3ème trimestre** | **4ème trimestre** |
| 2002 | 64 | 82 | 76 | 68 |
| 2003 | 60 | 80 | 70 | 66 |
| 2004 | 58 | 76 | 66 | 65 |
| 2005 | 57 | 73 | 64 | 63 |

a) Déterminer l'équation du Trend linéaire ;
b) Désaisonnaliser la série brute ;
c) Calculer les prévisions pour l'année 2007.

**Solution** : a) L'équation du trend est : $\hat{y}_t = -0{,}60\ t + 73{,}08$

c)

| **Année 2007** | **Prévisions** |
|---|---|
| **Trimestre 1** | 52 |
| **Trimestre 2** | 68 |
| **Trimestre 3** | 60 |
| **Trimestre 4** | 57 |

### 5.8.10. Exercice.

Le tableau suivant donne, pour 15 trimestres consécutifs, les valeurs des deux variables suivantes :

X : l'indice d'offre d'emploi.
Y : le taux de chômage.

| Années | Trimestres | X | Y |
|---|---|---|---|
| 2002 | 1 | 159 | 8,40 |
| | 2 | 154 | 8,50 |
| | 3 | 161 | 8,40 |
| | 4 | 187 | 8,16 |
| 2003 | 1 | 175 | 7,96 |
| | 2 | 186 | 7,70 |
| | 3 | 198 | 7,13 |
| | 4 | 196 | 7,23 |
| 2004 | 1 | 204 | 7,50 |
| | 2 | 195 | 7,70 |
| | 3 | 204 | 7,50 |
| | 4 | 210 | 7,40 |
| 2005 | 1 | 231 | 7,30 |
| | 2 | 221 | 7,15 |
| | 3 | 241 | 7,13 |
| | 4 | 252 | 7,11 |

Etudier les deux séries chronologiques : l'indice d'offre d'emploi et le taux de chômage et déterminer quels devraient être l'indice d'offre d'emploi et le taux de chômage pour les 4 trimestres de 2006.

Indice de l'offre d'emploi :

Equation du Trend : $\hat{y}_t = 5{,}76t + 149{,}43$

Coefficients saisonniers :

| Trimestres | CS |
|---|---|
| Trimestre 1 | 1,0141 |
| Trimestre 2 | 0,9692 |
| Trimestre 3 | 0,9991 |
| Trimestre 4 | 1,0176 |

Prévisions 2006 :

| Trimestres | Prévisions |
|---|---|
| Trimestre 1 | 251 |
| Trimestre 2 | 245 |
| Trimestre 3 | 259 |
| Trimestre 4 | 269 |

Taux de chômage :

Equation du Trend : - 0,09 t + 8,41
Coefficients saisonniers :

| Trimestres | CS |
|---|---|
| Trimestre 1 | 1,0035 |
| Trimestre 2 | 1,0116 |
| Trimestre 3 | 0,9947 |
| Trimestre 4 | 0,9902 |

Prévisions 2006 :

| Trimestres | Prévisions |
|---|---|
| Trimestre 1 | 6,9 |
| Trimestre 2 | 6,9 |
| Trimestre 3 | 6,7 |
| Trimestre 4 | 6,5 |

# CHAPITRE 6
# INDICES STATISTIQUES

Les indices sont des instruments de mesure de l'évolution des grandeurs, ils sont habituellement exprimés en pourcentage.

Un indice est donc destiné à comparer deux grandeurs ou les valeurs d'une même grandeur à deux moments ou dans deux espaces différents. Ces grandeurs peuvent être soit simples, et l'indice est dit élémentaire ou simple, soit des grandeurs complexes, et l'indice est dit synthétique.

## 6.1. LES INDICES ELEMENTAIRES.

### 6.1.1. Définition.

Considérons une grandeur simple, G, mesurée par un nombre qui caractérise directement une situation, si nous notons Go la valeur de la grandeur G à la date 0, appelée date ou période de base ou de référence et Gt sa valeur à la date t, appelée date ou période courante, l'indice élémentaire de la grandeur G à la date t, par rapport à la date 0 est :

$$I_{t/0} = \frac{G_t}{G_0} \times 100$$

**Exemple 1** : On considère les prix successifs du Kg de sucre, à des dates différentes :

| **Dates** | **1999** | **2000** | **2001** | **2002** |
|---|---|---|---|---|
| **Prix (DH/kg)** | 4,50 | 4,65 | 4,97 | 5,12 |

On pourra calculer les indices du prix du sucre, selon les périodes, avec comme date de référence 1999, on a :

| **Dates** | **1999** | **2000** | **2001** | **2002** |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| **Indices des prix avec 1999 base** | 100 | 103,33 | 110,44 | 113,78 |

Cette façon de faire permet de remplacer la suite des prix du sucre, à différentes périodes, par la suite des indices, plus facile à manipuler.

Pour mieux comprendre cette affirmation on considère l'évolution du prix de la tonne du fuel domestique sur plusieurs années :

**Exemple 2** : On donne l'évolution du prix du fuel domestique entre 1999 et 2005. On demande de calculer les indices du prix du fuel domestique pour les mêmes dates avec comme date de base 1999.

| **Dates** | **1999** | **2001** | **2003** | **2005** |
|---|---|---|---|---|
| **Prix (DH/t)** | 4926,84 | 5237,77 | 5876,34 | 6735,98 |
| **Indices des prix avec 1999 base** | 100 | 106,31 | 119,27 | 136,72 |

Ainsi, au lieu de manipuler des prix qui sont des nombres à plusieurs chiffres, on se contente, avec les indices, de ne manipuler que des pourcentages qui sont faciles à transcrire et à mémoriser. D'où l'intérêt considérable des indices.

**Remarques :**

1) Il ne faut jamais oublier qu'un indice est un pourcentage. Bien qu'il soit noté conventionnellement, par exemple, 121,67 ou 95,32 il faut avoir, constamment à l'esprit qu'en fait il s'agit de 121,67% c'est-à-dire 1,2167 ou 95,32% c'est-à-dire 0,9532.

2) Lorsqu'on manipule des indices et conformément à la première remarque, il faut, selon le cas, utiliser la notation en pourcentage (121,67% ou 94,32%) ou la notation en décimale (1,2167 ou 0,9532).

3) Pour nous résumer et être le plus explicite possible, il est important de comprendre et d'accepter les notations suivantes, même si elles paraissent, à première vue, incorrectes :

- Pour l'addition d'indices :
121,67 + 95,32 = 216,99 = 216,99% = 2,1699
- Pour la multiplication d'indices :
121,67 x 95,32 = 115,98 = 115,98% = 1,1598

### 6.1.2. Propriétés des indices élémentaires.

#### 6.1.2.1. Dimension d'un indice.

Un indice n'a pas de dimension du fait que, par définition, il est le rapport d'une même grandeur à deux dates différentes ou dans deux endroits différents.

#### 6.1.2.2. Indicateur de l'évolution de la grandeur.

Un indice simple est un indicateur du sens de l'évolution de la grandeur à laquelle il est rattaché, en effet

- si I > 100 la grandeur a accusé une augmentation ;
- si I = 100 la grandeur n'a pas varié ;
- si I < 100 la grandeur a accusé une diminution.

#### 6.1.2.3. Propriété d'identité.

La propriété d'identité s'exprime par la relation simple suivante :

$$I_{o/o} = \frac{G_0}{G_0} \times 100 = 100\ \%$$

#### 6.1.2.4. Propriété de réversibilité.

La propriété de réversibilité s'exprime par la relation :

$$I_{0/t} = \frac{1}{I_{t/0}}$$

En effet, on a : $I_{0/t} = \frac{G_0}{G_t} = \frac{1}{\frac{G_t}{G_0}} = \frac{1}{I_{t/0}}$

#### 6.1.2.5. Propriété de circularité.

La propriété de circularité s'exprime par la relation suivante :

$$I_{t/0} = I_{t/t'} \times I_{t'/0}$$

En effet, on a :

$$I_{t/0} = \frac{G_t}{G_0} = \frac{G_t}{G_{t'}} \times \frac{G_{t'}}{G_0} = I_{t/t'} \times I_{t'/0}$$

Cette propriété de circularité est essentielle pour les indices simples car elle permet :

- de changer de date de base, c'est-à-dire de date de référence ;

- de comparer des indices ayant une même date de base ;
- de comparer des indices ayant des dates de base différentes ;
- de calculer l'indice simple moyen entre deux périodes.

#### 6.1.2.5.1. Changement de date de base.

Le changement de date de référence est une opération courante dans la manipulation des indices, nous en donnerons plusieurs exemples dans les paragraphes suivants.

Pour le moment nous allons expliquer, par un exemple simple, comment procéder.

**Exemple 3** : On considère les indices du prix du sucre de l'exemple 1 et l'on voudrait prendre comme date de référence 2000 au lieu de 1999.

Rappelons le tableau de l'exemple 1.

| **Dates** | **1999** | **2000** | **2001** | **2002** |
|---|---|---|---|---|
| **Indices des prix avec 1999 base** | 100 | 103,33 | 110,44 | 113,78 |

Pour changer la date de base, nous utilisons la propriété de circularité des indices simples et nous essayons de calculer l'indice des prix du sucre avec comme date de base 2000 à partir des indices du prix du sucre ayant comme date de base 1999.

$$I_{t/2000} = \frac{G_t}{G_{2000}} = \frac{G_t}{G_{1999}} \times \frac{G_{1999}}{G_{2000}} = \frac{\frac{G_t}{G_{1999}}}{\frac{G_{2000}}{G_{1999}}} = \frac{I_{t/1999}}{I_{2000/1999}}$$

Nous pouvons alors dresser le tableau des indices du prix du sucre, avec comme base de référence 2000, à partir du tableau des indices du prix ayant comme base 1999.

| **Dates** | **1999** | **2000** | **2001** | **2002** |
|---|---|---|---|---|
| **Indices des prix avec 1999 base** | 100 | 103,33 | 110,44 | 113,78 |
| **Indices des prix avec 2000 base** | 96,78 | 100 | 106,88 | 110,11 |

#### 6.1.2.5.2. Comparaison de deux indices ayant même date de base.

La propriété de circularité permet aussi de comparer deux indices, ayant les mêmes dates de base, en effet considérons l'exemple suivant :

**Exemple 4** : On considère deux indices relatifs à deux grandeurs différentes, ayant la même date de référence 2001 et ayant les valeurs suivantes ; on demande lequel des 2 indices a augmenté le plus entre 2003 et 2006.

| **Dates** | **2003** | **2006** |
|---|---|---|
| **Indice $I_1$ ayant 2001 comme base** | 124 | 145 |
| **Indice $I_2$ ayant 2001 comme base** | 117 | 137 |

Pour faire une telle comparaison, il est nécessaire de changer de base de référence et de prendre comme nouvelle base, 2003. Le tableau précédant devient dans ce cas :

| **Dates** | **2003** | **2006** |
|---|---|---|
| **Indice $I_1$ ayant 2003 comme base** | 100 | 116,94 |
| **Indice $I_2$ ayant 2003 comme base** | 100 | 117,09 |

Pour calculer la valeur des indices, en 2006, on utilise la propriété de circularité des indices, à savoir :

$$I_{2006/2003} = \frac{I_{2006/2001}}{I_{2003/2001}}$$

On voit sur ce nouveau tableau que le deuxième indice a augmenté plus que le premier.

#### 6.1.2.5.3. Comparaison de deux indices ayant des dates de base différentes.

La propriété de circularité permet aussi de comparer deux indices, ayant des dates de base différentes, en effet considérons l'exemple suivant :

**Exemple 5** : On considère les indices des quantités consommées d'orge et de blé, $I_o$ et $I_b$ et on demande laquelle de ces quantités a subi la plus forte augmentation, entre 2000 et 2004, sachant que les indices $I_o$ et $I_b$ qui ont des dates de base différentes ont les valeurs suivantes :

| **Dates** | **2000** | **2004** |
|---|---|---|
| **Quantités d'orge consommées en Kg** | 2 345 965,00 | 2 607 070,90 |
| **Indice $I_o$ ayant 1998 comme base** | 124,87 | 138,77 |
| **Quantités de blé consommées en Kg** | 1 634 961,00 | 1 729 461,75 |
| **Indice $I_b$ ayant 1997 comme base** | 132,65 | 140,32 |

Afin de faire une telle comparaison, il est nécessaire de changer, pour les 2 indices, les dates de base de référence et de prendre comme nouvelle base, 2000. Le tableau précèdent devient dans ce cas :

| **Dates** | **2000** | **2004** |
|---|---|---|

| | | |
|---|---|---|
| **Quantités d'orge consommées en Kg** | 2 345 965,00 | 2 607 070,90 |
| **Indice $I_o$ ayant 2000 comme base** | 100 | 111,13 |
| **Quantités de blé consommées en Kg** | 1 634 961,00 | 1 729 461,75 |
| **Indice $I_b$ ayant 2000 comme base** | 100 | 105,78 |

Pour calculer la valeur des indices, en 2004 avec 2000 comme date de référence, on utilise la propriété de circularité des indices, à savoir :

$$I_{o2004/2000} = \frac{I_{2004/1998}}{I_{2000/1998}} \quad \text{et} \quad I_{b2004/2000} = \frac{I_{2004/1997}}{I_{2000/1997}}$$

On voit, sur ce nouveau tableau, que le premier indice a augmenté plus que le deuxième ; c'est-à-dire qu'entre 2000 et 2004, la quantité consommée d'orge a augmenté, en pourcentage, plus que celle du blé.

### 6.1.2.5.4. Détermination de l'indice simple moyen.

La détermination d'un indice simple moyen est nécessaire lorsque des données relatives à certaines périodes sont manquantes.

**Exemple 6** : En effet prenons l'exemple 2 relatif aux indices du prix du fuel domestique.

| **Dates** | **1999** | **2001** | **2003** | **2005** |
|---|---|---|---|---|
| **Prix (DH/t)** | 4926,84 | 5237,77 | 5876,34 | 6735,98 |
| **Indices des prix avec 1999 base** | 100 | 106,31 | 119,27 | 136,72 |

Dans cet exemple, les indices de prix relatifs aux années 2000, 2002 et 2004 manquent ; la question qui se pose est la suivante : Comment déterminer les indices de prix des années 2000, 2002 et 2004 ?

Pour ce faire, nous devons faire une hypothèse vraisemblable ; elle consiste à supposer qu'entre 1999 et 2001, le prix du fuel domestique a augmenté régulièrement, c'est-à-dire qu'il a subi le même taux d'augmentation entre 2000 et 2001 qu'entre 1999 et 2000.

Soit t ce taux moyen d'augmentation annuel du prix du fuel domestique entre 2000 et 2001 puis entre 1999 et 2000.

On a, si l'on se rappelle qu'un indice de prix est justement le taux de variation du prix entre deux périodes et qu'il est donné en pourcentage :

$I_{2001/1999} = I_{2001/2000} \times I_{2000/1999} = t^2$

On a supposé que $I_{2001/2000} = I_{2000/1999} = t$

106,31% = 1,0631 = $t^2$ => t = 1,03107 = 103,11%

Le taux de variation du prix du fuel, entre 1999 et 2000 qui est supposé égal au taux de variation du prix entre 2000 et 2001 est égal à 103,11%.

## 6.2. LES INDICES SYNTHETIQUES.

Nous n'avons considéré, dans le paragraphe précédent, pour étudier les indices simples, que le cas simple et particulier de grandeurs simples, comme prix, quantités, etc. Or habituellement, dans la vie courante des affaires, on est amené à considérer, en même temps, plusieurs grandeurs et à essayer de discuter de la variation de l'ensemble de ces grandeurs.

Considérons alors n grandeurs simples notées $G_i$ (avec i = 1,…., n) et posons $G_{i0}$ (toujours avec i = 1,…., n) les valeurs des grandeurs simples pour les différentes grandeurs i relevées à la date 0 et $G_{it}$ les valeurs des mêmes grandeurs simples i relevés à la date t. Pour étudier l'évolution de l'ensemble des grandeurs $G_i$, nous avons besoin de définir des indices globaux ou synthétiques.

Ces grandeurs peuvent être des prix $P_i$, des quantités $Q_i$, des valeurs globales $Q_iP_i$, etc. Elles ont donc toutes la même dimension (Kg, m, $m^3$, l, DH, etc.).

On peut définir plusieurs types d'indices moyens synthétiques relatifs à l'ensemble des grandeurs i observées aux dates 0 et t :

- un indice des moyennes ;
- un indice moyenne des indices

### 6.2.1. Indice synthétique des moyennes.

#### 6.2.1.1. Définition.

L'expression de l'indice des moyennes est donnée, par définition, par la formule suivante :

$$I_{t/0} = \frac{\frac{\sum_{i=1}^{n} G_{it}}{n}}{\frac{\sum_{i=1}^{n} G_{i0}}{n}} = \frac{\sum_{i=1}^{n} G_{it}}{\sum_{i=1}^{n} G_{i0}}$$

Par un tel indice des moyennes des grandeurs entre l'instant t et la date de référence, on estime donner une image de l'évolution de l'ensemble des grandeurs $G_i$.

#### 6.2.1.2. Propriétés de l'indice des moyennes.

L'indice synthétique simple des moyennes possède la propriété :

- de réversibilité ;
- de circularité.

Nous pouvons montrer cela dans les exemples 7 et 8 suivants.

**Exemple 7** : **Propriété de réversibilité de l'indice des moyennes :** Reprenons l'exemple 5 relatif aux quantités consommées d'orge et de blé entre 1998 et 2004 et posons-nous la question suivante : Comment évaluer l'évolution des quantités consommées de céréales entre 1998 et 2004 ?

Rappelons le tableau qui nous a servi pour les calculs de l'exemple 5.

| **Dates** | **2000** | **2004** |
|---|---|---|
| **Quantités d'orge consommées en Kg** | 2 345 965,00 | 2 607 070,90 |
| **Indice $I_o$ ayant 2000 comme base** | 100 | 111,13 |
| **Quantités de blé consommées en Kg** | 1 634 961,00 | 1 729 461,75 |
| **Indice $I_b$ ayant 2000 comme base** | 100 | 105,78 |

Calculons, pour ce cas, l'indice des moyennes.

| **Dates** | **2000** | **2004** |
|---|---|---|
| **Quantités d'orge consommées en Kg** | 2 345 965,00 | 2 607 070,90 |
| **Quantités de blé consommées en Kg** | 1 634 961,00 | 1 729 461,75 |
| **Sommes des quantités** | 3 980 926,00 | 4 336 532,65 |
| **$I_{2004\ /\ 2000}$ (moyenne des quantités)** | 100 x 4 336 532,65 / 3 980 926,00 = 108,93 | |
| **$I_{2000\ /\ 2004}$ (moyenne des quantités)** | 100 x 3 980 926,00 / 4 336 532,65 = 91,80 | |

On a bien (1/1,0893) = 0,9180

On peut montrer, d'une façon générale, que l'indice synthétique simple des moyennes est réversible, en effet, cet indice se calcule comme suit :

$$I_{2000/2004} = \frac{\sum_{i=1}^{n} G_{i\,2000}}{\sum_{i=1}^{n} G_{i\,2004}} = \frac{1}{I_{2004/2000}}$$

**Exemple 8** : **Propriété de circularité de l'indice des moyennes :** Reprenons donc l'exemple 5 relatif aux quantités consommées d'orge et de blé, pour 2000, 2004 et 2006 et calculons les différents indices synthétiques simples.

Rappelons le tableau qui nous a servi pour les calculs de l'exemple 7.

**Indice des moyennes :**

| **Dates** | **2000** | **2004** | **2006** |
|---|---|---|---|
| **Quantités d'orge consommées en Kg** | 2 345 965,00 | 2 607 070,90 | 2 876 554,12 |
| **Quantités de blé consommées en Kg** | 1 634 961,00 | 1 729 461,75 | 2 347 885,23 |
| **Sommes des quantités** | 3 980 926,00 | 4 336 532,65 | 5 224 439,35 |
| $\mathbf{I_{2006/2000}}$ **(moyenne des quantités)** | 5 224 439,35 / 3 980 926,00 = 131,24 | | |
| $\mathbf{I_{2006/2004}}$ **(moyenne des quantités)** | - - - | 5 224 439,35 / 4 336 532,65 = 120,48 | |
| $\mathbf{I_{2004/2000}}$ **(moyenne des quantités)** | 4 336 532,65 / 3 980 926,00 = 108,93 | - - - | |

0n voit bien que : $I_{2006/2004} \times I_{2004/2000} = 1{,}2048 \times 1{,}0893 = 131{,}24$

Et que : $I_{2006/2000} = 131{,}24$

Ce qui fait : $I_{2006/2000} = I_{2006/2004} \times I_{2004/2000}$

L'indice synthétique simple des moyennes des grandeurs possède la propriété de circularité.

On peut montrer, d'une façon générale, que l'indice synthétique simple des moyennes possède la propriété de circularité, en effet, cet indice se calcule comme suit :

$$I_{2006/2000} = \frac{\sum_{i=1}^{n} G_{i\,2006}}{\sum_{i=1}^{n} G_{i\,2000}} = \frac{\sum_{i=1}^{n} G_{i\,2006}}{\sum_{i=1}^{n} G_{i\,2004}} \times \frac{\sum_{i=1}^{n} G_{i\,2004}}{\sum_{i=1}^{n} G_{i\,2000}}$$

$$= I_{2006/2004} \text{ x } I_{2004/2000}$$

### 6.2.2. Indice synthétique moyenne des indices.

#### 6.2.2.1. Définition.

L'expression de l'indice synthétique moyenne des indices est donnée, par définition, par la formule suivante :

$$I_{t/0} = \frac{1}{n}\sum_{i=1}^{n} \frac{G_{it}}{G_{i0}} = \frac{\sum_{i=1}^{n} I_{i\,t/0}}{n}$$

Par un tel indice moyenne des indices des grandeurs Gi entre l'instant t et la date de référence, on estime donner une image de l'évolution de l'ensemble des grandeurs $G_i$.

#### 6.2.2.2. Propriétés de l'indice synthétique moyenne des indices.

L'indice synthétique moyenne des indices ne possède :

- ni la propriété de réversibilité ;
- ni la propriété de circularité.

Nous pouvons montrer cela dans les exemples suivants.

**Exemple 9** : Reprenons l'exemple 5 relatif aux quantités consommées d'orge et de blé entre 2000 et 2004 et posons-nous la question suivante : Comment évaluer l'évolution des quantités consommées de céréales entre 2000 et 2004.

Rappelons le tableau qui nous a servi pour les calculs de l'exemple 5.

| **Dates** | **2000** | **2004** |
|---|---|---|
| **Quantités d'orge consommées en Kg** | 2 345 965,00 | 2 607 070,90 |
| **Indice $I_o$ ayant 2000 comme base** | 100 | 111,13 |
| **Quantités de blé consommées en Kg** | 1 634 961,00 | 1 729 461,75 |
| **Indice $I_b$ ayant 2000 comme base** | 100 | 105,78 |

Calculons, pour ce cas, l'indice moyenne des indices.

| **Dates** | **2000** | **2004** |
|---|---|---|
| **Indice $I_o$ ayant 2000 comme base** | 100 | 111,13 |
| **Indice $I_b$ ayant 2000 comme base** | 100 | 105,78 |
| **Sommes des indices** | 200 | 216,91 |
| **$I_{2004\,/\,2000}$ (moyenne des indices)** | 216,91 / 2 = 108,46 | |

Pour le calcul de l'indice synthétique moyenne des indices $I_{2000\,/\,2004}$, nous devons, d'abord, reprendre le tableau ci-dessus et calculer les indices simples avec 2004 comme date de référence :

| **Dates** | **2000** | **2004** |
|---|---|---|
| **Indice $I_o$ ayant 2004 comme base** | 89,98 | 100 |
| **Indice $I_b$ ayant 2004 comme base** | 94,54 | 100 |
| **Sommes des indices** | 184,52 | 200 |
| **$I_{2000\,/\,2004}$ (moyenne des indices)** | 184,52 / 2 = 92,26 | |

L'indice synthétique moyenne des indices n'est donc pas réversible.

On peut aussi montrer, d'une façon générale, que l'indice synthétique moyenne des indices n'est pas réversible, en effet, cet indice se calcule comme suit :

$$I_{t/0} = \frac{\sum_{i=1}^{n} I_{i\,t/0}}{n} \qquad \text{et} \qquad I_{0/t} = \frac{\sum_{i=1}^{n} I_{i\,0/t}}{n}$$

Ces deux expressions sont très différentes puisque si l'on a bien

$I_{i\,t/0} = \frac{1}{I_{i\,0/t}}$ on a, en général, $\sum_{i=1}^{n} I_{i\,t/0} \neq \sum_{i=1}^{n} I_{i\,0/t}$

Pour montrer que l'indice synthétique moyenne des indices ne possède pas la propriété de circularité, nous conservons l'exemple précédent en y ajoutant les données de l'année 2006.

**Exemple 10** : Reprenons donc l'exemple 5 relatif aux quantités consommées d'orge et de blé, pour 2000, 2004 et 2006 et calculons les différents indices synthétiques simples.

Rappelons le tableau qui nous a servi pour les calculs de l'exemple 7.

| **Dates** | **2000** | **2004** | **2006** |
|---|---|---|---|
| **Indice $I_o$ ayant 2000 comme base** | 100 | 111,13 | 122,62 |
| **Indice $I_b$ ayant 2000 comme base** | 100 | 105,78 | 143,60 |
| **Sommes des indices** | 200 | 216,91 | 266,22 |

| | | |
|---|---|---|
| **$I_{2006\ /\ 2000}$ (moyenne des indices)** | (122,62 + 143,60) / 2 = 133,11 | |
| **$I_{2006\ /\ 2004}$ (moyenne des indices)** | - - - | (110,34 + 135,75) / 2 = 123,05 (1) |
| **$I_{2004\ /\ 2000}$ (moyenne des indices)** | (111,13 + 105,78) / 2 = 108,46 | - - - |

Pour le calcul de $I_{2006\ /\ 2004}$ indice moyenne des indices, pour l'année 2006, avec comme date de référence 2004, nous devons changer de dates de base des indices du tableau, et prendre l'année 2004 comme date de référence :

| **Dates** | **2000** | **2004** | **2006** |
|---|---|---|---|
| **Indice $I_o$ ayant 2004 comme base** | 89,98 | 100 | 110,34 |
| **Indice $I_b$ ayant 2004 comme base** | 94,54 | 100 | 135,75 |
| **$I_{2006\ /\ 2004}$ (moyenne des indices)** | - - - | (110,34 + 135,75) / 2 = 123,05 | |

0n voit bien que : $I_{2006\ /\ 2004} \times I_{2004\ /\ 2000} = 1,2305 \times 1,0846$
$= 1,3346$

Et que : $I_{2006\ /\ 2000} = 1,3311$

Ce qui fait : $I_{2006\ /\ 2000} \neq I_{2006\ /\ 2004} \times I_{2004\ /\ 2000}$

L'indice synthétique moyenne des indices ne possède donc pas la propriété de circularité.

Mais de tels indices, quoique synthétiques, restent simples. On leur préfère d'autres indices plus explicites. Ce sont les indices synthétiques pondérés.

### 6.3. LES INDICES SYNTHETIQUES PONDERES.

Si les grandeurs simples $G_i$ sont de même nature (même unité) mais n'ont pas la même importance, on associe à chaque grandeur $G_i$ un poids différent. Si l'on note $\alpha_i$ le coefficient de pondération affecté à la grandeur $G_i$, la formule retenue pour le calcul de l'indice synthétique devient :

#### 6.3.1. Indice synthétique pondéré des moyennes.

$$I_{t/0} = \frac{\sum_{i=1}^{n} \alpha_i G_{it}}{\sum_{i=1}^{n} \alpha_i G_{i0}}$$

### 6.3.2. Indice synthétique pondéré des moyennes des indices.

$$I_{t/0} = \frac{\sum_{i=1}^{n} \alpha_i \left(\frac{G_{it}}{G_{i0}}\right)}{\sum_{i=1}^{n} \alpha_i}$$

Nous verrons, dans la suite du cours, que le problème le plus important qui se pose au statisticien est justement la pertinence du choix des coefficients de pondération $\alpha_i$.

## 6.4. LES PRINCIPAUX INDICES SYNTHETIQUES.

Les indices synthétiques les plus couramment utilisés sont les indices de LASPEYRES et de PAASCHE.

### 6.4.1. Indice de LASPEYRES.

L'indice de LASPEYRES adopte des coefficients de pondération de la période de base, soit $\alpha_{i0}$, il est égal à la moyenne arithmétique des indices élémentaires, pondérés par les coefficients de la période de référence. Sa formule est donc :

Pour l'indice de LASPEYRES, moyenne pondérée des grandeurs :

$$L_{t/0} = \frac{\sum_{i=1}^{n} \alpha_{i0} G_{it}}{\sum_{i=1}^{n} \alpha_{i0} G_{i0}}$$

Pour l'indice de LASPEYRES, moyenne pondérée des indices :

$$L_{t/0} = \frac{\sum_{i=1}^{n} \alpha_{i0} \left(\frac{G_{it}}{G_{i0}}\right)}{\sum_{i=1}^{n} \alpha_{i0}}$$

**REMARQUE.**

Le choix des coefficients de pondération, pour les indices LASPEYRES, ceux relatifs à la période 0, fait que les indices de LASPEYRES ne sont représentatifs de la réalité que dans la mesure où les valeurs des coefficients de pondération restent stables, avec le temps, ou varient dans les mêmes proportions ou varient très peu. Cela nous permet de comparer des indices à des dates $t_1$ et $t_2$ différentes, bien que les indices de LASPEYRES ne possèdent pas la propriété de circularité.

Dans le cas où les coefficients de pondération varient significativement beaucoup, on est en droit de parler de durée de vie d'un indice, c'est-à-dire le temps au bout duquel les coefficients de pondération ont tellement varié au point que la situation, à l'instant t, soit très différente par rapport à l'instant zéro.

On effectue, à ce moment là, pour les indices de LASPEYRES, un changement de date de référence pour prendre comme nouvelle base, la date à laquelle les coefficients de pondération ont beaucoup varié, c'est-à-dire, à l'expiration de la durée de vie de l'indice.

Mais se pose alors la question de circularité des indices de LASPEYRES pour pouvoir relier les indices ayant différentes dates de référence. Et, traditionnellement, bien que l'on sache pertinemment que les indices LASPEYRE ne possèdent pas la propriété de circularité, nous faisons comme s'ils la possédaient parce que nous ne pouvons pas faire autrement.

### 6.4.2. Indice de PAASCHE.

L'indice de PAASCHE adopte des coefficients de pondération de la période courante, soit $\alpha_{it}$, il est égal à la moyenne harmonique des indices élémentaires, pondérés par les coefficients de la période courante. Sa formule est donc :

Pour l'indice de PAASCHE, moyenne pondérée des grandeurs :

$$P_{t/0} = \frac{\sum_{i=1}^{n} \alpha_{it} G_{it}}{\sum_{i=1}^{n} \alpha_{it} G_{i0}}$$

Pour l'indice de PAASCHE, moyenne pondérée des indices :

$$P_{t/0} = \frac{\sum_{i=1}^{n} \alpha_{it}}{\sum_{i=1}^{n} \alpha_{it} \left(\frac{G_{i0}}{G_{it}}\right)}$$

**Remarque.**

Le choix des coefficients de pondération, pour les indices PAASCHE, ceux relatifs à la période t, fait que les indices de PAASCHE ne sont représentatifs de la réalité que dans la mesure où les valeurs des coefficients de pondération de l'instant t soient les mêmes que ceux des périodes antérieures.

Dans le cas contraire, on est en droit de parler de durée de vie d'un indice, c'est-à-dire le temps en deçà duquel les coefficients de pondération sont tellement différents par rapport à ceux de la période t que la situation à ce moment là ne soit pas traduite assez fidèlement par des coefficients de la période t.

On pourrait alors effectuer, à ce moment là, pour les indices de PAASCHE, un changement de date de référence pour prendre comme nouvelle base, la date à laquelle les coefficients de pondération sont très différents par rapport à ceux de la période t.

Mais se pose alors la question de circularité des indices de PAASCHE pour pouvoir relier les indices ayant différentes dates de référence. Et, traditionnellement, bien que l'on sache pertinemment que les indices PAASCHE ne possèdent pas la propriété de circularité, nous faisons comme s'ils la possédaient parce que nous ne pouvons pas faire autrement.

### 6.4.3. Relation entre indice de LASPEYRES et indice de PAASCHE.

Les indices de LASPEYRES et de PAASCHE ne satisfont ni la condition de réversibilité, ni celle de circularité, ils ont la propriété de s'échanger l'un contre l'autre lorsqu'on permute la date de référence et la date courante. En effet :

$$L_{0/t} = \frac{\sum_{i=1}^{n} \alpha_{it} \left(\frac{G_{i0}}{G_{it}}\right)}{\sum_{i=1}^{n} \alpha_{it}} = \frac{1}{P_{t/0}}$$

$$P_{0/t} = \frac{\sum_{i=1}^{n} \alpha_{i0}}{\sum_{i=1}^{n} \alpha_{i0} \left(\frac{G_{it}}{G_{i0}}\right)} = \frac{1}{L_{t/0}}$$

### 6.4.4. Indice de FISCHER.

L'indice de FISCHER est la moyenne géométrique des deux indices de PAASCHE et de LASPEYRES, soit :

$$F_{t/0} = \sqrt{L_{t/0} \times P_{t/0}}$$

$$F_{0/t} = \sqrt{L_{0/t} \times P_{0/t}} = \sqrt{\frac{1}{L_{t/0}} \times \frac{1}{P_{t/0}}}$$

$$F_{0/t} = \frac{1}{\sqrt{L_{t/0} \times P_{t/0}}} = \frac{1}{F_{t/0}}$$

L'indice de FISCHER possède, de ce fait, la propriété de réversibilité mais il ne possède pas celle de la circularité puisque ni les indices de LASPEYRES ni ceux de PAASCHE ne la possèdent.

### 6.5. L'INDICE DES PRIX A LA CONSOMMATION.

Encore appelé indice des prix de détail ou indice du coût de la vie, l'indice des prix à la consommation est calculé à partir des prix d'un ensemble d'articles spécifiques représentant les produits de consommation et les services essentiels (le «panier de la ménagère»). Il est utilisé pour mesurer les variations, dans le temps, des prix payés pour ces produits et services par un ménage type et sert de mesure la plus courante à l'inflation.

La composition de l'indice des prix à la consommation est généralement défini à la suite d'études gouvernementales sur les dépenses de ménages types. Les composantes sont pondérées en fonction de leur poids respectif dans ces dépenses.

Considérons les dépenses du ménage aux dates 0 et t. Soient $p_i$ le prix du produit i et $q_i$ la quantité achetée.

A la date 0 les dépenses du ménage en produit i sont :

$$d_{i0} = p_{i0}\, q_{i0}$$

La dépense totale à la date 0 est donc :

$$d_0 = \sum_{i=1}^{n} p_{i0} q_{i0}$$

A la date t les dépenses du ménage en produit i sont :

$$d_{it} = p_{it}\, q_{it}$$

La dépense totale à la date t est donc :

$$d_t = \sum_{i=1}^{n} p_{it} q_{it}$$

A la date t, les prix et les quantités ont varié. On peut calculer pour chaque produit :

**Indices élémentaires de prix du produit i :**

$$I_{pi\,t/0} = \frac{p_{it}}{p_{i0}}$$

**Indices élémentaires de quantité du produit i :**

$$I_{qi\,t/0} = \frac{q_{it}}{q_{i0}}$$

**Indices élémentaires de dépenses du produit i :**

$$I_{di\,t/0} = D^{i}_{t/0} = \frac{p_{it} q_{it}}{p_{i0} q_{i0}} = I_{qi\,t/0} \times I_{pi\,t/0}$$

On affecte à chaque indice élémentaire i un coefficient de pondération qui exprime la part du produit i dans les dépenses totales du ménage, cette part est égale :

A la date 0, à $\alpha_{i0} = \frac{p_{i0} q_{i0}}{\sum_{i=1}^{n} p_{i0} q_{i0}}$

A la date t, à $\alpha_{it} = \frac{p_{it} q_{it}}{\sum_{i=1}^{n} p_{it} q_{it}}$

La somme des coefficients de pondération étant égale à 1 :

Car : $\sum_{i=1}^{n} \alpha_{i0} = \sum_{i=1}^{n} \alpha_{it} = 1$

On peut dès lors écrire les indices de LASPEYRES et de PAASCHE des prix et des quantités.

### 6.5.1. Indices de prix.

**Indice LASPEYRES de prix :**

$$L_{p_{t/0}} = \sum_{i=1}^{n} \alpha_{i0} \left(\frac{p_{it}}{p_{i0}}\right) = \sum_{i=1}^{n} \frac{p_{i0} q_{i0}}{\sum_{i=1}^{n} p_{i0} q_{i0}} \times \left(\frac{p_{it}}{p_{i0}}\right)$$

$$\text{Soit :} \quad L_{p_{t/0}} = \frac{\sum_{i=1}^{n} p_{it} q_{i0}}{\sum_{i=1}^{n} p_{i0} q_{i0}}$$

**Indice PAASCHE de prix :**

$$P_{p_{t/0}} = \frac{1}{\sum_{i=1}^{n} \alpha_{it} \left(\frac{p_{i0}}{p_{it}}\right)} = \frac{1}{\sum_{i=1}^{n} \frac{p_{it} q_{it}}{\sum_{i=1}^{n} p_{it} q_{it}} \times \left(\frac{p_{i0}}{p_{it}}\right)}$$

$$\text{Soit :} \quad P_{p_{t/0}} = \frac{\sum_{i=1}^{n} p_{it} q_{it}}{\sum_{i=1}^{n} p_{i0} q_{it}}$$

### 6.5.2. Indices de quantités.

**Indice LASPEYRES de quantités :**

$$L_{q_{t/0}} = \sum_{i=1}^{n} \alpha_{i0} \left(\frac{q_{it}}{q_{i0}}\right) = \sum_{i=1}^{n} \frac{p_{i0} q_{i0}}{\sum_{i=1}^{n} p_{i0} q_{i0}} \times \left(\frac{q_{it}}{q_{i0}}\right)$$

$$\text{Soit :} \quad L_{q_{t/0}} = \frac{\sum_{i=1}^{n} q_{it} p_{i0}}{\sum_{i=1}^{n} q_{i0} p_{i0}}$$

**Indice PAASCHE de quantités :**

$$P_{q_{t/0}} = \frac{1}{\sum_{i=1}^{n} \alpha_{it} \left(\frac{q_{i0}}{q_{it}}\right)} = \frac{1}{\sum_{i=1}^{n} \frac{p_{it} q_{it}}{\sum_{i=1}^{n} p_{it} q_{it}} \times \left(\frac{q_{i0}}{q_{it}}\right)}$$

$$\text{Soit :} \quad P_{q_{t/0}} = \frac{\sum_{i=1}^{n} q_{it} p_{it}}{\sum_{i=1}^{n} q_{i0} p_{it}}$$

**Exemple 11** : Considérons un ménage dont la consommation de 5 produits et/ou services, au cours des 3 dernières années, a évolué comme le montre le tableau synthétique suivant :

| N° | Période 1 | | Période 2 | | Période 3 | |
|---|---|---|---|---|---|---|
| | Quantité | Prix | Quantité | Prix | Quantité | Prix |
| | $q_1$ | $p_1$ | $q_2$ | $p_2$ | $q_3$ | $p_3$ |
| 1 | 2,5 | 10,45 | 3,2 | 10,65 | 4,4 | 11,32 |
| 2 | 5,9 | 43,87 | 6,8 | 43,88 | 6,7 | 43,90 |
| 3 | 4,8 | 120,78 | 5,7 | 121,76 | 6,1 | 135,99 |
| 4 | 1,2 | 156,98 | 1,6 | 166,87 | 1,7 | 178,91 |
| 5 | 0,5 | 548,67 | 0,7 | 650,88 | 0,8 | 700,76 |

On demande de calculer, pour le cas de ce ménage, les indices prix et les indices quantités de LASPEYRES relatifs aux 3 dernières années.

On demande aussi d'évaluer, pour chaque type d'indice, l'ordre de grandeur de l'erreur qu'on commet en appliquant injustement la propriété de circularité aux indices de LASPEYRES.

Il s'agit, en fait, d'un cas très particulier de calcul de l'indice du coût de la vie.

a) Calculons les indices de prix de LASPEYRES, pour ce faire, on dresse le tableau de calculs suivant :

| N° | $p_{i1}q_{i1}$ | $p_{i2}q_{i1}$ | $p_{i3}q_{i1}$ | $p_{i2}q_{i2}$ | $p_{i3}q_{i2}$ |
|---|---|---|---|---|---|
| 1 | 26,13 | 26,63 | 28,30 | 34,08 | 36,22 |
| 2 | 258,83 | 258,89 | 259,01 | 298,38 | 298,52 |
| 3 | 579,74 | 584,45 | 652,75 | 694,03 | 775,14 |
| 4 | 188,38 | 200,24 | 214,69 | 266,99 | 286,26 |
| 5 | 274,34 | 325,44 | 350,38 | 455,62 | 490,53 |

| | | | | | |
|---|---|---|---|---|---|
| $\sum$ | 1327,41 | 1395,65 | 1505,13 | 1749,10 | 1886,68 |
| **$Lp_{2/1}$** | | 1,0514 | - - - | - - - | - - - |
| **$Lp_{3/1}$** | | | 1,1339 | - - - | - - - |
| **$Lp_{3/2}$** | | | | | 1,0787 |

Nous pouvons alors calculer l'erreur qu'on commet en appliquant, injustement, la propriété de circularité à l'indice de prix de LASPEYRES, en effet :

$L_{p3/2}$ x $L_{p2/1}$ = 1,0787 x 1,0514 = 1,1341 or $L_{p3/1}$ = 1,1339

$$\frac{L_{p3/2} \times L_{p2/1} - L_{p3/1}}{L_{p3/1}} = \frac{1,1341 - 1,1339}{1,1339} = 0,0002 = 0,02\%$$

On voit bien que l'erreur est minime puisqu'elle est à peine égale à 0,02%.

b) Calculons maintenant les indices de quantités de LASPEYRES, pour ce faire, on dresse le tableau de calculs suivant :

| **N°** | **qi1pi1** | **qi2pi1** | **qi3pi1** | **qi2pi2** | **qi3pi2** |
|---|---|---|---|---|---|
| 1 | 26,13 | 33,44 | 45,98 | 34,08 | 46,86 |
| 2 | 258,83 | 298,32 | 293,93 | 298,384 | 293,996 |
| 3 | 579,74 | 688,45 | 736,76 | 694,032 | 742,736 |
| 4 | 188,38 | 251,17 | 266,87 | 266,992 | 283,679 |
| 5 | 274,34 | 384,07 | 438,94 | 455,616 | 520,704 |
| $\sum$ | 1327,41 | 1655,44 | 1782,47 | 1749,104 | 1887,975 |
| **$Lq_{2/1}$** | | 1,2471 | - - - | - - - | - - - |
| **$Lq_{3/1}$** | | | 1,3428 | - - - | - - - |
| **$Lq_{3/2}$** | | | | | 1,0794 |

Nous pouvons alors calculer l'erreur qu'on commet en appliquant, injustement, la propriété de circularité à l'indice de quantité de LASPEYRES, en effet :

$L_{q3/2}$ x $L_{q2/1}$ = 1,0794 x 1,2471 = 1,3461 or $L_{3/1}$ = 1,3428

$$\frac{L_{q3/2} \times L_{q2/1} - L_{q3/1}}{L_{q3/1}} = \frac{1,3461 - 1,3428}{1,3428} = 0,25\%$$

On voit bien que l'erreur est minime puisqu'elle est à peine égale à 0,25%.

**Exemple 12** : Reprenons les données de l'exemple 11, On demande de calculer, pour le cas de ce ménage, les indices prix et les indices quantités de PAASCHE relatifs aux 3 dernières années.

On demande aussi d'évaluer, pour chaque type d'indice, l'ordre de grandeur de l'erreur qu'on commet en appliquant injustement la propriété de circularité aux indices de PAASCHE.

a) Calculons les indices de prix de PAASCHE, pour ce faire, on dresse le tableau de calculs suivant :

| N° | $p_{i1}q_{i2}$ | $p_{i2}q_{i2}$ | $p_{i1}q_{i3}$ | $p_{i3}q_{i3}$ | $p_{i2}q_{i3}$ |
|---|---|---|---|---|---|
| 1 | 33,4 | 34,08 | 45,98 | 49,81 | 46,86 |
| 2 | 298,3 | 298,38 | 293,93 | 294,13 | 293,996 |
| 3 | 688,4 | 694,03 | 736,76 | 829,54 | 742,736 |
| 4 | 251,2 | 266,99 | 266,87 | 304,15 | 283,679 |
| 5 | 384,1 | 455,62 | 438,94 | 560,61 | 520,704 |
| $\sum$ | 1655,4 | 1749,10 | 1782,47 | 2038,23 | 1887,98 |
| $Pp_{2/1}$ | | 1,0566 | | - - - | - - - |
| $Pp_{3/1}$ | | | | 1,1435 | - - - |
| $Pp_{3/2}$ | | | | | 1,0796 |

Nous pouvons alors calculer l'erreur qu'on commet en appliquant, injustement, la propriété de circularité à l'indice de prix de PAASCHE, en effet :

$P_{p3/2}$ x $P_{p2/1}$ = 1,0796 x 1,0566 = 1,1407 or $P_{p3/1}$ = 1,1435

$$\frac{P_{p3/2} \times P_{p2/1} - P_{p3/1}}{P_{p3/1}} = \frac{1,1407 - 1,1435}{1,1435} = -0,0025 = -0,25\%$$

On voit bien que l'erreur est minime puisqu'elle est à peine égale à 0,25%.

b) Calculons les indices de quantités de PAASCHE, pour ce faire, on dresse le tableau de calculs suivant :

| N° | $q_{i1}p_{i2}$ | $q_{i2}p_{i2}$ | $q_{i1}p_{i3}$ | $q_{i3}p_{i3}$ | $q_{i2}p_{i3}$ |
|---|---|---|---|---|---|
| 1 | 26,63 | 34,08 | 28,30 | 49,81 | 36,22 |
| 2 | 258,89 | 298,38 | 259,01 | 294,13 | 298,52 |
| 3 | 584,45 | 694,03 | 652,75 | 829,54 | 775,14 |
| 4 | 200,24 | 266,99 | 214,69 | 304,15 | 286,26 |
| 5 | 325,44 | 455,62 | 350,38 | 560,61 | 490,53 |
| $\sum$ | 1395,6 | 1749,10 | 1505,13 | 2038,23 | 1886,68 |
| $Pq_{2/1}$ | | 1,2533 | - - - | - - - | - - - |
| $Pq_{3/1}$ | | | | 1,3542 | - - - |

| $Pq_{3/2}$ | 1,0803 |
|---|---|

Nous pouvons alors calculer l'erreur qu'on commet en appliquant, injustement, la propriété de circularité à l'indice de quantité de PAASCHE, en effet :

$P_{q3/2}$ x $P_{q2/1}$ = 1,0803 x 1,2503 = 1,3507 or $P_{q3/1}$ = 1,3542

$$\frac{P_{q3/2} \times P_{q2/1} - P_{q3/1}}{P_{q3/1}} = \frac{1,3507 - 1,3542}{1,3542} = -0,26\%$$

On voit bien que l'erreur est minime puisqu'elle est à peine égale à 0,26%.

**Exemple 13** : Reprenons les données de l'exemple 11, On demande de calculer, pour le cas de ce ménage, les indices prix et les indices quantités de FISCHER relatifs aux 3 dernières années.

On demande aussi d'évaluer, pour chaque type d'indice, l'ordre de grandeur de l'erreur qu'on commet en appliquant injustement la propriété de circularité aux indices de FISCHER.

a) Calculons les indices de prix de FISCHER, pour ce faire, on dresse le tableau de calculs suivant :

| $L_{p2/1}$ | $P_{p2/1}$ | $L_{p3/1}$ | $P_{p3/1}$ | $L_{p3/2}$ | $P_{p3/2}$ |
|---|---|---|---|---|---|
| 1,0514 | 1,0566 | 1,1339 | 1,1435 | 1,0787 | 1,0796 |
| $F_{p2/1}$ | | $F_{p3/1}$ | | $F_{p3/2}$ | |
| 1,0540 | | 1,1387 | | 1,0791 | |

Nous pouvons alors calculer l'erreur qu'on commet en appliquant, injustement, la propriété de circularité à l'indice des prix de FISCHER, en effet :

$F_{p3/2}$ x $F_{p2/1}$ = 1,0791 x 1,0540 = 1,1374 or $F_{p3/1}$ = 1,1387

$$\frac{F_{p3/2} \times F_{p2/1} - F_{p3/1}}{F_{p3/1}} = \frac{1,1374 - 1,1387}{1,1387} = -0,0011 = -0,11\%$$

On voit bien que l'erreur est minime puisqu'elle est à peine égale à 0,11%.

b) Calculons les indices de quantité de FISCHER, pour ce faire, on dresse le tableau de calculs suivant :

| $L_{q2/1}$ | $P_{q2/1}$ | $L_{q3/1}$ | $P_{q3/1}$ | $L_{q3/2}$ | $P_{q3/2}$ |
|---|---|---|---|---|---|
| 1,2471 | 1,2533 | 1,3428 | 1,3542 | 1,0794 | 1,0803 |
| $F_{q2/1}$ | | $F_{q3/1}$ | | $F_{q3/2}$ | |
| 1,2502 | | 1,3485 | | 1,0798 | |

Nous pouvons alors calculer l'erreur qu'on commet en appliquant, injustement, la propriété de circularité à l'indice des quantités de FISCHER, en effet :

$F_{q3/2}$ x $F_{q2/1}$ = 1,0798 x 1,2502 = 1,3500 or $F_{q3/1}$ = 1,3485

$$\frac{F_{q3/2} \times F_{q2/1} - F_{q3/1}}{F_{q3/1}} = \frac{1{,}3500 - 1{,}3485}{1{,}3485} = 0{,}11\%$$

On voit bien que l'erreur est minime puisqu'elle est à peine 0,11%.

On voit bien sur cet exemple que tant pour l'indice prix que pour l'indice quantité de FISCHER, l'application de la propriété de circularité induit de faibles erreurs.

### 6.5.3. Indice des valeurs globales.

Les indices synthétiques de prix et de quantités de LASPEYRES et de PAASCHE, peuvent être combinés deux à deux pour retrouver l'indice des dépenses totales ou indice des valeurs globales.

Cet indice des dépenses totales est le rapport des valeurs globales aux prix et quantités de la période t sur les valeurs globales aux prix et quantités de la période 0.

Il est égal, par définition à :

$$D_{t/0} = \frac{\sum_{i=1}^{n} p_{it} q_{it}}{\sum_{i=1}^{n} p_{i0} q_{i0}}$$

Nous pouvons calculer cet indice en fonction des indices de prix et de quantités de LASPEYRES et de PAASCHE.

En effet multiplions le numérateur et le dénominateur de $D_{t/0}$ par $\sum_{i=1}^{n} p_{it} q_{i0}$ :

$$D_{t/0} = \frac{\sum_{i=1}^{n} p_{it} q_{it}}{\sum_{i=1}^{n} p_{i0} q_{i0}} \times \frac{\sum_{i=1}^{n} p_{it} q_{i0}}{\sum_{i=1}^{n} p_{it} q_{i0}} = \frac{\sum_{i=1}^{n} p_{it} q_{i0}}{\sum_{i=1}^{n} p_{i0} q_{i0}} \times \frac{\sum_{i=1}^{n} p_{it} q_{it}}{\sum_{i=1}^{n} p_{it} q_{i0}}$$

Ce qui donne : $D_{t/0} = L_{p\,t/0}$ x $P_{q\,t/0}$

De même, on aurait pu multiplier le numérateur et le dénominateur de $D_{t/0}$ par $\sum_{i=1}^{n} p_{i0}q_{it}$ :

$$D_{t/0}=\frac{\sum_{i=1}^{n} p_{it}q_{it}}{\sum_{i=1}^{n} p_{i0}q_{i0}}\times\frac{\sum_{i=1}^{n} p_{i0}q_{it}}{\sum_{i=1}^{n} p_{i0}q_{it}}=\frac{\sum_{i=1}^{n} q_{it}p_{i0}}{\sum_{i=1}^{n} q_{i0}p_{i0}}\times\frac{\sum_{i=1}^{n} p_{it}q_{it}}{\sum_{i=1}^{n} p_{i0}q_{it}}$$

Ce qui donne : $D_{t/0} = L_{q\,t/0} \times P_{p\,t/0}$

Les deux égalités qu'on vient d'établir entre 3 indices, à savoir, ceux de LASPEYRES, de PAASCHE et des valeurs globales, permettent de calculer l'un des 3 indices si l'on connaît les deux autres.

L'indice des valeurs globales possède la propriété de circularité :

$D_{t/t'} \times D_{t'/0} = D_{t/0}$

En effet on a :

$$D_{t/t'} \times D_{t'/0} = \frac{\sum_{i=1}^{n} p_{it}q_{it}}{\sum_{i=1}^{n} p_{it'}q_{it'}}\times\frac{\sum_{i=1}^{n} p_{it'}q_{it'}}{\sum_{i=1}^{n} p_{i0}q_{i0}} = \frac{\sum_{i=1}^{n} p_{it}q_{it}}{\sum_{i=1}^{n} p_{i0}q_{i0}} = D_{t/0}$$

**Exemple 14** : Reprenons les données de l'exemple 11, On demande de calculer, pour le cas de ce ménage, l'indice des valeurs globales relatif aux 3 dernières années.

Calculons les indices de valeurs globales, pour ce faire, on utilisera l'une des deux formules qu'on vient d'établir comme indiqué dans le tableau de calculs suivant :

| $L_{p2/1}$ | $P_{q2/1}$ | $L_{p3/1}$ | $P_{q3/1}$ | $L_{p3/2}$ | $P_{q3/2}$ |
|---|---|---|---|---|---|
| 1,0514 | 1,2533 | 1,1339 | 1,3542 | 1,0787 | 1,0803 |
| **$D_{2/1}$** | | **$D_{3/1}$** | | **$D_{3/2}$** | |
| 1,3177 | | 1,5355 | | 1,1653 | |

$D_{3/2} \times D_{2/1} = 1,1653 \times 1,3177 = 1,5355$ or $D_{3/1} = 1,5355$

On voit bien que l'indice des valeurs globales possède la propriété de circularité puisque :

$D_{3/2} x D_{2/1} = D_{3/1}$

Les indices de prix servent aussi à déterminer les indices de révision de prix, dans les marchés de travaux dont la durée de réalisation s'étale sur plusieurs années. Ces marchés comportent, la plupart du temps, des formules de révision de prix simples ou complexes.

### 6.5.4. Formules de révision des prix d'un marché.

Une formule de révision des prix est un indice synthétique qui permet de calculer les prix, à la date de réalisation des travaux, à partir des prix à la date de signature du contrat.

Le principe de révision des prix d'un marché vient du fait que le contrat est signé, à une date 0 et les travaux sont réalisés, à des dates ultérieures $t_1$, $t_2$, … $t_n$, il est donc normal de recalculer les nouveaux prix auxquels doivent être facturés les travaux.

La formule générale de révision des prix d'un marché est un indice synthétique qui donne le rapport de prix $P_t / P_o$ entre les instants t et 0, elle s'écrit de la façon suivante :

$$P_t / P_o = \alpha_0 + \sum_{i=1}^{n} \alpha_{i0} \left(\frac{I_{it}}{I_{i0}}\right)$$

$$\text{Avec } \alpha_0 + \sum_{i=1}^{n} \alpha_{i0} = 1$$

Les rapports $I_{it}$ et $I_{i0}$ donnent l'évolution de l'indice d'un constituant du marché : main d'œuvre, matières premières, produits finis ou semi finis, etc. Ces indices peuvent être simples ou synthétiques. En général, on admet que 10 à 20% du montant du marché ne soit pas révisable et que le reste le soit au prorata des montants des différents corps d'état dans le montant total du marché

**Exemple 15** : On considère un marché passé, en mars 2004, entre la société SAMTOL et l'entreprise BATIMAROC pour la construction du local pour stockage de la société SAMTOL. Le montant de ce marché se décompose comme suit :

| Intitulés | | Montants HT |
|---|---|---|
| - Génie civil | : | 2 358 500,00 |
| - Electricité | : | 452 360,00 |
| - Plomberie | : | 235 125,00 |
| - Menuiserie | : | 354 750,00 |
| - Appareillages électriques | : | 175 855,00 |
| Total | : | 3 576 590,00 |

On suppose que 15% du marché ne sont pas révisables et que le reste l'est au prorata des montants des différents corps d'état que sont le génie civil, l'électricité, la plomberie, la menuiserie et l'appareillage électrique dont les indices sont respectivement $I_{gc}$, $I_{élec}$, $I_{pl}$, $I_{me}$ et $I_{ap}$

Pour des raisons d'autorisations administratives, les travaux de ce marché n'ont démarré qu'en mai 2005, ont duré 3 mois et ont été facturés selon l'avancement des travaux comme suit :

- Juin 2005 : 1 249 965,00 DH
- Juillet 2005 : 1 103 769,00 DH
- Août 2005 : 1 222 856,00 DH

On demande de donner la formule de révision de prix de ce marché et de déterminer le montant total de la révision de prix si l'on suppose que les indices des différents corps d'état ont évolué, entre mars 2004 et les mois de réalisation, comme l'indique le tableau suivant :

| **Intitulés/mois** | **Mars 04** | **Juin 05** | **Juillet 05** | **Août 05** |
|---|---|---|---|---|
| - $I_{gc}$ Génie civil | 425,32 | 471,22 | 475,52 | 482,61 |
| - $I_{élec}$ Electricité | 256,54 | 281,62 | 293,22 | 301,05 |
| - $I_{pl}$ Plomberie | 356,23 | 392,26 | 394,66 | 400,02 |
| - $I_{me}$ Menuiserie | 332,56 | 382,12 | 390,21 | 392,35 |
| - $A_{ap}$ Appar élec | 517,31 | 550,21 | 562,38 | 581,54 |

a) Déterminons la formule de révision des prix du marché conformément à la formule générale de révision de prix d'un marché, c'est-à-dire une formule de la forme :

$$\frac{P_t}{P_0} = \alpha_0 + \sum_{i=1}^{n} \alpha_{i0} \left(\frac{I_{it}}{I_{i0}}\right)$$

Avec $\alpha_0 = 0{,}15$ et les autres $\alpha$ déterminés comme suit, au prorata des montants :

| **Intitulés** | **Montants HT** | **En % de 85%** |
|---|---|---|
| - Génie civil | 2 358 500,00 | 56,05 |
| - Electricité | 452 360,00 | 10,75 |
| - Plomberie | 235 125,00 | 5,59 |
| - Menuiserie | 354 750,00 | 8,43 |
| - Appareillages électriques | 175 855,00 | 4,18 |
| **Total** | **3 576 590,00** | **85%** |

La formule de révision des prix pour ce marché est donc :

$$\frac{P_t}{P_0} = 0,15 + 0,5605\frac{I_{gc\,t}}{I_{gc\,0}} + 0,1075\frac{I_{élec\,t}}{I_{élec\,0}} + 0,0559\frac{I_{pl\,t}}{I_{pl\,0}} + 0,0843\frac{I_{me\,t}}{I_{me\,0}} + 0,0418\frac{I_{ap\,t}}{I_{ap\,0}}$$

On trouve bien une formule similaire avec la somme des $\alpha$ égale à 1, en effet :

0,15 + 0,5605 + 0,1075 + 0,0559 + 0, 0843 + 0,0418 = 1

b) Pour calculer le montant total de la révision des prix de ce marché, on doit d'abord calculer les taux d'évolution des différents indices, selon les mois de réalisation et appliquer ces taux aux montants de factures :

**Facture de juin 2005**

| **Indices** | **Mars 04** | **Juin 05** | **coeff** | **taux** | **Coeff x taux** |
|---|---|---|---|---|---|
| Invariant | - - - | - - - | 0,1500 | 1,0000 | 0,1500 |
| - $I_{gc}$ | 425,32 | 471,22 | 0,5605 | 1,1079 | 0,6215 |
| - $I_{élec}$ | 256,54 | 281,62 | 0,1075 | 1,0978 | 0,1180 |
| - $I_{pl}$ | 356,23 | 392,26 | 0,0559 | 1,1011 | 0,0616 |
| - $I_{me}$ | 332,56 | 382,12 | 0, 0843 | 1,1490 | 0,0969 |
| - $A_{ap}$ | 517,31 | 550,21 | 0,0418 | 1,0636 | 0,0445 |
| **Total de la révision pour juin 2005** | | | | | 1,0920 |

**Facture de juillet 2005**

| **Indices** | **Mars 04** | **Juillet 05** | **coeff** | **taux** | **Coeff x taux** |
|---|---|---|---|---|---|
| Invariant | - - - | - - - | 0,1500 | 1,0000 | 0,1500 |
| - $I_{gc}$ | 425,32 | 475,52 | 0,5605 | 1,1180 | 0,6267 |
| - $I_{élec}$ | 256,54 | 293,22 | 0,1075 | 1,1430 | 0,1229 |
| - $I_{pl}$ | 356,23 | 394,66 | 0,0559 | 1,1079 | 0,0619 |
| - $I_{me}$ | 332,56 | 390,21 | 0, 0843 | 1,1734 | 0,0989 |
| - $A_{ap}$ | 517,31 | 562,38 | 0,0418 | 1,0871 | 0,0454 |
| **Total de la révision pour juillet 2005** | | | | | 1,1058 |

**Facture d'août 2005**

| **Indices** | **Mars 04** | **Août 05** | **coeff** | **taux** | **Coeff x taux** |
|---|---|---|---|---|---|
| Invariant | - - - | - - - | 0,1500 | 1,0000 | 0,1500 |
| - $I_{gc}$ | 425,32 | 482,61 | 0,5605 | 1,1347 | 0,6360 |

| | | | | | |
|---|---|---|---|---|---|
| - $I_{élec}$ | 256,54 | 301,05 | 0,1075 | 1,1735 | 0,1262 |
| - $I_{pl}$ | 356,23 | 400,02 | 0,0559 | 1,1229 | 0,0628 |
| - $I_{me}$ | 332,56 | 392,35 | 0, 0843 | 1,1798 | 0,0995 |
| - $A_{ap}$ | 517,31 | 581,54 | 0,0418 | 1,1242 | 0,0470 |
| **Total de la révision pour août 2005** | | | | | 1,1215 |

Ainsi la révision de prix, pour l'ensemble du marché, s'élève à :

$$\Delta P = P_1 + P_2 + P_3 - P_0$$

or P1 = 1 249 965,00 x 1,0920 = 1 364 961,78 DH
P2 = 1 103 769,00 x 1,1058 = 1 220 547,76 DH
P3 = 1 222 856,00 x 1,1215 = 1 371 433,00 DH
Total = 3 956 942,54 DH

Ce qui donne : $\Delta P = 3956942,54 - 3576590,00 = 380352,54 \text{ DH}$

Soit + 10,63 % du montant total du marché.

On voit là, l'intérêt de prévoir des formules de révision de prix pour des marchés qui ne sont pas réalisés dans des délais assez courts, en effet du fait que les délais sont, parfois longs, les prix des différents produits et services augmentent et il n'est pas raisonnable d'exiger de l'entrepreneur de réaliser des travaux ou de fournir des produits à des prix qui n'ont plus aucune réalité.

Dans notre exemple, l'entrepreneur peut facturer des révisions de prix d'un montant total de $380352,54 DH$ qui doit représenter les augmentations de prix qu'il a subies.

### 6.6. INDICES BOURSIERS.

Un indice boursier est un indice synthétique, il est calculé quotidiennement et correspond à la moyenne pondérée du cours des valeurs boursières sélectionnées de manière à refléter la tendance générale des cours des valeurs immobilières à la Bourse.

Les principaux indices boursiers dans le monde sont :
- à la Bourse de New York (Wall Street), le Dow Jones ;
- à Londres (Stock Exchange), le Footsie ;
- à Tokyo (Kabuto cho), le Nikkei ;
- à Francfort, le Dax ;
- à Paris, le C.A.C. 40 ;
- etc.

En toute rigueur, un indice synthétique de la bourse doit être un indice de valeurs globales, sous la forme :

$$I_{t/0} = \sum_{i=1}^{n} N_{it} \times \frac{I_{it}}{I_{i0}}$$

Avec :

- $I_{t/0}$ indice boursier à t par rapport à l'instant 0 ;
- $N_{it}$ nombre d'action i existant en bourse ;
- $I_{it}$ indice actuel de l'action i ;
- $I_{i0}$ indice de départ de l'action i.

Cependant et traditionnellement, dans le calcul des indices boursiers, on se contente de ne considérer que les valeurs mobilières les plus significatives, c'est-à-dire celle relatives aux entreprises les plus importantes en capitalisation mobilières (c'est-à-dire les plus fortes sommes : $\sum_{i=1}^{n} N_{it} P_{it}$ : Nombre d'action i multiplié par le prix de cette action à l'instant t).

Ceci explique, par exemple, la dénomination explicite de CAC40 de l'indice boursier de la place de Paris, pour lequel on retient les 40 entreprises les plus importantes. Ce panel de 40 entreprises peut évidemment changer, selon les périodes.

Dans le cas particulier de la bourse des valeurs de Casablanca, on fait appel à deux types d'indices boursiers, le MASI et le MADEX.

Le MASI est un indice boursier synthétique global qui traduit l'évolution de l'ensemble des valeurs cotées, à la bourse de Casablanca, par contre, le MADEX est un indice boursier synthétique partiel qui, par la traduction de l'évolution de quelques valeurs importantes (20) ambitionne de traduire assez fidèlement l'évolution de l'ensemble de la place de Casablanca. Le choix d'un indice boursier synthétique partiel est intéressant dans la mesure où il résume assez fidèlement l'évolution boursière de la place qu'il représente et que son calcul est assez rapide et facile.

## 6.7. EXERCICES D'APPLICATION.

### 6.7.1. Exercice.

Les prix ainsi que les quantités consommées des produits A et B sont donnés dans le tableau suivant, selon les années :

| **Produits/Années** | | **2000**<br>**t = 0** | **2002**<br>**t = 2** | **2004**<br>**t = 4** | **2006**<br>**t = 6** |
|---|---|---|---|---|---|
| **A** | **Prix** | 132,00 | 125,00 | 121,00 | 130,00 |

| | | | | | |
|---|---|---|---|---|---|
| | **Quantités** | 25 | 30 | 31 | 35 |
| **B** | **Prix** | 112,00 | 121,00 | 126,00 | 137,00 |
| | **Quantités** | 26 | 33 | 34 | 36 |

a) Calculer les indices élémentaires de prix des biens A et B avec l'année 2000 comme date de base.

b) Calculer les indices de prix de LASPEYRES suivants : $Lp_{2/0}$ et $Lp_{4/0}$.

c) Calculer l'indice de PAASCHE suivant : $Pp_{6/0}$.

**Solution** :

a) Indices élémentaires de prix des biens A et B avec l'année 2000 comme date de base :

| **Produits/Années** | **2000** (t = 0) | **2002** (t = 2) | **2004** (t = 4) | **2006** (t = 6) |
|---|---|---|---|---|
| A | 100 | 94,70 | 91,67 | 98,48 |
| B | 100 | 108,04 | 112,5 | 122,32 |

b) Indices de prix de LASPEYRES : $Lp_{2/0}$ et $Lp_{4/0}$.

| **Produits** | $p_{i0}q_{i0}$ | $p_{i2}q_{i0}$ | $p_{i4}q_{i0}$ |
|---|---|---|---|
| A | 3300 | 3125 | 3025 |
| B | 2912 | 3146 | 3276 |
| $\sum$ | 6212 | 6271 | 6301 |
| $\mathbf{Lp_{2/0}}$ | | 100,95 | |
| $\mathbf{Lp_{4/0}}$ | | | 101,43 |

c) Indice de PAASCHE : $Pp_{6/0}$.

| **Produits** | $p_{i6}q_{i6}$ | $p_{i0}q_{i6}$ |
|---|---|---|
| A | 4550 | 4620 |
| B | 4932 | 4032 |
| $\sum$ | 9482 | 8652 |
| $\mathbf{Pp_{6/0}}$ | | 109,59 |

### 6.7.2. Exercice.

On donne les relevés des prix et des quantités consommés pour deux groupes de produits, alimentation et habillement, à deux périodes différentes : 2002 et 2006.

| **Périodes** / **Groupe de produits** | **2002** | | **2006** | |
|---|---|---|---|---|
| | **Prix** | **Quantités** | **Prix** | **Quantités** |
| **Alimentation** | 21,00 | 29 | 22,00 | 27 |
| **Habillement** | 18,00 | 19 | 25,00 | 21 |

En prenant l'année 2002 comme date de référence :

a) Déterminer les indices élémentaires de prix pour chacun des groupes.
b) Déterminer les indices élémentaires de quantités pour chacun des groupes de produits.
c) À partir des indices calculés aux deux questions précédentes, déterminer l'indice des prix LASPEYRES et l'indice des quantités PAASCHE de l'année 2006.

**Solution** :
a) Indices élémentaires de prix pour chacun des groupes :

| **Groupe de produits** | $\mathbf{I_{2006/2002}}$ |
|---|---|
| **Alimentation** | 104,76 |
| **Habillement** | 138,89 |

b) Indices élémentaires de quantités pour chacun des groupes de produits.

| **Groupe de produits** | $\mathbf{I_{2006/2002}}$ |
|---|---|
| **Alimentation** | 93,10 |
| **Habillement** | 110,53 |

c) - Indice des prix LASPEYRES de l'année 2006 à partir des indices calculés aux deux questions précédentes.

| **Groupe de produits** | $\mathbf{p_{i2002}q_{i2002}}$ | $\mathbf{I_{2006/2002}}$ | $\mathbf{p_{i2002}q_{i2002} \times I_{2006/2002}}$ |
|---|---|---|---|
| **Alimentation** | 609 | 104,76 | 63798,84 |
| **Habillement** | 342 | 138,89 | 47500,38 |
| $\sum$ | 951 | - | 111299,22 |
| $\mathbf{Lp_{2006/2002}}$ | | | 117,03 |

- Indice des quantités PAASCHE de l'année 2006 à partir des indices calculés aux deux questions précédentes.

| **Groupe de produits** | $\mathbf{p_{i2006}q_{i2006}}$ | $\mathbf{I_{2006/2002}}$ | $\mathbf{p_{i2006}q_{i2006} / I_{2006/2002}}$ |
|---|---|---|---|
| **Alimentation** | 594 | 93,10 | 6,3802 |
| **Habillement** | 525 | 110,53 | 4,7498 |
| $\sum$ | 1119 | - | 11,1300 |
| $\mathbf{Pq_{2006/2002}}$ | | | 100,54 |

### 6.7.3. Exercice.

On donne les prix et les consommations suivantes pour 4 produits A, B, C et D.

| **Produits** | **A** | **B** | **C** | **D** |
|---|---|---|---|---|
| **Prix en 2002** | 35,00 | 15,00 | 93,00 | 278,00 |

| | | | | |
|---|---|---|---|---|
| **Prix en 2004** | 40,00 | 18,00 | 110,00 | 301,00 |
| **Consommation en 2002** | 93 | 110 | 30 | 171 |

Calculer un indice de prix global base 100 en 2002. Justifier votre choix et interpréter votre résultat.

**Solution** :

On calcule l'indice de prix LASPEYRES puisqu'on ne dispose que de la consommation de l'année de base.

| **Produits** | $p_{i2002}q_{i2002}$ | $p_{i2004}q_{i2002}$ |
|---|---|---|
| A | 3255 | 3720 |
| B | 1650 | 1980 |
| C | 2790 | 3300 |
| D | 47538 | 51471 |
| $\sum$ | 55233 | 60471 |
| $Lp_{2004/2002}$ | | 109,48 |

### 6.7.4. Exercice.

On donne les relevés des prix et des quantités consommés pour deux groupes de produits, logement et transport, à deux périodes différentes : 2002 et 2006.

| **Périodes** | **2002** | | **2006** | |
|---|---|---|---|---|
| **Groupe de produits** | **Prix** | **Quantités** | **Prix** | **Quantités** |
| **Logement** | 25,00 | 125 | 26,0 | 125 |
| **Transport** | 7,25 | 56 | 7,85 | 60 |

En prenant l'année 2002 comme date de référence :
a) Déterminer les indices élémentaires de prix pour chacun des groupes.
b) Déterminer les indices élémentaires de quantités pour chacun des groupes de produits

c) En supposant que ces indices évoluent régulièrement, déterminer les mêmes indices pour les années 2003, 2004 et 2005.

**Solution** :

a) Indices élémentaires de prix pour chacun des groupes.

| **Groupe de produits** | $I_{2006/2002}$ |
|---|---|
| **Logement** | 104 |
| **Transport** | 108,28 |

b) Indices élémentaires de quantités pour chacun des groupes de produits.

| **Groupe de produits** | $I_{2006/2002}$ |
|---|---|

| | |
|---|---|
| **Logement** | 100 |
| **Transport** | 107,14 |

c) Indices pour les années 2003, 2004 et 2005.

Indice des prix de logement annuel moyen = 100,99

Indice des prix de transport annuel moyen = 102,01

Indice des quantités de logement annuel moyen = 100

Indice des quantités de transport annuel moyen = 101,74

| | **Groupe de produits** | **2002** | **2003** | **2004** | **2005** | **2006** |
|---|---|---|---|---|---|---|
| **Indice des prix** | **Logement** | 100 | 100,99 | 101,98 | 102,99 | 104 |
| | **Transport** | 100 | 102,01 | 104,06 | 106,15 | 108,28 |
| **Indice des quantités** | **Logement** | 100 | 100 | 100 | 100 | 100 |
| | **Transport** | 100 | 101,74 | 103,51 | 105,31 | 107,14 |

### 6.7.5. Exercice.

Considérons un portefeuille de valeurs mobilières, composée de 2 actions X et Y dont les cours sont donnés dans le tableau suivant :

| **Cours des actions** | **31/12/2001** | **31/12/2005** |
|---|---|---|
| **X** | 625 | 700 |
| **Y** | 1000 | 1800 |

a) Calculer les indices simples pour l'année 2005 avec comme date de base 2001.

b) Calculer les indices synthétiques pour l'année 2005 avec comme date de base 2001.

c) Interpréter les résultats des questions a) et b)

d) En supposant que le 1er indice évolue régulièrement, déterminer le même indice pour les années 2002, 2003 et 2004.

**Solution** :

a) Indices simples pour l'année 2005 avec comme date de base 2001.

| **Cours des actions** | $\mathbf{I_{2005/2001}}$ |
|---|---|
| **X** | 112 |
| **Y** | 180 |

b) Indices synthétiques pour l'année 2005 avec comme date de base 2001.

- Moyenne des indices : $\mathbf{I_{2005/2001}} = \frac{112+180}{2} = 146$

- Indice des moyennes : $\mathbf{I_{2005/2001}} = \frac{700+1800}{625+1000} = 153{,}85$

c) Interprétation des résultats des questions a) et b)

Entre le 31/12/2001 et le 31/12/2005, le cours de l'action X a augmenté de 12 %

Entre le 31/12/2001 et le 31/12/2005, le cours de l'action Y a augmenté de 80 %

Entre le 31/12/2001 et le 31/12/2005, les cours des action X et Y ont connu une augmentation moyenne de 46 %

Entre le 31/12/2001 et le 31/12/2005, le cours moyen des actions X et Y a augmenté de 53,85 %

d) Indices simples pour les années 2002, 2003 et 2004.

Indice annuel moyen de l'action X = 102,87

Indice annuel moyen de l'action Y = 115,83

### 6.7.6. Exercice.

On a enregistré, à différentes périodes, les prix et les quantités de 2 produits A et B.

| **Périodes** | **2001** | | **2003** | | **2005** | |
|---|---|---|---|---|---|---|
| **Produits** | **Prix** | **Quantités** | **Prix** | **Quantités** | **Prix** | **Quantités** |
| **A** | 11 | 35 | 12 | 41 | 15 | 42 |
| **B** | 18 | 10 | 21 | 12 | 24 | ? |

a) Quelle quantité de B a été enregistrée en 2005 sachant que l'indice de PAASCHE des quantités de 2005 par rapport à 2001 était égal à 126,27 % ?

b) Donner les indices de prix des 2 produits pour les années 2003 et 2005 en prenant 2001 comme date de référence.

c) Calculer les indices de prix de LASPEYRES et de PAASCHE, à partir des indices simples de la question b).

**Solution** :

a) Quantité de B enregistrée en 2005 sachant que l'indice de PAASCHE des quantités de 2005 par rapport à 2001 était égal à 126,27 % ?

$$P_{q2005/2001} = \frac{\sum_{i=1}^{n} q_{i2005} p_{i2005}}{\sum_{i=1}^{n} q_{i2001} p_{i2005}} = \frac{42 \times 15 + q \times 24}{35 \times 15 + 10 \times 24} = 1,2627 \text{ soit : } q = 14$$

b) Indices de prix des 2 produits pour les années 2003 et 2005 en prenant 2001 comme date de référence.

| **Produits/Années** | **2001** | **2003** | **2005** |
|---|---|---|---|
| A | 100 | 109,09 | 136,36 |
| B | 100 | 116,67 | 133,33 |

c) Indices de prix de LASPEYRES et de PAASCHE, à partir des indices simples de la question b).

- Indices de prix LASPEYRES.

| **Produits** | $p_{i2001}q_{i2001}$ | $I_{2003/2001}$ | $p_{i2001}q_{i2001} \times I_{2003/2001}$ | $I_{2005/2001}$ | $p_{i2001}q_{i2001} \times I_{2005/2001}$ |
|---|---|---|---|---|---|
| A | 385 | 109,09 | 41999,65 | 136,36 | 52498,60 |
| B | 180 | 116,67 | 21000,60 | 133,33 | 23999,40 |
| $\sum$ | 565 | - | 63000,25 | - | 76498,00 |
| $Lp_{2003/2001}$ | | | 111,50 | | |
| $Lp_{2005/2001}$ | | | | | 135,39 |

- Indices de prix PAASCHE.

| **Produits** | $p_{i2003}q_{i2003}$ | $I_{2003/2001}$ | $p_{i2003}q_{i2003} / I_{2003/2001}$ | $p_{i2005}q_{i2005}$ | $I_{2005/2001}$ | $p_{i2005}q_{i2005} / I_{2005/2001}$ |
|---|---|---|---|---|---|---|
| A | 492 | 109,09 | 4,51 | 630 | 136,36 | 4,62 |
| B | 252 | 116,67 | 2,16 | 336 | 133,33 | 2,52 |
| $\sum$ | 744 | - | 6,67 | 966 | - | 7,14 |
| $Pp_{2003/2001}$ | | | 111,54 | | | |
| $Pp_{2005/2001}$ | | | | | | 135,29 |

**6.7.7. Exercice.**

On considère une place boursière avec 10 entreprises cotées. On se propose de définir et de calculer, pour le jour j, un certain nombre d'indices boursiers relatifs à cette place.

a) Déterminer la valeur de l'indice global de cette bourse qui tient compte de toutes les actions cotées.

b) Déterminer la valeur de l'indice partiel de cette bourse relatif aux 4 plus fortes capitalisations.

c) Déterminer la valeur de l'indice partiel de cette bourse relatif aux 4 plus fortes valeurs liquides en valeur.

Le tableau récapitulatif des cotations du jour j est donné ci-dessous :

| **Noms** | **Nombre actions** | **Valeur $V_{(j-1)}$** | **Valeur $V_{(j)}$** | **Mvt Nbre** | **Total capitalisation** | **Mvt DH** |
|---|---|---|---|---|---|---|
| **Alma** | 250 | 52,23 | 53,11 | 21 | 13277,50 | 1115,31 |
| **Blal** | 200 | 31,00 | 31,25 | 62 | 6250,00 | 1937,50 |
| **Cali** | 125 | 52,00 | 55,25 | 38 | 6906,25 | 2099,50 |
| **Dile** | 230 | 36,12 | 37,86 | 150 | 8707,80 | 5679,00 |
| **Elma** | 410 | 19,85 | 19,85 | 10 | 8138,50 | 198,50 |
| **Faty** | 210 | 21,13 | 19,22 | 51 | 4036,20 | 980,22 |
| **Grès** | 230 | 28,36 | 25,41 | 21 | 5844,30 | 533,61 |
| **Hély** | 245 | 46,32 | 46,32 | 23 | 11348,40 | 1065,36 |
| **Ikam** | 185 | 71,11 | 70,08 | 0 | 12964,80 | 0,00 |
| **Joly** | 245 | 39,46 | 45,96 | 150 | 11260,20 | 6894,00 |
| **Total** | 2330 | | | | 88733,95 | 20503,00 |

**Solution** :

a) Valeur de l'indice global de cette bourse qui tient compte de toutes les actions cotées.

$$I_{j/j-1} = \frac{2349,49}{2330} \times 100 = 101\ \%$$

b) Valeur de l'indice partiel de cette bourse relatif aux 4 plus fortes capitalisations.

$$I_{j/j-1} = \frac{966,89}{925} \times 100 = 105\ \%$$

c) Valeur de l'indice partiel de cette bourse relatif aux 4 plus fortes valeurs liquides en valeur.

$$I_{j/j-1} = \frac{814,34}{805} \times 100 = 101\ \%$$

**6.7.8. Exercice.**

Le tableau suivant donne quelques produits importés par le Maroc à partir de la France en 2002 et 2006.

| Produits | Quantités en 1000 tonnes | | Prix en 1.000.000 DH | |
|---|---|---|---|---|
| | **2002** | **2006** | **2002** | **2006** |
| **Acier** | 256 | 352 | 23 | 41 |
| **Aluminium** | 25 | 36 | 126 | 195 |
| **Cuivre** | 75 | 70 | 201 | 168 |

a) Calculer l'indice de quantité selon la pondération de LASPEYRES, base 100 à l'année 2002.

b) Calculer l'indice de prix selon la pondération de PAASCHE, base 100 à l'année 2002.

c) Calculer l'indice de Fisher.

d) Calculer ces mêmes indices pour l'année 2004 en supposant que les indices simples de prix et de quantités des différents produits évoluent régulièrement entre 2002 et 2006.

e) Calculer ces mêmes indices pour l'année 2004 en supposant qu'ils évoluent régulièrement entre 2002 et 2006.

**Solution** :

a) Indice de quantité selon la pondération de LASPEYRES, base 100 à l'année 2002.

$$Lq_{2006/2002} = 110,74$$

b) Calculer l'indice de prix selon la pondération de PAASCHE, base 100 à l'année 2002.

$$Pp_{2006/2002} = 124,38$$

c) Indice de Fisher. : $Fp_{2006/2002} = 120,12$ et $Fq_{2006/2002} = 114,67$

d) Calcul des mêmes indices pour l'année 2004 en supposant que les indices simples de prix et de quantités des différents produits évoluent régulièrement entre 2002 et 2006.

Indices simples de prix et de quantités pour chacun des produits.

| Produits | $Ip_{2006/2002}$ | $Iq_{2006/2002}$ |
|---|---|---|
| **Acier** | 178,26 | 137,5 |
| **Aluminium** | 154,76 | 144 |
| **Cuivre** | 83,58 | 93,33 |

| Produits | $Ip_{2004/2002}$ | $Iq_{2004/2002}$ |
|---|---|---|
| **Acier** | 133,51 | 117,26 |
| **Aluminium** | 124,40 | 120 |
| **Cuivre** | 91,42 | 96,61 |

$Lq_{2004/2002} = 104,31$ ; $Pp_{2004/2002} = 108,37$ ; $Lp_{2004/2002} = 106,42$ et $Pq_{2004/2002} = 106,22$

$Fp_{2006/2002} = 107,39$ et $Fq_{2006/2002} = 105,26$

e) Calcul des mêmes indices pour l'année 2004 en supposant qu'ils évoluent régulièrement entre 2002 et 2006.

| Indices | 2006/2002 | Indice annuel moyen | 2004/2002 |
|---|---|---|---|
| **Lp** | 116 | 103,78 | 107,7 |
| **Pp** | 124,38 | 105,61 | 111,53 |
| **Lq** | 110,74 | 102,58 | 105,23 |
| **Pq** | 118,74 | 104,39 | 108,97 |
| **Fp** | 120,12 | 104,69 | 109,60 |
| **Fq** | 114,67 | 103,48 | 107,08 |

### 6.7.9. Exercice.

Un marché passé, en mars 2002 n'a été exécuté qu'en octobre 2002. Calculer la révision de prix à faire si la formule de révision est donnée par :

$P_t = P_0 ( 0,20 + 0,15Al_t/Al_0 + 0,30ACu_t/Cu_0 + 0,35Fe_t/Fe_0)$

Les indices $Fe_t/Fe_0$ , $Al_t/Al_0$ et $Cu_t/Cu_0$ sont ceux du fer, de l'aluminium et du cuivre, principales fournitures du marché qui ont évolué de mars à octobre, respectivement de 5%, de 7% et de 4%.

**Solution** : $P_{oct\ 2002} / P_{mars\ 2002} = 104\ \%$

### 6.7.10. Exercice.

L'administration a signé un marché avec l'entreprise SOTAG pour la réalisation d'un projet sur plusieurs mois. MATAG facture ses travaux tous les deux mois.

Calculer les révisions de prix dues pour toutes les factures que SOTAG soumet au paiement, sachant que :

Date de signature du marché : Mars 2001.
Début des travaux : Septembre 2001.
Fin des travaux : Mars 2002.
Base de référence des indices : Janvier 2000
Formule de révision des prix : $P_t = P_0(0,25 + 0,25S_t/S_0 + 0,30GO_t/GO_0 + 0,20CS_t/CS_0)$

L'évolution des indices est donnée par le tableau suivant:

| Mois / Année | $S_i$ | $GO_i$ | $CS_i$ |
|---|---|---|---|
| **Mars 2001** | 124 | 345 | 225 |

| | | | |
|---|---|---|---|
| **Septembre 2001** | 125 | 345 | 233 |
| **Novembre 2001** | 125 | 355 | 245 |
| **Janvier 2002** | 126 | 365 | 256 |
| **Mars 2002** | 130 | 370 | 261 |

Avec $S_i$ : indice des salaires
$GO_i$ : indice des gros oeuvres
$CS_i$ : indice des corps d'état secondaires

Les factures établies par MATAG ont été comme suit :

F1 = 234 345,98 DH à fin septembre 2001
F2 = 543 768,56 DH à fin novembre 2001
F3 = 354 621,34 DH à fin janvier 2002
F4 = 147 869,24 DH à fin mars 2002

**Solution** :
Formule de révision des prix : $P_t = P_0(0,25 + 0,25S_t/S_0 + 0,30GO_t/GO_0 + 0,20CS_t/CE_0)$

Pour calculer le montant total de la révision des prix de ce marché, on doit d'abord calculer les taux d'évolution des différents indices, selon les mois de réalisation et appliquer ces taux aux montants de factures :

**Facture de fin septembre 2001**

| **Indices** | **Mars 01** | **Septembre 01** | **coeff** | **taux** | **Coeff x taux** |
|---|---|---|---|---|---|
| Invariant | - - - | - - - | 0,25 | 1,0000 | 0,2500 |
| - $I_S$ | 124 | 125 | 0,25 | 1,0081 | 0,2520 |
| - $I_{GO}$ | 345 | 345 | 0,30 | 1 | 0,30 |
| - $I_{CS}$ | 225 | 233 | 0,20 | 1,0356 | 0,2071 |
| **Total de la révision pour fin septembre 2001** | | | | | 1,0091 |

**Facture de fin novembre 2001**

| **Indices** | **Mars 01** | **Novembre 01** | **coeff** | **taux** | **Coeff x taux** |
|---|---|---|---|---|---|
| Invariant | - - - | - - - | 0,25 | 1,0000 | 0,2500 |
| - $I_S$ | 124 | 125 | 0,25 | 1,0081 | 0,2520 |
| - $I_{GO}$ | 345 | 355 | 0,30 | 1,0290 | 0,3087 |
| - $I_{CS}$ | 225 | 245 | 0,20 | 1,0889 | 0,2178 |
| **Total de la révision pour fin novembre 2001** | | | | | 1,0285 |

**Facture de fin janvier 2002**

| **Indices** | **Mars 01** | **janvier 2002** | **coeff** | **taux** | **Coeff x taux** |
|---|---|---|---|---|---|
| Invariant | - - - | - - - | 0,25 | 1,0000 | 0,2500 |
| - $I_S$ | 124 | 126 | 0,25 | 1,0161 | 0,2540 |

| | | | | | |
|---|---|---|---|---|---|
| - $I_{GO}$ | 345 | 365 | 0,30 | 1,0580 | 0,3174 |
| - $I_{CS}$ | 225 | 256 | 0,20 | 1,1378 | 0,2276 |
| **Total de la révision pour fin janvier 2002** | | | | | 1,0490 |

**Facture de fin fin mars 2002**

| **Indices** | **Mars 01** | **mars 2002** | **coeff** | **taux** | **Coeff x taux** |
|---|---|---|---|---|---|
| Invariant | - - - | - - - | 0,25 | 1,0000 | 0,2500 |
| - $I_S$ | 124 | 130 | 0,25 | 1,0484 | 0,2621 |
| - $I_{GO}$ | 345 | 370 | 0,30 | 1,0725 | 0,3217 |
| - $I_{CS}$ | 225 | 261 | 0,20 | 1,1600 | 0,2320 |
| **Total de la révision pour fin mars 2002** | | | | | 1,0658 |

Ainsi la révision de prix, pour l'ensemble du marché, s'élève à :

$\Delta P$ = P1 + P2 + P3 + P4 – P0 = 1 325 341,31 – 1 280 605,12 = 44 736,19 DH

Soit + 3,5 % du montant total du marché.

# BIBLIOGRAPHIE


| TITRES | AUTEURS | EDITIONS |
|---|---|---|
| AIDE MEMOIRE DE PROBABILITES ET STATISTIQUES | J. MARCEIL | ELLIPSES 92 |
| ANALYSE STATISTIQUE DES DONNEES APPLICATIONS ET CAS POUR LE MARKETING | H. FENNETEAU | ELLIPSES 93 |
| COURS DE STATISTIQUE | G. HERNIAUX | MASSON 71 |
| COURS DE STATISTIQUE DESCRIPTIVE | G. CALOT | DUNOD 73 |
| DE L'ANALYSE A LA PREVISION | D. SCHLACTHER | ELLIPSES 86 |
| ELEMENT DE MATHEMATIQUES ET STATISTIQUES POUR L'ECONOMIE TOME 1 ET TOME 2 | NAJIB MIKOU | WALLADA 93-94 |
| ETUDE STATISTIQUE DES DEPENDANCES | S. AIVAZIAN | MOSCOU 70 |
| EXERCICES CORRIGES DE STATISTIQUES DESCRIPTIVE | B. GRAIS | DUNOD 83 |
| EXERCICES DE PROBABILITES ET STATISTIQUE | D. DACCUNHA | MASSO 96 |
| EXERCICES ET PROBLEMES RESOLUS DE STATISTIQUE PROBABILITE | M. ELLATIFI | AFRIQUE ORIENT 84 |
| EXERCICES ET PROBLEMES RESOLUS DE STATISTIQUES | M. ELLATIFI | AFRIQUE ORIENT 84 |
| EXERCICES RESOLUS DE STATISTIQUES APPLIQUEES A L'ECONOMIE | J. FOURASTIE | MASSON 93 |
| FORMULAIRE DE PROBABILITES ET DE STATISTIQUES | J. RENAULT | DUNOD 92 |
| INTRODUCTION A LA METHODE STATISTIQUE | B. GOLDFARB | DUNOD 99 |
| INTRODUCTION A LA STATISTIQUE | J. P. BELISLE | GAETAN MORIN 83 |
| INTRODUCTION A LA STATISTIQUE APPLIQUEE | S. ALALOUF | WESLEY 90 |
| INTRODUCTION A LA STATISTIQUE DESCRIPTIVE | G. BAILLAGEON | S.M.G. 81 |
| INTRODUCTION AUX PROBABILITES ET A LA STATISTIQUE | E. AMIOT | GAETAN MORIN 90 |

| TITRES | AUTEURS | EDITIONS |
|---|---|---|
| METHODES STATISTIQUES | P. TASSI | ECONOMICA 89 |
| METHODES STATISTIQUES | B. GRAIS | DUNOD 2000 |
| METHODES STATISTIQUES EN GESTION | M. TENENHAUS | DUNOD 96 |
| PREVISION Approche empirique d'une méthode statistique | M. DAVID | MASSON 89 |
| PROBABILITES ET STATISTIQUE ET TECHNIQUES DE REGRESSION | G. BAILLARGEON | S.M.G 89 |
| PROBABILITES ET STATISTIQUES | J. FOURASTIE | DUNOD 87 |
| PROBABILITES ET STATISTIQUES | AUDET, BOUCHER | GAETAN MORIN 93 |
| PROBABILITES ET STATISTIQUES COURS DE MATHEMATIQUES | L. CACOGNE | EYROLLES 90 |
| REGRESSION Nouveaux regards sur une ancienne méthode statistique | R. TOMASSONE | MASSON 92 |
| STATISTIQUE APPLIQUEE | G. BAILLARGEON | SMG 79 |
| STATISTIQUE CONCEPTS ET METHODES AVEC EXERCICES CORRIGES | S. LESSARD | MASSON 93 |
| STATISTIQUE DESCRIPTIVE | B. GRAIS | DUNOD 91 |
| STATISTIQUE DESCRIPTIVE | BERNARD PY | ECONOMICA 88 |
| STATISTIQUE DESCRIPTIVE - MANUEL | B. GRAIS | DUNOD 94 |
| STATISTIQUE DESCRIPTIVE : EXERCICES CORRIGES | B. GRAIS | DUNOD 99 |
| STATISTIQUE DESCRIPTIVE EXERCICES RESOLUS | I. ABBASSI<br>A. EL MARHOUM | LA SOURCE 94 |
| STATISTIQUE ET CALCUL DES PROBABILITES | W. MASIERI | SIREY 82 |
| STATISTIQUE ET PROBABILITE : TRAVAUX DIRIGES | J. P. LECOUTRE | DUNOD 2000 |
| STATISTIQUE ET PROBABILITES | ERIC FAVRO | DUNOD 91 |
| STATISTIQUE EXERCICES CORRIGES AVEC RAPPELS DE COURS TOMME 1 ET TOME 2 | C. LABROUSSE | DUNOD |
| STATISTIQUE INITIATION PRATIQUE | J. P. CABANNES | HACHETTE 90 |
| STATISTIQUE RESUME DE COURS-EXERCICES-PROBLEMES | P. JAFFARD | MASSON 90 |
| STATISTIQUE SANS MATHEMATIQUE | J. BADIA | ELLIPSES 97 |
| STATISTIQUES : ANNALES CORRIGES | G. PUPION | DUNOD 94 |
| STATISTIQUES ET PROBABILITES EN MATHEMATIQUES | C. M. BAUMONT | ELLIPSES 90 |
| STATISTIQUES EXERCICES CORRIGES AVEC RAPPELS DE COUR | C. LABROUSSE | DUNOD 78 |
| STATISTIQUES POUR L'ECONOMIE | J. HUBLER | BREAL 96 |
| STATISTIQUES UN OUTIL DU MANAGEMENT | C. RAMEAU | ORGANISATION 7 |