



**Observatoire Economique et Statistique d'Afrique Subsaharienne**

**Formation des agents non statisticiens travaillant dans les services  
statistiques des systèmes statistiques nationaux**

**Tome 2**

**Initiation à la statistique descriptive**

***Série méthode***

***mars 2019***

## Table des matières

<i>Avant propos</i> .....	3
<b>Chapitre 1 : Généralités : principaux concepts de base</b> .....	4
<b>Chapitre 2 : Résumé tabulaire des données statistiques</b> .....	10
2.1. Généralités et définitions.....	10
2.2. Distribution à un et deux caractères : Tableaux statistiques à simple et à double entrées .....	10
3.1. Représentation graphique des distributions à une variable .....	17
3.2. Représentation graphique d'une distribution à deux variables .....	22
4.1. Résumés numériques des distributions à une variable .....	24
4.1.1. Caractéristiques de valeur centrale .....	24
Les quantiles les plus courants :.....	31
5.1. Définition et caractéristiques.....	35
5.2. Notion de désaisonnalisation et correction des variations saisonnières.....	36
Annexes.....	38

## *Avant propos*

En septembre 2008, l'Observatoire a d'abord élaboré et mis à la disposition de la communauté statistique, un programme régional de formation continue répondant aux multiples attentes des utilisateurs. Il est à l'époque un véritable cadre de référence « à partir duquel les divers organismes fournisseurs d'offres de formation continue peuvent bâtir des plans de formation harmonisés et coordonnés en faveur des pays de la zone d'intervention d'AFRISTAT ». Les formations proposées par ce cadre étaient destinées aux personnels statisticiens ou assimilés. Ensuite, avec le concours de l'Association CESD-Statisticiens pour le développement, AFRISTAT s'est intéressé au cas spécifique des cadres non statisticiens travaillant dans les services statistiques nationaux. A leur intention, deux modules ont été élaborés pour servir de supports de formation pour leur fournir des notions de base nécessaires à la compréhension des tâches qu'ils exécutent quotidiennement dans leurs services respectifs.

Le document, intitulé « Formation des agents non statisticiens travaillant dans les Services statistiques des Systèmes statistiques nationaux (SSN) » est subdivisé en deux tomes : le tome 1 étant « Généralités sur travail statistique » et le tome 2 « Introduction à la statistique descriptive ». Il a pour objectif de faire comprendre l'environnement de l'exercice du métier de statisticien. Il constitue une démarche basée essentiellement sur l'initiation au vocabulaire et aux méthodes de travail du statisticien orientée vers les agents non statisticiens travaillant dans les services statistiques.

Un tel outil est apparu opportun pour répondre aux sollicitations de plus en plus nombreuses des Etats membres et non membres quant au développement des capacités statistiques en vue d'améliorer la qualité des données qu'ils produisent ; surtout dans le contexte d'insuffisance du nombre de personnels spécialisés affectés dans les services statistiques.

La réalisation des deux tomes de ce document est effective grâce aux apports de MM. Julien Amegandjin et Zomana Camara, Consultants internationaux, et de M. Xavier Charoy de l'association CESD-statisticiens pour le développement. Les experts d'AFRISTAT y ont également contribué de façon significative.

La Direction générale d'AFRISTAT tient à exprimer sa reconnaissance à tous ceux qui lui ont apporté leurs appuis multiformes pour la concrétisation de ce document, en particulier l'association CESD-Statisticiens pour le développement qui a fourni les moyens pour mobiliser l'expertise extérieure.

Enfin, nous formulons le vœu que cet ouvrage soit à la hauteur des attentes des agents non statisticiens qui les exploiteront ainsi que des responsables des Systèmes statistiques nationaux utilisateurs de ces personnels. Que par le biais de son utilisation, la qualité des données produites par les cadres non statisticiens formés favorise et contribue à la prise de bonnes décisions.

Cosme Z. VODOUNOU, Ph. D.

Directeur Général d'AFRISTAT

## Chapitre 1 : Généralités : principaux concepts de base

La **Statistique descriptive** décrit un ensemble d'observations. Elle a pour but d'analyser les données, de les ordonner et d'en dégager certaines caractéristiques du domaine observé.

Son objectif est de **décrire tant visuellement que numériquement un ensemble relativement important de données** ; en d'autres termes, c'est de résumer ou représenter par des tableaux, des graphiques ou des « indicateurs statistiques » les données nombreuses. Pour ce faire, la Statistique descriptive utilise un ensemble de concepts. Toutefois, même s'ils ont parfois la même dénomination dans le langage courant, ces concepts sont définis de façon précise et sans ambiguïté aucune.

Nous allons passer en revue les principaux concepts de base à partir d'exemples concrets issus des travaux de structures productrices du système statistique national du Mali.

### **Exemple N°1 : Caractère qualitatif nominal**

Considérons l'enregistrement en 2008 des offres d'emplois des entreprises maliennes selon la qualification professionnelle auprès des services d'intermédiation et de placement publics et privés (cf. *tableau N°01*).

**Tableau N°01** : Offres d'emploi enregistrées auprès des services d'intermédiation en 2008 selon la qualification professionnelle.

Qualifications demandées par les entreprises	Nombre	Pourcentage (%)
Cadre supérieur (CS)	457	22,6%
Cadre moyen et agent de maîtrise (CM/AM)	481	23,8%
Ouvrier	351	17,4%
Employé	318	15,8%
Manœuvre et assimilé	411	20,4%
<b>Total</b>	<b>2018</b>	<b>100,0%</b>

Source : BSIMT 04 & 05 / DOEF-ANPE

Ces statistiques (qu'on désignera par *données brutes statistiques* ou *série statistique*), collectées par l'Observatoire de l'emploi et de la formation (OEF) de l'Agence nationale pour l'emploi (ANPE), ont été classées selon les qualifications professionnelles demandées par les entreprises.

L'*ensemble*, objet de l'enregistrement est constitué de toutes les offres d'emploi que les entreprises ont fait enregistrer auprès des services d'intermédiation (ANPE, Bureaux et agences d'intermédiation et de placement payant). On appelle cet ensemble la **population statistique**. Ses *éléments*, constitués des « *offres d'emploi enregistrés* », sont appelés **individus** ou **unités statistiques**.

Le nombre **2018** qui est le *nombre total* des offres enregistrées en 2008 auprès des services d'intermédiation est l'**effectif total** de la population ou de l'ensemble statistique.

Comme signalé ci-dessous, les offres d'emploi individuelles sont regroupées et synthétisées par l'OEF selon la « *qualification professionnelle* » qui comprend les « catégories » suivantes : Cadres Supérieurs, Cades Moyens et Agents de Maîtrise, Ouvriers, Employés, Manœuvres et assimilés.

### Encadré 1

#### Données brutes ou série statistique

Les **données statistiques brutes** ou **série statistique** sont un ensemble de mesures ou d'observations sur un ensemble d'individus et concernant l'état ou l'évolution d'un phénomène.

Ce sont les valeurs prises par une ou plusieurs variables sur un certain nombre d'individus. C'est l'ensemble des résultats d'une étude : valeurs du caractère et effectifs correspondants.

On représente habituellement une série statistique sous forme d'un tableau.

La « *qualification professionnelle* » est appelée **caractère** ou **variable statistique**. Une population peut être étudiée selon un ou plusieurs caractères.

Selon la nomenclature adoptée par l'ANPE, la qualification professionnelle d'une offre correspond à une et une seule des cinq (5) catégories ci-dessus citées. Autrement dit, la catégorie professionnelle d'une offre d'emploi ne peut prendre qu'une de ces 5 valeurs appelées **modalités** du caractère « *qualification professionnelle* ». Ce caractère dont les « valeurs » sont des « appellations nominales » est dit **caractère qualitatif nominal**.

Sur le tableau N°01, on constate qu'en face de chaque catégorie, correspond un nombre. Ainsi, on constate que les entreprises offre **457** emplois de cadre supérieurs, **481** emplois de cadres moyens et agents de maîtrise, **351** emplois d'ouvriers, **318** emplois d'employés et **411** emplois de manœuvres et assimilés.

Ces nombres d'emplois offerts par catégorie sont appelés **effectifs des modalités ou valeurs absolues**. C'est ainsi que le nombre 457 représente l'effectif de la modalité « *cadre supérieur* » du caractère « *qualification professionnelle* ».

#### Exercices d'application :

##### Exercice N° 1

A partir du **tableau N°01.1** ci-dessus relatif aux demandes d'emplois enregistrés en 2008, compléter les phrases suivantes :

- a) la population étudiée est .....
- b) le(s) caractère(s) étudié(s) est/sont .....
- c) le nombre **1 403** représente .....
- d) le nombre **10,9%** représente .....
- e) le nombre **8 638** représente .....
- f) le nombre **10 758** représente .....

**Tableau N°01.1** : Demandes d'emploi enregistrées en 2008 par les services d'intermédiation selon la qualification du demandeur et par sexe

Qualification du demandeur d'emploi	Homme		Femme		Total	
	Effectif	%	Effectif	%	Effectif	%
Cadre supérieur	2 163	25,0%	541	25,5%	2 704	25,1%
Cadre moyen et agent de maîtrise	2 620	30,3%	1006	47,5%	3 626	33,7%
Ouvrier	<b>1 403</b>	16,2%	209	9,9%	1 612	15,0%
Employé	1 384	16,0%	232	<b>10,9%</b>	1 616	15,0%
Manœuvre et assimilé	1 068	12,4%	132	6,2%	1 200	11,2%
<b>Total</b>	<b>8 638</b>	<b>100,0%</b>	<b>2120</b>	<b>100,0%</b>	<b>10 758</b>	<b>100,0%</b>

Source : BSIMT 04 &amp; 05 / DOEF-ANPE

**Exercice N° 2****2°) Soit la répartition des pays de l'UEMOA selon le Produit intérieur brut (PIB). On demande de déterminer :**

- a) la population étudiée ;
- b) l'unité statistique ;
- c) le caractère étudié et sa nature.

**Exercice N° 3****3°) On considère les données météorologiques relatives aux températures moyennes mensuelles du Mali au cours de l'année 2001.****Même questions que l'exercice N°2 ci-dessus.****Encadré 2****Définition de quelques concepts**

Une définition statistique doit être précise, ne comportant aucune ambiguïté ; en somme elle doit être opérationnelle, de sorte que deux observateurs différents doivent aboutir aux mêmes résultats en utilisant cette même définition.

**Définition 1 : population statistique, individu/unité statistique, effectif total**

L'ensemble sur lequel on effectue des observations en Statistique descriptive est appelé **population** ou **ensemble statistique** ou **Univers**.

Si cet ensemble est trop vaste, on en restreint l'étude à une partie appelée **échantillon** ou **sous population**.

Les éléments de la population sont appelés **individus** ou **unités statistiques**. Ils répondent à une ou plusieurs caractéristiques communes.

Le nombre total des individus d'une population statistique est appelé **effectif total**.

**N.B.** Les premières statistiques correctement élaborées ont été celles des *recensements démographiques*. Ainsi le vocabulaire statistique est essentiellement celui de la démographie.

**Exemple N°2 : Caractère qualitatif ordinal**

L'« *Enquête légère intégrée auprès des ménages* » au Mali en 2003 (ELIM 2003) réalisée par la Direction nationale de la statistique et de l'informatique (DNSI), actuel Institut National de la Statistique (INSAT) a étudié les conditions de vie des ménages.

Les données ci-dessous concernent les résultats de la *perception des ménages* quant à leur condition de vie de 2003 comparée à celle de 2002 (cf. *tableau N°02*).

**Tableau N°02** : Répartition (en %) des ménages du Mali selon leur perception comparée de la situation économique en 2003 du ménage par rapport à celle de 2002 et la situation matrimoniale.

Perception comparée de la situation actuelle	Etat matrimonial					Ensemble Mali
	Célibataire	Marié monogame	Marié polygame	Divorcé	Veuf	
Nettement pire	17,9	4,5	3,2	0,7	1,7	4,0
Pire	18,2	15,5	18,4	50,8	16,2	16,8
Même situation	11,5	11,0	11,6	9,8	18,2	11,8
Situation meilleure	48,6	63,5	58,4	37,0	60,5	61,3
Nettement meilleure	3,9	4,7	7,9	0,0	3,2	5,4
Ne sait pas	0,0	0,6	0,3	1,7	0,1	0,5
Non déclaré	0,0	0,2	0,2	0,0	0,0	6,2
Total	100,0	100,0	100,0	100,0	100,0	100,0

Source : ELIM 2003 DNSI

Le caractère « perception de la situation de 2003 comparée à celle de 2002 » prend ses valeurs dans les catégories littérales (modalités) « Nettement pire », « Pire », « Même situation », « Situation meilleure », « Nettement meilleure ».

Concernant les conditions de vie des ménages, on peut dire que dans un certain sens, la modalité « Nettement pire » décrit une perception de la pauvreté qui est moins désirée que celle décrite par la modalité « Pire », laquelle est moins désirée que la modalité « Même situation », qui à son tour est moins désirée que la modalité « Situation meilleure », elle-même moins désirée que la modalité « Nettement meilleure ».

La variable qualitative « perception de la situation de 2003 comparée à celle de 2002 » dont les modalités sont « linguistiquement » hiérarchisées selon un ordre total est un **caractère qualitatif ordinal**.

### **Exemple N°3** : Caractère quantitatif discret

Les données ci-dessous (cf. tableau N°03) sont issues de la quatrième édition de l'« Enquête démographique et de santé » de 2006 (EDSM – IV 2006). Elles concernent la distribution des ménages selon le « nombre de membres habituels du ménage » (taille du ménage) et selon qu'ils résident en milieu rural ou urbain.

**Tableau N°03** : Répartition (en %) des ménages du Mali selon le milieu de résidence et la taille du ménage.

Nombre de membres habituels (taille du ménage)	Pourcentage es ménages (%) (fréquences)		Total Mali
	Urbain	Rural	
0	0,2	0,0	0,1
1	6,7	4,3	5,0
2	8,8	9,0	8,9
3	11,9	13,8	13,2
4	13,7	14,2	14,1
5	15,5	13,3	14,0
6	11,9	11,5	11,7
7	9,6	9,8	9,7
8	5,0	7,3	6,6
9 ou plus	16,7	16,8	16,7
<b>Total</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>
<b>Taille moyenne des ménages</b>	<b>5,6</b>	<b>5,7</b>	<b>5,7</b>

Source : EDSM – IV – 2006 ; CPS-MS/DNSI/Macro International Inc., décembre 2007

Dans cet exemple, le caractère « *taille du ménage* » prend des valeurs numériques isolés, c'est-à-dire 0, 1, 2, 3, 4, 5, etc.... les effectifs de chaque modalité sont représentés par des pourcentages (**fréquences**). C'est un **caractère quantitatif discret**.

On notera qu'il y a deux sous ensemble de la population totale du Mali. Il s'agit des milieux urbain et rural.

#### **Exemple N°4 : Caractère quantitatif continu**

A la suite des résultats du « *Recensement général de la population et de l'habitat de 1998* » (RGPH-98), la DNSI a élaboré les « perspectives de population 1999-2024 ».

Les données suivantes sont le regroupement en quatre (04) groupes d'âge des perspectives de 1999 pour les populations féminine, masculine et l'ensemble hommes et femmes (cf. *tableau N°04*).

**Tableau N°04** : Population du Mali en 1999 selon l'âge (regroupé en 4 groupes d'âge) et le sexe

Age	Sexe	Homme	Femme	Total	Pourcentage du total par groupe d'âge
<b>Moins de 15 ans</b>		2 366 443	2 250 770	4 617 213	<b>46,3%</b>
<b>15 à 39 ans</b>		1 599 697	1 827 839	3 427 536	<b>34,4%</b>
<b>40 à 64 ans</b>		771 631	759 638	1 531 269	<b>15,4%</b>
<b>65 ans ou plus</b>		196 466	196 449	392 915	<b>3,9%</b>
<b>Total</b>		<b>4 934 237</b>	<b>5 034 696</b>	<b>9 968 933</b>	<b>100,0%</b>
<b>Pourcentage par sexe</b>		<b>49,5%</b>	<b>50,5%</b>	<b>100,0%</b>	

Source : RGPH98- Perspectives de la population (1999-2024) DNSI

Dans ce dernier exemple, le caractère groupe d'âge prend ses valeurs dans les intervalles « *moins de 15 ans* », « *de 15 à 19 ans* », « *de 40 à 64 ans* », « *de 65 ans ou plus de 65 ans* ». Ainsi, parmi les quatre millions et demi de maliens âgés de moins de 15 ans, on peut trouver des personnes de tout âge compris dans cet intervalle (par exemple 2 mois ; 7,5 mois ; 8 ans ; 10 ans ; 14,5 ans).

Le caractère considéré dans cet exemple (« *âge* ») est un **caractère quantitatif continu**.



### Encadré 3

#### Le caractère et ses différentes natures

Le **caractère statistique** ou une **variable statistique** est une propriété distinctive des individus d'une population. Il désigne une grandeur ou un attribut, observable sur un individu et susceptible de *varier* en prenant différentes valeurs possibles dites **modalités ou variantes**.

L'ensemble des valeurs possibles ou des modalités est appelé le **domaine de la variable**.

Du point de vue nature, on distingue les caractères **qualitatifs** des caractères **quantitatifs** (appelés dans ce cas **variables**).

- Le **caractère est dit qualitatif** s'il ne peut être mesuré ni repéré par un nombre. C'est un caractère dont les modalités sont des catégories (des mots). On distingue :
  - les *caractères qualitatifs nominaux* ou non ordonnés, c'est un caractère qui permet de *nommer les individus* à décrire. C'est-à-dire que les modalités sont exprimables par des noms ; elles ne sont pas ordonnées (hiérarchisées), c'est-à-dire qu'elles n'ont pas d'ordre logique.
  - les *caractères qualitatifs ordinaux* ou ordonnés, dont les modalités ont un ordre logique, c'est-à-dire qu'elles peuvent être hiérarchisées, rangées dans un ordre croissant ou décroissant.
- Le **caractère est quantitatif** s'il peut être mesuré ou repéré par un nombre. Toutes ses valeurs possibles sont des nombres. On distingue :
  - les **caractères quantitatifs continus**, s'ils peuvent prendre un très grand nombre de valeurs ;
  - les **caractères quantitatifs discontinus ou discrets** s'ils ne peuvent prendre que des valeurs isolées, souvent entières, dans l'intervalle de variation (domaine des valeurs de la variable). Ils ne peuvent prendre qu'un nombre limité de valeurs entières.

Les **observations** constituent l'ensemble des valeurs observées sur chaque individu. Ces valeurs peuvent être décrites soit **individuellement** (*données individuelles*), soit **regroupées simplement** (*données groupées*), soit **regroupées en classes** (*données groupées et classées*).

Le nombre d'unités d'observation (nombre des individus de la population) est dit **effectif total**.

On appelle **effectif** associé à une valeur de la variable, le nombre de fois où cette valeur apparaît.

On appelle **fréquence d'une modalité** associée à une valeur de la variable, le rapport de l'effectif de ladite modalité à l'effectif total observé. On rencontre également l'appellation de *fréquence relative* pour désigner ce rapport ou fréquence absolue lorsqu'il s'agit de l'effectif tout court.

## Chapitre 2 : Résumé tabulaire des données statistiques

### 2.1. Généralités et définitions

Une **distribution statistique** indique la présentation des individus selon la valeur du caractère. La distribution observée est généralement représentée sous la forme d'un **tableau statistique**.

Bien qu'ils soient une des matières premières importantes du travail statistique, les **tableaux** sont cependant réducteurs de l'information.

Un tableau statistique représente déjà une certaine structuration de l'information. La ligne "**total**" et/ou la colonne "**total**" font déjà la synthèse, le condensé des données figurant dans les autres cellules du tableau.

#### Encadré 4

##### Les différents éléments d'un tableau

Un tableau statistique comporte **5 éléments** ou caractéristiques essentiels :

- 1° le **titre** qui permet d'identifier l'objet du document (population étudiée). Il donne une indication souvent précise du thème abordé par le tableau ;
- 2° la **source** qui précise l'origine des informations et souvent leur date de publication ;
- 3° les **intitulés de lignes et de colonnes** qui permettent de connaître selon quels caractères la population est étudiée. Ils doivent être simples, concis et lisibles ;
- 4° les **unités de mesure** retenues. Il peut s'agir de kg, mètres, années, milliers, milliards, etc.
- 5° la **date de publication** ou d'élaboration. Elle situe les données du tableau dans le temps.

### 2.2. Distribution à un et deux caractères : Tableaux statistiques à simple et à double entrées

#### Exemple 1 : Tableaux à simple entrée

Une **distribution statistique à un caractère** ou **tableau à simple entrée** indique la présentation des individus selon la valeur du caractère. La distribution observée est représentée sous la forme d'un tableau statistique à deux colonnes indiquant la *nomenclature* ou *libellé* et donnant le *nombre d'unités statistiques (effectif)* de chaque modalité.

On considère la répartition des demandes d'emplois enregistrées par les services d'intermédiation du Mali en 2007 selon l'âge des demandeurs d'emplois (cf. *tableau N°05*).

**Tableau N°05** : Répartition des demandes d'emploi enregistrées en 2007 selon l'âge

Groupe d'âge (1)	Effectif (2)	Pourcentage (fréquence en %) (3)
< 15 ans	46	0,3
15 à 19 ans	333	2,5
20 à 24 ans	2 558	19,1
25 à 29 ans	5 142	38,4
30 à 39 ans	3 871	28,9
40 à 59 ans	1 319	9,9
60 ans et plus	47	0,4
Non déclaré	61	0,5
<b>Total</b>	<b>13 377</b>	<b>100,0</b>

Source : BSIMT 02 et 03 / DOEF/ANPE

- La première colonne comprend les modalités du caractère « âge »,
- la seconde les effectifs de ces modalités et
- la troisième les pourcentages correspondants (**fréquences des modalités**).

La troisième colonne peut ne pas être donnée dans le tableau original. Dans ce cas, il est obtenu par calculs élémentaires en cas de besoin. Ainsi, la première fréquence 0,34% est obtenue par le rapport :

$$\frac{46}{13377} * 100 = 0,34\%. \text{ La somme de la colonne des fréquences donne } 100.$$

**NB.** Il est possible de reconstituer la colonne des effectifs connaissant les fréquences des modalités et l'effectif total (ici 13 377).

#### Encadré 5

##### Tableau statistique, données individuelles, données groupées et données classées

Le **tableau statistique à un caractère/variable** encore nommé *tableau à une dimension* ou *tableau unidimensionnel* ou *univarié* ou du *premier ordre* est le **dénombrement**, en effectifs et/ou fréquences, des individus observés en fonction des modalités du caractère considéré.

Le tableau à simple entrée (ou tableau à une variable, ou tableau à une dimension) du premier ordre comprend, dans sa forme la plus simple, deux colonnes (ou deux lignes selon la présentation) :

- la première colonne contient les *valeurs* (ou *classe de valeurs*) ou *modalités* du caractère ;
- la seconde colonne enregistre les *effectifs* (*fréquences*) correspondants.

Ainsi, une distribution empirique d'une variable associe à chaque valeur ou modalité du caractère les effectifs (ou les fréquences) observées dans une population ou une partie de la population. Les distributions empiriques sont le domaine d'étude des statistiques descriptives.

Une série statistique peut être étudiée selon deux caractères possédés par les unités statistiques. On est alors en présence d'une distribution à deux caractères ou à deux dimensions, c'est-à-dire que les deux variables sont associées comme dans un repère cartésien. On dit qu'on a croisé deux caractères. On peut croiser des caractères de même nature ou de natures différentes. Le tableau de présentation des unités est un **tableau à double entrée** ou **tableau de contingence**.

Notons que les distributions à deux variables permettent de connaître l'influence d'une variable sur l'autre.

Le **tableau à double entrée** ou **tableau de contingence** est celui qui fait le croisement entre deux caractères dont l'un en ligne et l'autre en colonne. C'est-à-dire qu'à chaque effectif est associé une modalité de chacun des deux caractères étudiés.

##### Données individuelles, données groupées et données classées

On parle de *données individuelles* si, pour chaque individu on connaît la valeur du caractère.

Les *données groupées ou agrégées* sont celles dont on ne connaît que le nombre total d'individus selon les différentes valeurs ou modalités. On ne connaît pas la valeur du caractère pour chaque individu.

Les *données groupées et classées* sont des données groupées qui prennent leurs valeurs dans des intervalles.

#### **Exemple 2 : Tableaux à double entrée**

##### ❖ Définition et Présentation générale

Considérons maintenant le *tableau N°06*, extrait des résultats de l'ELIM-2006, qui donne la répartition des ménages selon leur nombre de personnes habituellement résidentes et selon le milieu de résidence (urbain / rural).

Tableau N°06 : Répartition des ménages selon la taille, selon le sexe

Milieu Taille	Urbain	Rural	Ensemble Mali (total par ligne)
<b>1 à 2 personnes</b>	36846	29405	66251
<b>3 à 4 personnes</b>	118275	113254	231529
<b>5 à 6 personnes</b>	118363	198834	317197
<b>7 ou plus</b>	265832	562101	827933
<b>Total (total par colonne)</b>	<b>539316</b>	<b>903594</b>	<b>1442910</b>
Taille moyenne	7,2	9,3	8,5

Source : ELIM 2006, volume 1, DNSI, juillet 2007

Les deux caractères sont : le « milieu de résidence » des ménages dont les modalités sont 'Urbain' et 'Rural' et la 'Taille' des ménages qui a pour modalités '1 à 2 personnes', '3 à 4 personnes', '5 à 6 personnes', '7 ou plus de 7 personnes'.

❖ Lecture du Tableau à double entrée (tableau N°06)

1°) les cellules (cases) des données :

- le croisement de la 1<sup>ère</sup> ligne « une à deux personnes » avec la 2<sup>ème</sup> colonne « Rural » donne le chiffre « 29 405 ». Ce chiffre se lit de la façon suivante : « **29405 ménages ruraux ont habituellement 'une à deux personnes par ménage'** » ;
- le croisement de la 4<sup>ème</sup> ligne et la 3<sup>ème</sup> colonne, c'est-à-dire le chiffre « 827933 » représente l'effectif des ménages du Mali qui ont une taille de '7 personnes ou plus;
- etc....

2°) les totaux par ligne

Les totaux par ligne représentent les effectifs totaux des ménages selon la taille quel que soit le milieu de résidence. On les appelle les **effectifs marginaux de la variable taille**.

Associés à la colonne des modalités du caractère « taille », ils donnent la **distribution marginale** de la variable « taille ».

3°) les totaux par colonne

De même les totaux par colonne sont les **effectifs marginaux de la variable milieu de résidence**. Ils donnent la **distribution marginale du caractère 'milieu de résidence'**, comme si l'on étudiait seulement ce caractère.

4°) Equilibre lignes - colonnes

Le total des « totaux par ligne » est égal au total des « totaux par colonne ». C'est l'**équilibre lignes - colonnes**. Il sert de vérification pour la justesse des totaux respectifs de la ligne et de la colonne, l'inexistence d'égalité appelant au recalcul de ces totaux.

**Encadré 6****Opérations sur les lignes et les colonnes**

1°) Equilibre lignes-colonnes : Total des totaux par ligne = Total des totaux par colonne.

ou

Total des *effectifs marginaux* du caractère *taille* = Total *effectifs marginaux* du caractère *milieu*.

2°) La somme des lignes fournit la *distribution marginale* du caractère « *taille* ». C'est comme si l'on étudiait seulement le caractère « *taille* ».

3°) La somme des colonnes fournit la *distribution marginale* du caractère « *milieu* », comme si l'on étudiait seulement le caractère « *milieu* ».

3°) Les lignes et colonnes conjointes

En se déplaçant sur la première ligne, on parcourt toutes les colonnes. On ne s'intéresse dans ce cas qu'aux seuls ménages de taille « *1 et 2 personnes* ». C'est la *distribution de la variable milieu de résidence pour les ménages de taille '1 et 2 personnes'*. On l'appelle **distribution conditionnelle** du milieu de résidence par rapport à la taille « *1 et 2 personnes* ».

**Exercices d'application :**

**Exercice N° 1 – Retrouver les concepts de base sur un tableau**

Soit le **tableau N°07** ci-dessous, issu de l'Enquête malienne d'évaluation de la pauvreté de 2001 de la DNSI et représentant la structure des dépenses moyennes des ménages selon les grands postes de dépenses et le degré de pauvreté de la localité résidentielle du ménage.

**Tableau N°07** : Structure en 2001 des dépenses moyennes des ménages par groupe de consommation pour l'ensemble du Mali

Degré de pauvreté Postes de dépenses	Moins pauvres	Vulnérables	Pauvres	Très pauvres	Ensemble Mali
<b>Alimentation</b>	245 323	134 810	99 038	<b>58 898</b>	<b>122 744</b>
<b>Habillement</b>	20 457	8 985	6 048	3 555	<b>8951</b>
<b>Logement, Energie, Eau</b>	36 153	11 534	6 835	2 579	<b>12 765</b>
<b>Meubles et ménagers</b>	6 176	2 261	1 401	712	<b>2387</b>
<b>Santé</b>	9 984	<b>4 222</b>	2 500	1 281	<b>4 056</b>
<b>Transport Communication</b>	26 535	6 306	3 282	1 411	<b>8 507</b>
<b>Loisirs et Education</b>	<b>5 136</b>	1 169	681	317	<b>1 662</b>
<b>Autres biens et services</b>	22 158	7 762	5 075	2 003	<b>8 262</b>
<b>Total</b>	<b>371 922</b>	<b>177 049</b>	<b>124 860</b>	<b>70 756</b>	<b>169 334</b>

Source : EMEP 2001, Résultats définitifs, DNSI novembre 2003

**Questions** : Déterminer :

- la population statistique étudiée ;
- l'individu statistique ;
- le(s) caractère(s) étudié(s) ;
- les modalités du (des) caractère(s) étudiés.

**Exercice N° 2 - Exercez vous à lire les tableaux du second ordre**

Repérer et lire les données suivantes indiquées dans le **tableau N°07** :

- a) 3<sup>ème</sup> ligne, 4<sup>ème</sup> colonne ;
- b) 5<sup>ème</sup> ligne, 2<sup>ème</sup> colonne ;
- c) 7<sup>ème</sup> ligne, 1<sup>ère</sup> colonne.

**Exercice N° 3 – Calcul des effectifs à partir des fréquences (pourcentages)**

On donne ci-dessous le tableau représentant la répartition (en %) des chefs de ménages selon leur situation matrimoniale et leur sexe (données issues de l'ELIM 2003),

**Tableau N°08** : Répartition des chefs de ménages (en %) selon leur situation matrimoniale

<b>Etat matrimonial</b>	<b>Sexe</b>	<b>Homme</b>	<b>Femme</b>	<b>Total</b>
<i>Célibataire</i>		0,9	2,8	1,1
<i>Marié(e) monogame</i>		66,7	9,6	60,7
<i>Marié(e) polygame</i>		30,3	5,7	27,7
<i>Divorcé(e)/séparé(e)</i>		0,4	8,0	1,2
<i>Veuf(ve)</i>		1,7	73,9	9,3
<b>Total</b>		100,0	100,0	100,0

Source : ELIM 2003 DNSI

**Travail à faire** : À partir des données ci-dessus, **remplir les cellules vides du tableau** suivant :

**Tableau N°08 bis** : Répartition des chefs de ménages selon leur situation matrimoniale

<b>Etat matrimonial</b>	<b>Sexe</b>	<b>Homme</b>	<b>Femme</b>	<b>Total</b>
<i>Célibataire</i>				
<i>Marié(e) monogame</i>				
<i>Marié(e) polygame</i>				
<i>Divorcé(e)/séparé(e)</i>				
<i>Veuf(ve)</i>				
<b>Total</b>		<b>1 100 110</b>	<b>130 216</b>	<b>1 230 326</b>

Source : ELIM 2003 DNSI

**Travaux Dirigés sur le chapitre 2****1°) Restitution des données de cellules de tableaux**

Des données ont été délibérément effacées dans des cellules des deux *tableaux N°09 et N°10*. Il vous est demandé de les restituer dans les cellules vides (grisées) à partir des données restantes.

**Tableau N°09** : Répartition des ménages par zone de résidence selon a) leur taille, b) le sexe de leur chef.

Caractéristique	Urbain		Rural		Ensemble Mali	
	Effectif	Pourcentage	Effectif	Pourcentage	Effectif	Pourcentage
<b>a) Taille</b>						
1 et 2 personnes		2,7		5,1		4,4
3 et 4 personnes		11,1		13,1		12,5
5 et 6 personnes		19,6		21,1		20,7
7 personnes et plus		66,6		60,6		62,4
<b>TOTAL</b>		<b>100,0</b>		<b>100,0</b>		<b>100,0</b>
Taille moyenne	<b>9,2</b>	–	<b>8,9</b>	–	<b>9,0</b>	–
Taille médiane	<b>8,0</b>	–	<b>8,0</b>	–	<b>8,0</b>	–
<b>b) Sexe du Chef de ménage</b>						
Homme	301 967	83,3				89,4
Femme	60 390	16,7	69 826	8,0		10,6
<b>TOTAL</b>		<b>100,0</b>			<b>1 230 326</b>	<b>100,0</b>

Source : ELIM 2003 DNSI

**Tableau N°10** : Demandes d'emploi enregistrées par qualification du demandeur et par sexe en 2008

Qualification du demandeur d'emploi	Homme		Femme		Total	
	Effectif	%	Effectif	%	Effectif	%
Cadre supérieur		25,0		25,5		
Cadre moyen et agent de maîtrise		30,3		47,5		
Ouvrier		16,2		9,9		
Employé		16,0		10,9		
Manœuvre et assimilé		12,4		6,2		
<b>Total</b>		<b>100,0</b>	<b>2120</b>	<b>100,0</b>	<b>10 758</b>	<b>100,0</b>

Source : BSIMT 04 &amp; 05 / DOEF-ANPE

**2°) Restitution des effectifs des modalités à partir des fréquences**

A l'aide du tableau N°11 ci-dessus, remplissez les cellules vides du tableau N°11 bis suivant

**Tableau N°11** : Répartition en pourcentage des demandes d'emploi enregistrées en 2007 par groupe d'âge et par sexe.

	Homme	Femme	Total
< 15 ans	0,32%	0,45%	0,34%
15 à 19 ans	2,56%	2,20%	2,49%
20 à 24 ans	18,41%	21,97%	19,12%
25 à 29 ans	37,16%	43,54%	38,44%
30 à 39 ans	29,23%	27,78%	28,94%
40 à 59 ans	11,45%	3,54%	9,86%
60 ans et plus	0,43%	0,04%	0,35%
Non déclaré	0,45%	0,48%	0,46%
<b>Total</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>

Source : BSIMT 02 et 03 / DOEF-ANPE

**Tableau N°11bis** : Répartition des effectifs des demandes d'emploi enregistrées en 2007 par groupe d'âge et par sexe

	Homme	Femme	Total
< 15 ans			
15 à 24 ans			
25 à 39 ans			
40 à 59 ans			
60 ans et plus			
Non déclaré			
<b>Total</b>	<b>10 692</b>		<b>13 377</b>

Source : BSIMT 02 et 03 / DOEF-ANPE

### 3°) Détermination d'un effectif à partir d'informations extérieures

Le *tableau N°12* ci-dessous donne la répartition des actifs en emploi en 2007 à Bamako selon leurs revenus mensuels d'emploi et leur sexe.

**Tableau N°12** : Répartition en % des actifs en emploi selon les revenus mensuels d'emploi

Revenus	Sexe		Ensemble Bamako
	Homme	Femme	
Moins de 29 000 FCFA	18,5%	52,3%	32,5%
[ 29 000, 50 000 [	31,2%	17,9%	25,7%
[ 50 000, 75 000 [	14,6%	11,3%	13,2%
[ 75 000, 100 000 [	12,5%	7,1%	10,3%
[ 100 000, 200 000 [	14,4%	7,5%	11,5%
[ 200 000, 500 000 [	6,3%	2,3%	4,6%
500 000 FCF et plus	0,4%	0,0%	0,2%
Rien du tout	2,1%	1,7%	1,9%
<b>Total</b>	<b>100,0%</b>	<b>100,0%</b>	<b>100,0%</b>

Source : EPAM 2007/DOEF/

Par ailleurs, en cette même année 2007, on disposait des données suivantes :

- a) le nombre total d'actifs en emploi à Bamako était évalué à **4 108 779** dont **2 020 243** hommes.
- b) le **SMIG** était fixé à environ **29 000 F**.

**Question** : Calculer le nombre de femmes qui ont un revenu mensuel inférieur au SMIG.



## Chapitre 3 : Résumé graphique des données

### Introduction

La représentation graphique est très importante pour présenter une étude ou un rapport statistique. En effet, un graphique vaut mieux que mille mots ou mille chiffres. Cependant, mal faite, elle peut induire en erreur le lecteur en masquant les informations intéressantes de la distribution ou faire apparaître des faits mineurs sans importance.

#### Encadré 8

##### Graphique

Par rapport à un texte ou à un tableau, un **graphique** présente l'avantage de *permettre une lecture globale*, des comparaisons visuelles très rapides. Il réalise ainsi une description visuelle des séries statistiques. Cependant, le graphique est moins précis qu'un tableau. Mais, plus que les tableaux, les graphiques restituent la richesse de l'information. Il attire l'œil et focalise sur un point précis. Traduire un tableau de chiffres en un graphique peut rendre souvent la lecture plus facile.

Le **graphique** comme un tableau doit comporter un *titre*, si possible informatif, une *unité*, une *source* et, éventuellement, une *légende* (qui reprend les couleurs ou valeurs du graphique avec leur intitulé).

### 3.1. Représentation graphique des distributions à une variable

#### **Exemple 1 : Représentation graphique d'une distribution à un caractère qualitatif : diagrammes circulaire et à bande**

Représentons graphiquement la distribution des Offres d'emplois enregistrées en 2008 selon la qualification professionnelle (cf. tableau N°01)

## Encadré 9

## Représentation graphique

**Principe**

La logique des représentations les plus courantes est de faire correspondre aux effectifs (ou fréquences) des diverses modalités du caractère étudié, des surfaces proportionnelles, surfaces pouvant être rectangulaires ou circulaires.

**Caractères qualitatifs**

Les types de présentations les plus courantes pour les caractères qualitatifs sont les **diagrammes à barres** (ou en **tuyaux d'orgue**) et les **diagrammes circulaires à secteurs** (ou en "**Camembert**").

**Diagramme circulaire**

Dans ces diagrammes, les modalités sont représentées par des portions de disque proportionnelles à leur effectif, ou à leur fréquence. Autrement dit, on découpe un cercle en « tranches » dont la surface est proportionnelle à l'effectif ou la fréquence.

L'angle au centre qui détermine la part de secteur de chaque modalité est proportionnel à la fréquence (ou à l'effectif).

Ce type de diagramme est très utilisé surtout pour les caractères qualitatifs. On le rencontre beaucoup dans la vie de tous les jours (surtout dans la presse écrite). Il fait partie des graphiques de communication.

Quant au **diagramme en bande** ou **en barre** ou **en tuyaux d'orgue**, il est composé des bandes d'une certaine « épaisseur » dont les longueurs / hauteurs sont proportionnelles aux effectifs (ou aux fréquences) correspondants.

**Caractères quantitatifs**

Les caractères quantitatifs sont habituellement représentés par les **diagrammes en bâtons** (caractère quantitatif discret) et les **histogrammes** (caractère quantitatif continu : cf. *encadré 10*).

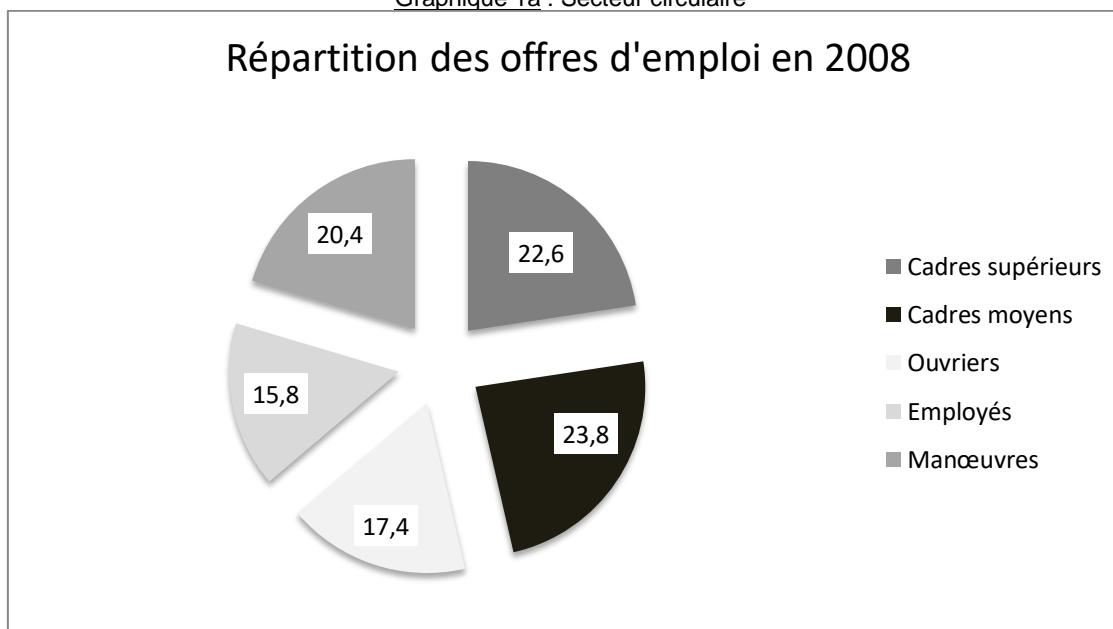
**Nota Bene**

Comme pour le caractère qualitatif, le **secteur circulaire** peut représenter un caractère quantitatif discret ou même continu. C'est un diagramme qui se prête très bien à la représentation des pourcentages.

Enfin, en plus du secteur circulaire, le diagramme à bande ou rectangulaire sont également utilisés pour représenter les caractères qualitatifs.

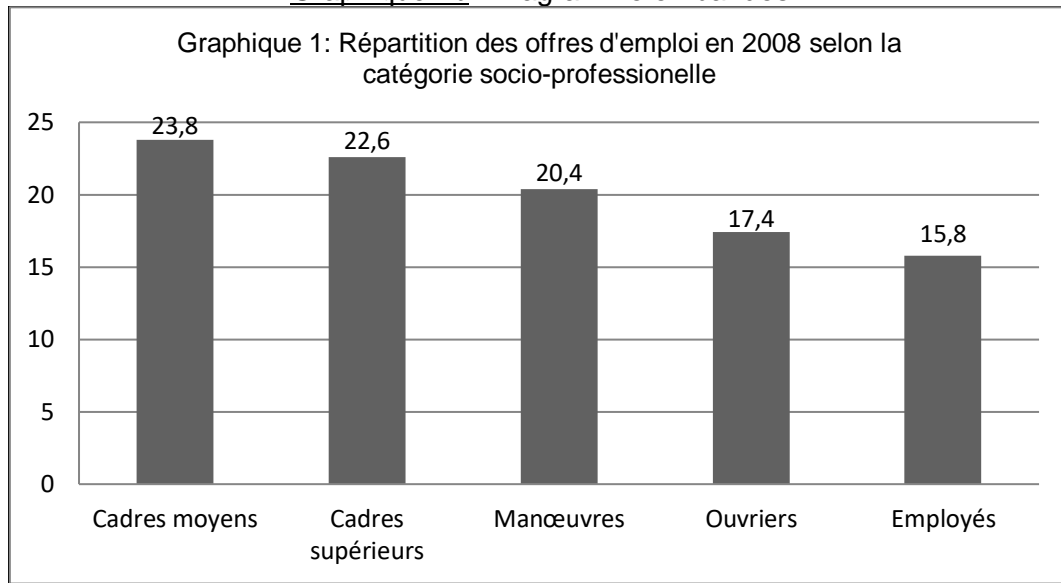
**Graphiques N°1a et N°1b : Offres d'emplois enregistrés en 2008**

Graphique 1a : Secteur circulaire



Source : BSIMT 02 et 03 / DOEF/ANPE

Graphique 1b : Diagramme en bandes



Source : BSIMT 02 et 03 / DOEF/ANPE

**Exemple 2 : Représentation graphique d'une distribution à un caractère quantitatif : diagramme en bâtons et histogramme**

**A - Caractères quantitatifs discrets : diagramme en bâtons**

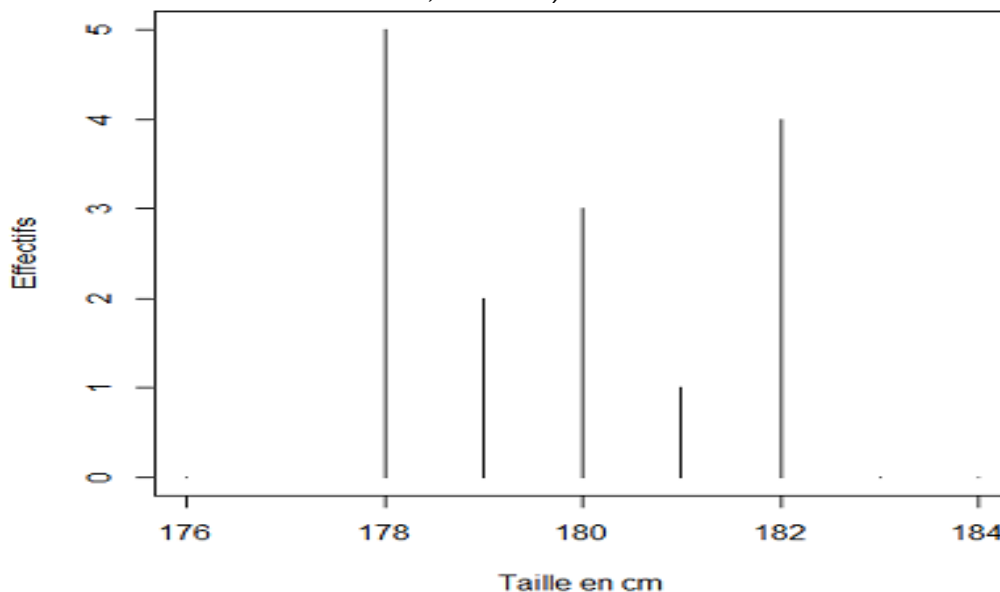
**Exemple 2 :** Distribution d'une population selon la taille.

**Tableau N°13 :** taille de 15 individus

Taille en cm	Effectif	Fréquences
178 cm	5	33%
179 cm	2	13%
180 cm	3	20%
181 cm	1	07%
182 cm	4	27%
<b>Total</b>	<b>15</b>	<b>100%</b>

En élevant, pour chaque mesure de taille, un trait vertical proportionnel au nombre d'élèves. Nous obtenons un diagramme en bâtons (graphique N°3).

**Graphique N°02 :** Diagramme en bâtons de la répartition d'une population de 15 individus selon la taille (en vertical, le nombre d'individus et en horizontale, les tailles).



**Exercice d'application**

Utiliser Excel pour représenter le tableau N°08 de la page 14.

**B – Caractère quantitatif continu : Histogramme**

La représentation graphique des effectifs ou des fréquences d'une distribution statistique d'une variable continue s'appelle **histogramme** (diagramme en barres accolées). La population d'une classe est représentée sous forme d'un rectangle dont la surface est proportionnelle à son effectif (ou à sa fréquence).

On représente l'effectif (ou la fréquence) de chaque classe par un rectangle dont la base est l'amplitude de la classe et dont l'aire est proportionnelle à l'effectif ou à la fréquence (selon qu'il s'agisse de tableaux d'effectifs ou de fréquences). *L'histogramme est la juxtaposition de ces différents rectangles.*

#### Encadré 10

##### Histogramme :

C'est le graphique permettant de représenter une distribution continue regroupée en classes : rectangles juxtaposés dont les bases sont les classes, et les surfaces sont proportionnelles aux effectifs (ou fréquences) associés.

Cas des classes d'amplitudes égales : Lorsque les classes sont d'amplitudes égales, on place en ordonnée les effectifs  $n_i$  (ou les fréquences  $f_i$ ).

Cas des classes d'amplitudes inégales : Lorsque les classes sont d'amplitudes inégales, il faut procéder au calcul des effectifs (ou fréquences) corrigé(e)s pour assurer leur proportionnalité par rapport aux aires des rectangles correspondants. On prend généralement l'amplitude la plus faible comme valeur de **référence**.

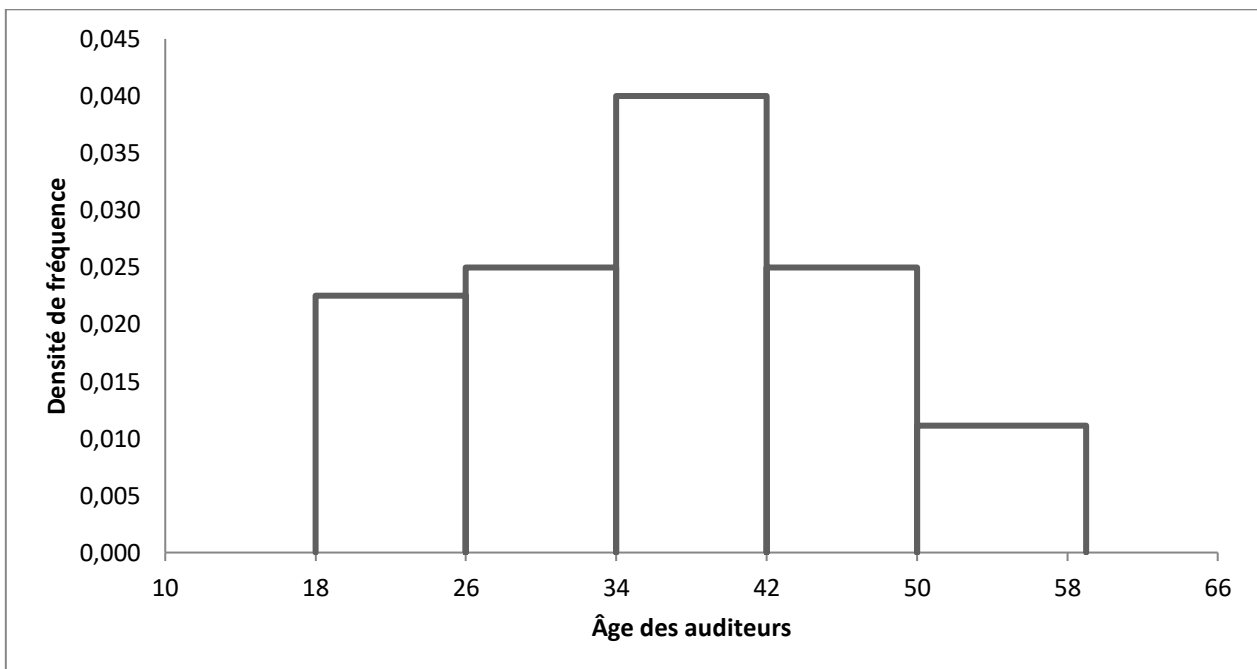
Si les amplitudes  $a_i$  sont différentes, on place  $n_i/a_i$  (ou  $f_i/a_i$ ). On parle d'**effectifs corrigés** (ou de **fréquences corrigés**).

$$\text{effectif corrigé} = \frac{\text{effectif de la classe}}{\text{amplitude de la classe}} \times \text{amplitude de référence} = \frac{n_i}{a_i} a_0$$

$$\text{fréquence corrigée} = \frac{\text{fréquence de la classe}}{\text{amplitude de la classe}} \times \text{amplitude de référence} = \frac{f_i}{a_i} a_0$$

#### **Exemple 2** : Distribution de 50 auditeurs selon les classes d'âge

**Graphique N°03** : Histogramme de la distribution de 50 auditeurs selon les classes d'âge



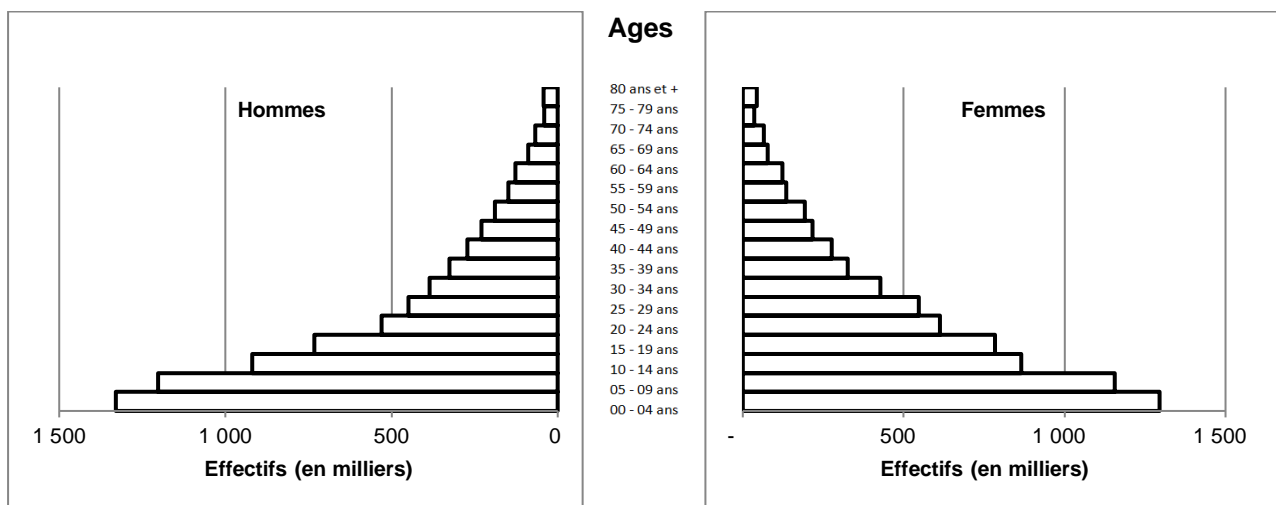
**Encadré 11****La pyramide des âges : un double histogramme**

La pyramide des âges est un outil utilisé pour analyser, à un moment donné, la composition d'une population répartie par groupe d'âges (généralement quinquennaux) et par sexe.

La pyramide des âges représente **deux histogrammes accolés** (placés dos à dos), l'un représentant la distribution par classe d'âge de la population masculine (placé généralement à gauche) et l'autre décrivant toujours par classe d'âge la structure de la population féminine (placé à droite).

Les fréquences sont portées sur les abscisses (partie gauche pour les hommes et la partie droite pour les femmes) et les âges sur l'axe des ordonnées.

**Exemple :** Pyramide des âges de la population résidente du Mali en 2009 (RGPH 2009)

**Graphique N°03 : Pyramide des âges de la population résidente**

Source : RGPH 2009, DNSI

**3.2. Représentation graphique d'une distribution à deux variables**

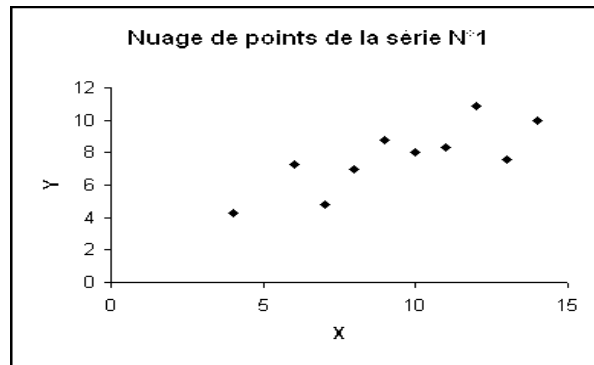
Les **nuages de points** et les **courbes linéaires** sont les deux types principaux de représentations graphiques des distributions à deux variables.

**Exemple :** Distribution conjointe des valeurs de 2 caractères X et Y de 11 individus

**Tableau N°14 :** distribution conjointe de 2 variables

N° des individus	Variable N°1 (X)	Variable N°1 (Y)
1	10	8,04
2	8	6,95
3	13	7,58
4	9	8,81
5	11	8,33
6	14	9,96
7	6	7,24
8	4	4,26
9	12	10,84
10	7	4,82
11	5	5,68

Source : Extrait du quartet d'Anscombe

**Graphique N°04** : Nuage des points de la répartition de la série double (X ; Y)

❖ => Notez bien

Lorsqu'on examine des séries à deux variables, **il est toujours recommandé d'effectuer au préalable une représentation graphique des nuages de points.** et d'éviter beaucoup de déboires dans les commentaires ou les conclusions.

## Chapitre 4 : Résumé numérique des données

### 4.1. Résumés numériques des distributions à une variable

#### Introduction

Les chapitres précédents ont concerné les descriptions visuelles des séries statistiques, toutes plus ou moins appropriées selon le contexte de la série. Le dernier niveau de description statistique est le résumé numérique d'une distribution statistique par ce qu'on appelle « **paramètres caractéristiques** ». Il s'agit de **caractériser, de résumer l'information contenue dans une série statistique par une valeur numérique**.

#### Encadré 12

##### **Les paramètres caractéristiques**

Enfin, pour des besoins d'analyse statistique, il est nécessaire de résumer des séries de répartition résultant d'un grand nombre d'observations en quelques paramètres qui les caractérisent : les « *résumés descriptifs* » ou « **paramètres caractéristiques** ». Ce sont des descriptions « plus compressées » d'une série statistique. Il existe à cet effet toute une gamme de statistiques que l'on peut utiliser à des fins descriptives. Il s'agit de **critères qui quantifient différentes caractéristiques** de la distribution des observations et qui répondent aux questions de savoir si les observations sont-elles *centrées* autour d'une valeur ou sont-elles *groupées* autour de certaines valeurs ?

Quelques statistiques simples permettent de décrire les observations. Il s'agit essentiellement (à un premier niveau) du mode, de la médiane, de la moyenne, de la variance et de l'écart type. On peut citer à un second niveau, les quantiles (généralisation de la médiane), la médiale, l'étendue de la distribution (appelée range), l'interquartile et du coefficient de variation (ou écart type relatif).

Le mode, la médiane, la médiale et la moyenne sont nommés **caractéristiques (ou indicateurs, critères, paramètres) de position ou tendances centrales** et les autres appartiennent à la catégorie des **caractéristiques (ou indicateurs, critères, paramètres) de dispersion**.

**N.B. : Excepté le mode, les caractéristiques ne peuvent être calculées que pour les distributions des caractères quantitatifs.**

#### 4.1.1. Caractéristiques de valeur centrale

Les caractéristiques de valeur centrale (ou critère de position) d'une population statistique étudiée, selon une variable quantitative, sont ses valeurs qui caractérisent globalement la position de cette population.

##### 4.1.1.1. Le Mode

**Exemple 1**: Offres d'emploi enregistrées auprès des services d'intermédiation en 2008 selon la qualification professionnelle en 2008 (*tableau N°01 page 2*).

L'effectif le plus élevé est 411 qui correspond à la qualification demandée « *Manœuvre et assimilé* ». Cette modalité du caractère « qualification demandée » qui a l'effectif le plus élevé est appelée **Mode**.



**Exemple 2 :** Répartition (en %) des ménages du Mali selon le milieu de résidence et la taille du ménage (cf. tableau N°03).

Le mode est « 9 membres habituels et plus » aussi bien pour chacun des deux milieux de résidence que pour l'ensemble du Mali.

**Tableau N°03.1 :** Répartition (en %) des ménages du Mali selon le milieu de résidence et la taille du ménage.

Nombre de membres habituels (taille du ménage) (1)	Urbain		Rural		Ensemble Mali	
	Pourcent (2)	Pourcent cumulé (3)	Pourcent (4)	Pourcent cumulé (5)	Pourcent (6)	Pourcent cumulé (7)
0	0,2	0,2	0,0	0	0,1	0,1
1	6,7	6,9	4,3	4,3	5,0	5,1
2	8,8	15,7	9,0	13,3	8,9	14
3	11,9	27,6	13,8	27,1	13,2	27,2
4	13,7	41,3	14,2	41,3	14,1	41,3
5	15,5	56,8	13,3	54,6	14,0	55,3
6	11,9	68,7	11,5	66,1	11,7	67
7	9,6	78,3	9,8	75,9	9,7	76,7
8	5,0	83,3	7,3	83,2	6,6	83,3
9 ou plus	16,7	100	16,8	100	16,7	100
<b>Total</b>	<b>100,0</b>	<b>0,2</b>	<b>100,0</b>	<b>0</b>	<b>100,0</b>	<b>0,1</b>
<b>Taille moyenne des ménages</b>		<b>5,6</b>		<b>5,7</b>		<b>5,7</b>

Source : EDSM – IV – 2006 ; CPS-MS/DNSI/Macro International Inc., décembre 2007

**Exemple 2 :** Répartition de la population de 12 ans et plus par degré de pauvreté selon l'état matrimonial.

Considérons le tableau N°15 ci-dessous issu de l'EMEP 2001 et relatif à la répartition de la population de 12 ans et plus par degré de pauvreté et selon l'état matrimonial.

Le mode correspondant à l'effectif le plus élevé, est : les modes pour les différents groupes (sous population) qui sont respectivement 514 327 pour les « moins pauvres », 413 622 pour le groupe « vulnérables », 310 051 pour celui des « pauvres » et 1 288 591 pour le groupe des « très pauvres »

#### Remarque sur les « groupes de pauvreté »

En partant du seuil de pauvreté fixé, les quatre groupes ont été définis comme suit :

Moins pauvres :  $\text{Dépenses/tête} \geq 150\% \text{ du seuil}$

Vulnérables :  $100\% \text{ du seuil} \leq \text{Dépenses/tête} < 150\%$

Pauvres :  $75\% \text{ du seuil} \leq \text{Dépenses/tête} < 100\%$

Très pauvres :  $\text{Dépenses/tête} < 75\% \text{ du seuil}$

**Tableau N°15 :** Répartition de la population de 12 ans et plus par degré de pauvreté et selon l'état matrimonial.

Degré de pauvreté	Moins pauvres	Vulnérables	Pauvres	Très pauvres	Total
<b>Situation matrimoniale</b>					
<i>Célibataire</i>	514 327	413 622	300 496	1 202 829	2 431 274
<i>Marié (e) monogame</i>	447 582	408 769	310 051	1 288 591	2 454 993
<i>Marié(e) polygame</i>	105 967	133 384	144 835	666 134	1 050 320
<i>Veuf (ve)</i>	67 678	66 311	44 242	173 221	351 452
<i>Divorcé(e)</i>	18 674	13 381	10 357	22 852	65 266
<b>Total</b>	<b>1 154 228</b>	<b>1 035 468</b>	<b>809 981</b>	<b>3 353 628</b>	<b>6 353 304</b>

Source : EMEP 2001, Résultats définitifs, DNSI novembre 2003

**Exemple 3** : Répartition des femmes de 15 à 49 ans selon l'âge.

On considère le *tableau N°16* extrait des résultats de l'EDSM-IV, relatif à la répartition des femmes de 15 à 49 ans.

**Tableau N°16** : Répartition des femmes de 15 à 49 ans selon l'âge

Age	Effectif de Femmes	Pourcentage de femmes	Fréquences cumulées croissantes	Fréquences cumulées décroissantes
15-19 ans	3 104	21,3%	21,3%	100,0%
20-24 ans	2 678	18,4%	39,7%	78,7%
25-29 ans	2 625	18,0%	57,7%	60,3%
30-34 ans	1 971	13,5%	71,2%	42,3%
35-39 ans	1 688	11,6%	82,8%	28,8%
40-44 ans	1 354	9,3%	92,0%	17,2%
45-49 ans	1 163	8,0%	100,0%	8,0%
<b>Total</b>	<b>14 583</b>	<b>100,0%</b>		

Source : EDSM – IV – 2006, CPS-MS/DNSI

Dans le cas ci-dessus, le mode n'est pas un nombre. Il correspond à un intervalle de valeurs : le groupe d'âge. « 15-19 ans ». Nous avons ce qu'on appelle une **classe modale**.

Soulignons que la dernière colonne « *cumul des pourcentages* » est obtenue de la façon suivante :

- le premier cumul correspond au premier pourcentage, c'est-à-dire 21,3% ;
- le second s'obtient en faisant la somme du premier pourcentage et du second, soit  $21,3\% + 18,4\% = 39,7\%$  ;
- le troisième est égal à la somme du second cumul et du troisième pourcentage, c'est-à-dire  $39,7\% + 18,0\% = 57,7\%$  et ainsi de suite ;
- etc....

Enfin, le cumul qui va de *haut en bas* est appelé « **cumul croissant** » et donne les **effectifs cumulés (ou fréquence cumulées) croissants**. Dans le sens contraire, on a le « **cumul décroissant** » et les **effectifs cumulés (ou fréquence cumulées) décroissants**.

### **Détermination du mode**

Le mode est l'indice le plus simple à déterminer. Sa détermination ne demande aucun calcul. Il suffit de regarder le tableau des effectifs ou de visualiser un graphique pour « l'apercevoir ».

#### **4.1.1.2. La médiane**

La médiane (notée **Me**) est la valeur centrale qui, lorsque la distribution est ordonnée (croissante ou décroissante), partage la population en deux groupes de même effectif tels que 50% lui soient inférieurs ou égaux et 50% lui soient supérieurs ou égaux. La médiane partage donc la population en deux sous population de tailles égales.

**Exemple 1 : Caractère quantitatif discret**

Reprenons le *tableau N°3.1* ci-dessus relatif à la répartition (en %) des ménages du Mali selon le milieu de résidence et la taille du ménage et considérons les colonnes des « cumuls ».

La médiane est la modalité « *5 membres habituels* » (dans les sous population Urbain (56,8%) et Rural (54,6%)) ainsi que l'ensemble du Mali (55,3%)) correspondant à l'atteinte de 50% des pourcentages cumulés.

**Exemple 2 : Caractère quantitatif continu : *classe médiane***

Reprenons le *tableau N°16*, relatif à la répartition des femmes de 15 à 49 ans et considérons la colonne du « *cumul des pourcentages* ».

On détermine la médiane à l'aide de la colonne des effectifs (ou fréquences) cumulées croissantes ou décroissantes.

Dans le *tableau N°16* on dispose du « *cumul croissant des pourcentages* ». La moitié des pourcentages cumulés (50%) se trouve dans la classe « *25 – 29 ans* » qui contient 57,7%. On dit qu'on a une **classe médiane**.

Ainsi, lorsque la variable quantitative est continue et que les modalités du caractère sont groupées en classes, on parle de ***classe médiane***.

**Remarque :**

Dans le cas de valeurs groupées en classe, *on peut calculer la médiane (comme le mode) par interpolation linéaire* (règle de trois ou calcul géométrique élémentaire), avec *l'hypothèse implicite de l'uniforme répartition des individus à l'intérieur de la classe médiane*.

**4.1.1.5. La moyenne**

La moyenne d'une distribution est la somme des valeurs observées divisée par leur nombre. La moyenne ne peut être définie que sur une variable quantitative.

**a) La moyenne arithmétique simple**

*Notion de moyenne*

Imaginons deux élèves dont un de l'école primaire (on le désignera par Toto) et l'autre de l'école secondaire (lycée) nommé Seyni avec leurs notes de composition.

**Exemple 1** : Notes moyennes d'élèves où toutes les matières ont la même importance (coefficients de pondération = 1)

$$\text{MoyennedeTOTO} = \frac{\text{Notecalcul} + \text{noteEcriture} + \text{noteLecture} + \text{noteRécitation}}{\text{nombretotaldematières}} =$$

$$= \frac{8 + 6 + 4 + 5}{4} = \frac{23}{4} = 5,75$$

Les tableaux des notes de nos deux écoliers peuvent être présentés dans des tableaux comme ci-dessous avec des coefficients par matière tous égaux à 1.

**Tableau N°17** : Notes de Toto élève à l'école primaire.

Matières	Notes de Toto	Coefficient des matières
Calcul	8	1
Ecriture	6	1
Lecture	4	1
Récitation	5	1
<b>Total</b>	<b>23</b>	<b>4</b>

**Remarque** : La moyenne n'est pas nécessairement une valeur possible de la variable.

### b) La moyenne arithmétique pondérée

**Exemple 2** : Notes de Seyni élève de lycée (classe de terminale math) où les matières ont des poids différents (coefficients de pondération différents).

**Tableau N°18** : Notes de Seyni, élève de lycée.

Matières	Notes de Séyni	Coefficients	Notes pondérées
Mathématique	14	6	84
Physique	13	3	39
Sciences de la nature	12	2	24
Histoire – Géographie	11	2	22
Littérature	9	2	18
Philosophie	8	3	24
Anglais	16	1	16
Chinois	18	1	18
<b>Total</b>	<b>–</b>	<b>20</b>	<b>245</b>

La colonne des notes indique les notes obtenues par Seyni à la composition. La colonne coefficient marque l'importance de la matière. La note de physique avec un poids de 3 équivaut à avoir 3 notes de physique répétées (13, 13, 13).

Pour calculer la moyenne des notes de Seyni, on multiplie sa note obtenue dans chaque matière par son coefficient (ou son poids). On obtient les notes coefficientées. On fait la somme de ces produits et on divise la somme des notes coefficientées par la somme des coefficients (ou somme des pondérations).

$$\begin{aligned} \text{MoyennedeSeyni} &= \frac{\text{Math} * \text{coefmath} + \text{Phy} * \text{coefphy} + \dots + \text{Ang} * \text{coefAng} + \text{Ch} * \text{coeCh}}{\text{coefMath} + \text{coefPhy} + \text{coefSc} + \dots + \text{coefAng} + \text{coefCh}} \\ &= \frac{14 * 6 + 13 * 3 + 12 * 2 + 11 * 2 + 9 * 2 + 8 * 3 + 16 * 1 + 18 * 1}{6 + 3 + 2 + 2 + 2 + 3 + 1 + 1} = \frac{245}{20} = 12,25 \end{aligned}$$

**La moyenne arithmétique est la somme des valeurs de la variable divisée par le nombre d'individus**

**Calcul de la moyenne d'une distribution avec des valeurs groupées en classe.**

**Exemple 2 :** Age moyen des femmes de 15 à 49 ans.

Reprenons le *tableau N°16*, relatif à la répartition des femmes de 15 à 49 ans et élaborons le *tableau 16.2* par l'ajout de trois (3) nouvelles colonnes.

**Tableau N°16.2 :** Répartition (en %) des femmes de 15 à 49 ans

Age (1)	Effectif de Femmes (2)	Pourcentage de femmes (3)	Centre de Classe (4)	Centre classe x effectif (5)	Centre classe x fréquence (6)
15-19 ans	3 104	21,3%	17	52768	3,621
20-24 ans	2 678	18,4%	22	58916	4,048
25-29 ans	2 625	18,0%	27	70875	4,86
30-34 ans	1 971	13,5%	32	63072	4,32
35-39 ans	1 688	11,6%	37	62456	4,292
40-44 ans	1 354	9,3%	42	56868	3,906
45-49 ans	1 163	8,0%	47	54661	3,76
<b>Total</b>	<b>14 583</b>	<b>100,0%</b>	-	<b>419616</b>	<b>28,807</b>

Source : EDSM – IV – 2006 ; CPS-MS/DNSI/Macro International Inc., décembre 2007

Pour calculer l'âge moyen, on détermine les **centres des classes** (*colonne 4*), c'est-à-dire les milieux des intervalles d'âge par la demi-somme des extrémités des classes. Ainsi, le premier centre de classe vaut :

$$\frac{15+19}{2} = 17, \text{ etc. les centres des classes représentent les classes en tant que modalité. Par la suite, on les}$$

multiplie par les effectifs de classe (*colonne 5*) ou par les fréquences de la classe (*colonne 6*).

La somme de ces produits « *centres de classe \* effectifs* » (ou « *centres de classe \* fréquences* ») rapportée à l'effectif total (fréquence totale) donne l'âge moyen des femmes.

$$\text{L'âge moyen} = \bar{X} = \frac{\text{sommedesproduits}(\text{centresdeclasse} * \text{effectif})}{\text{effectiftotaldesfemmes}} =$$

$$= \frac{52768 + 58919 + \dots + 54661}{14583} = \frac{419616}{14583} = 28,8 \text{ ans}$$

$$\text{Age moyen} = \bar{X} = \frac{419616}{14583} = 28,8 \text{ ans} = 28 \text{ ans } 9 \text{ mois et } 12 \text{ jours}$$

D'où l'âge moyen  $\bar{X}$  vaut environ **29 ans**

**Encadré 13 : Remarques**

1. Si les données sont groupées en classes, alors  $x_i$  est le centre de classe. Dans le tableau de distribution du caractère, on indique les centres de classe. Ce sont eux qui sont fondamentaux pour les calculs.
2. On rappelle que derrière cette représentation de la classe par le centre de classe, il y a l'hypothèse implicite de la répartition homogène des effectifs dans une classe. Le centre de classe est l'équivalent du centre de gravité d'un corps où toute la masse est supposée concentrée. Sans cette hypothèse implicite, beaucoup de pans de la Statistique (qu'elle soit descriptive ou inférentielle n'existerait pas).
3. Excepté le mode, les caractéristiques ne peuvent être calculées que pour les distributions des caractères quantitatifs.

**4.1.1.5. Les Quantiles**

Les quantiles sont d'autres types d'indicateurs de position.

**Notation :** Soit une série statistique  $X_n = \{x_1, x_2, \dots, x_n\}$ . On appelle statistique ordonnée de  $X_n$  la série  $X_{(n)} = \{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$  constituée des valeurs de  $X_n$  telle que  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ .

Exemple 1 :

**Tableau 1a : Notes en opérations calcul de six élèves d'une classe du primaire**

N° d'ordre alphabétique	Elève	Notes en calcul et opérations $X_6$
1	Aminata	7
2	Bouba	2
3	Fanta	3
4	Hamidou	9
5	Seyni	5
6	Zakaria	6

**Tableau 1b : Notes en opérations calcul ordonnées de six élèves d'une classe du primaire**

N° de rang	Elève	Notes en calcul et opérations ordonnées $X_{(6)}$
1	Bouba	2
2	Fanta	3
3	Seyni	5
4	Zakaria	6
5	Aminata	7
6	Hamidou	9

Dans l'exemple 1 (**Tableau 1a**),  $x_1$  correspond à la note du 1<sup>er</sup> élève sur la liste alphabétique soit celle de Aminata,  $x_1=7$ . Dans le **tableau 1b**, la note de Aminata est classée au 5<sup>ème</sup> rang, donc elle correspond à  $x_{(5)}$ . Autrement  $x_{(5)} = x_1$ . De même,  $x_{(1)}$  correspond à la plus petite des notes des élèves. Il s'agit de la note de Bouba,  $x_{(1)}=2$ . Cette note correspond dans le **Tableau 1a** à la note  $x_2$  ( $x_{(1)}=x_2$ ).

**Définition :** On appelle **quantile  $x_p$  d'ordre  $p$** , une valeur qui partage la série **statistique ordonnée** en deux sous-ensembles qui contiennent respectivement un **nombre** d'observations à peu près égal à  **$np$**  et  **$n-np$** .

$np=n(1-p)$ . En d'autres termes, les deux sous-ensembles contiennent respectivement une **proportion** d'observations à peu près égale **p** et **(1-p)**.

#### Les quantiles les plus courants :

Les quantiles les plus couramment utilisés sont :

- **La médiane :  $p=1/2$**

La médiane partage la série statistique ordonnée en deux sous-ensembles qui contiennent chacun la moitié des observations (ou à peu près).

- **Les quartiles :  $p=1/4$  (1<sup>er</sup> quartile),  $p=2/4$  (2<sup>ème</sup> quartile=médiane),  $p=3/4$  (3<sup>ème</sup> quartile).**

Les 3 quartiles partagent la série statistique ordonnée en 4 sous-ensembles qui contiennent chacun (environ) un quart (25%) des observations.

- **Les déciles :  $p=1/10, 2/10, \dots, 9/10$**

Les 9 déciles partagent la série statistique ordonnée en 10 sous-ensembles qui contiennent chacun (environ) un dixième (10%) des observations.

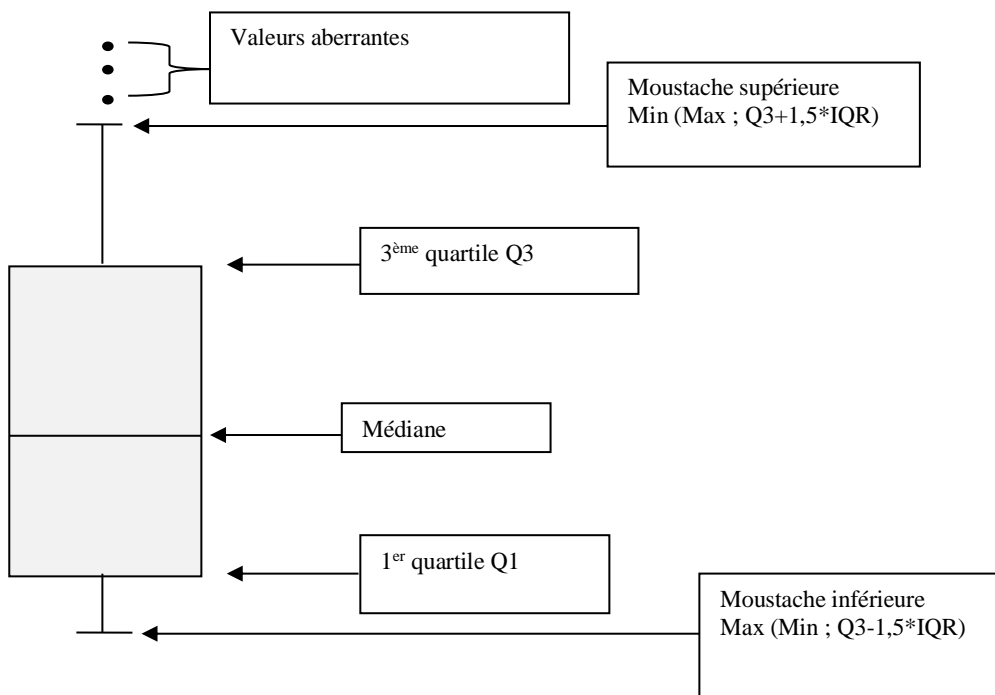
- **Les percentiles (ou centile forme abrégée) :  $p=1/100, 2/100, \dots, 99/100$**

Les 99 percentiles partagent la série statistique ordonnée en 100 sous-ensembles qui contiennent chacun (environ) un centième (1%) des observations.

#### 4.1.1.6. Diagnostics des valeurs extrêmes

Les représentations graphiques permettent de détecter d'éventuelles valeurs aberrantes (outliers) d'une distribution. Une valeur aberrante est une valeur extrême, anormalement différente de la distribution d'une variable. En d'autres termes, **la valeur de cette observation diffère grandement des autres valeurs de la même variable**. Il est important de traiter les valeurs aberrantes car elles influencent certains paramètres statistiques, comme la moyenne. La présence des outliers dans un jeu de données peut fausser la compréhension de ce dernier et conduire à émettre à son égard des hypothèses erronées.

### Détection univariée des valeurs aberrantes : La boîte à moustache ou box plot (anglais)



Dans une boîte à moustache, les points en dessous de la moustache inférieure et ceux au-dessus de la moustache supérieure sont appelés valeurs extrêmes. Dans le cas ci-dessus les trois supérieurs sont considérés comme valeurs aberrantes, aucun point n'étant en-dessous de la moustache inférieure. Les valeurs aberrantes doivent être examinées. Elles peuvent être des valeurs réelles (observées) comme des valeurs erronées.

**Définition** : Un outlier est une valeur est toute valeur extrême supérieure ou inférieure à **I** fois l'intervalle interquartile ( $IQR=Q3-Q1$ ).

### Détection multivariée des valeurs aberrantes : le nuage de points ou scatterplot

Lorsqu'on est amené à détecter les valeurs aberrantes en fonction de plusieurs données, on utilise la représentation graphique du nuage de points. Cette représentation permet de mettre en évidence les valeurs qui s'écartent du comportement général de deux ou plusieurs variables.

#### **4.1.2. Caractéristiques de dispersion**

Considérons l'exemple de la distribution des notes de devoir de Statistique descriptive obtenues par 50 étudiants d'une Ecole de Statistique (*tableau N°18bis*). Après leur premier devoir de Statistique descriptive, le responsable du cours se demande si ses étudiants qui sont hétérogènes par rapport au type de baccalauréat obtenu, sont des niveaux homogènes par rapport à ce devoir.

A partir de la distribution des notes de devoir, le professeur calcule la moyenne de la classe. Par rapport à ce paramètre de position centrale, peut-on affirmer que les étudiants de première année de l'Ecole de



Statistique ont, en Statistique descriptive, à peu près le même niveau ou que les niveaux sont très dispersés ?

**Tableau N°18bis : Notes en statistique descriptive des étudiants**

Elèves	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
BAC	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	C	C	C	C
Note	9	15	16	12	17	17	7	16	17	6	15	9	15	13	15	16	11	6	10	12	10	17	14	12	16
Elèves	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
BAC	C	C	C	C	C	C	C	C	C	C	C	C	C	C	G	G	G	G	G	G	G	G	G	G	G
Note	12	12	18	13	12	15	15	18	13	19	17	14	19	11	9	15	8	12	5	10	7	7	12	12	5

Les paramètres de dispersion permettent justement de mesurer la dispersion des notes de Statistique descriptive de cette classe d'étudiants.

### La variance ( $\sigma^2$ ) et l'écart type ( $\sigma$ )

La **variance** est la somme des carrés des écarts à la moyenne divisée par le nombre d'observations. C'est la moyenne des carrés des écarts à la moyenne.

Elle mesure la dispersion des valeurs autour de la moyenne. La variance étant une somme de carrés, donc une quantité positive dont on peut prendre la racine carrée. La quantité ainsi obtenue s'appelle **écart type**.

#### Encadré 14

##### Les caractéristiques de dispersion

Les paramètres de position ont montré où se situent les valeurs de la variable statistique.

On peut alors se poser la question de savoir si les valeurs sont groupées ou dispersées. La distribution est-elle homogène ? Comment les valeurs du caractère se répartissent :

- les unes par rapport aux autres ;
- les unes par rapport à une valeur donnée (valeur centrale par exemple).

Les paramètres de dispersion d'un caractère quantitatif caractérisent la variabilité des données dans la population.

**L'Etendue ou le Range** est l'intervalle entre la plus petite et la plus grande valeur de la série. C'est la différence entre la plus grande et la plus petite valeur observée.

**Etendue = valeur maximale – valeur minimale = Max – Min**

##### La variance et l'écart type

**moyenne des carrés des écarts à la moyenne.** C'est la différence entre la moyenne des carrés et le carré de la moyenne.

La racine carrée de la variance est **écart type**.

##### Le coefficient de variation ou écart type relatif

L'écart type et la moyenne s'exprimant dans la même unité que le caractère sont dits de dimension un. Leur quotient donne un nombre sans dimension. Ce qui peut permettre aisément d'effectuer des comparaisons de distributions de caractères, quels que soient leurs unités et leurs ordres de grandeur.

Le quotient  $\frac{\text{écart type}}{\text{moyenne}} = \sigma_x / \bar{X}$  est dit **coefficient de variation** ou **écart type relatif**

On peut l'exprimer en pourcentage.

Ce nombre est sans unité, c'est une des raisons pour lesquelles il est parfois préféré à l'écart type qui lui ne l'est pas. En effet, pour comparer deux séries de données d'unités différentes, l'utilisation du coefficient de variation est plus judicieuse



## Chapitre 5 : Notion sur Séries temporelles

### 5.1. Définition et caractéristiques

**Exemple 1** : considérons le dénombrement du nombre journalier de motocyclettes au passage d'un point de Bamako entre 08H et 13H durant trois semaines de juin 1992.

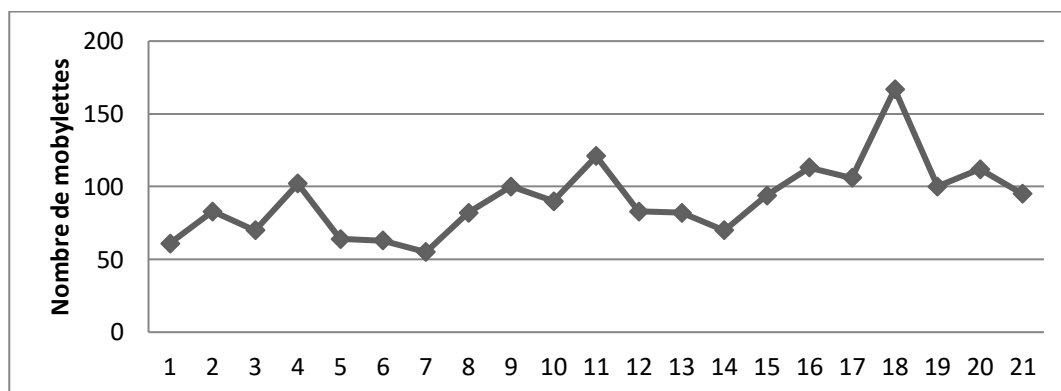
**Tableau N°19** : Dénombrement du nombre journalier de motocyclettes

N°	Jour de la semaine	Nombre mobylette	N°	Jour de la semaine	Nombre mobylette	N°	Jour de la semaine	Nombre mobylette
01	Lundi	61	08	Lundi	82	15	Lundi	94
02	Mardi	83	09	Mardi	100	16	Mardi	113
03	Mercredi	70	10	Mercredi	90	17	Mercredi	106
04	Jeudi	102	11	Jeudi	121	18	Jeudi	167
05	Vendredi	64	12	Vendredi	83	19	Vendredi	100
06	Samedi	63	13	Samedi	82	20	Samedi	112
07	Dimanche	55	14	Dimanche	70	21	Dimanche	95

Source : Direction régionale du plan et de la statistique du District de Bamako

La représentation graphique donne la figure ci-dessous.

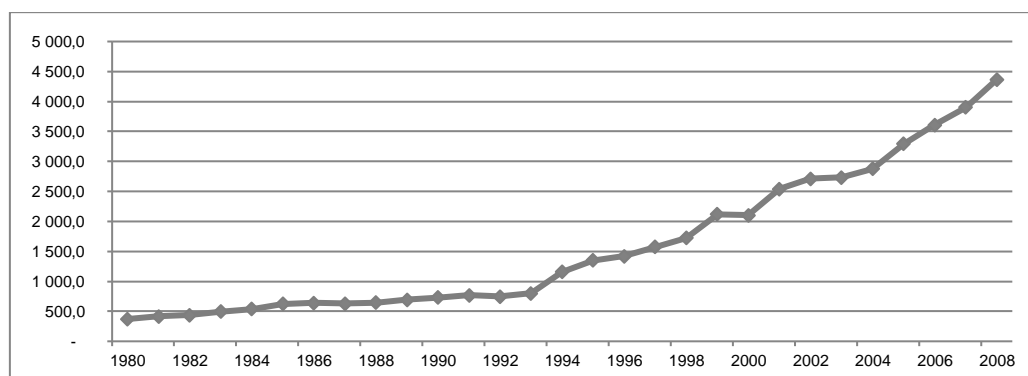
**Graphique N°05** : Chronique du passage hebdomadaire de moto



**Exemple 2** : Série annuelle du Produit intérieur brut.

Considérons l'évolution du Produit intérieur brut nominal/réel du Mali de 1980 à 2008. Les données proviennent des comptes économiques du Mali et extraites de la Base de données nationales de l'INSAT « Malikunnafoni ».

**Graphique N°05** : Evolution du PIB réel (en milliards de F CFA) du Mali de 1980 à 2008



Ces deux exemples de données statistiques mesurées dans le temps (série journalière de passage de motocyclettes, série du PIB) sont appelées **séries chronologiques** ou **chroniques** ou **séries temporelles**. Dans la série journalière de passage de motocyclettes, on remarque sur le graphique que, chaque semaine, le graphique présente une forme identique mais avec des valeurs qui vont croissantes : le phénomène se reproduit dans sa forme. On a un « **phénomène saisonnier** » qui se reproduit à l'identique toutes les semaines avec une tendance à la croissance.

Quant à la série du PIB, ce phénomène ci-dessus signalé n'apparaît pas. Seulement, on constate globalement une **tendance** à la croissance des valeurs du PIB.

#### Encadré 15

##### **Définition et composantes d'une série chronologique**

Une **série temporelle** (ou **série chronologique** ou **chronique**) est une séquence de mesures visant une variable observée au fil du temps, mesures qui sont dépendantes entre elles. C'est une suite de valeurs numériques représentant l'évolution d'un phénomène au cours du temps. La série peut contenir des données observées sur une période infra annuelle (jour, semaine, mois ou trimestre, semestre).

Les observations étant fonction du temps, le traitement des séries temporelles se ramène à la recherche de cette fonction dans le but d'effectuer les prévisions des valeurs futures du caractère.

La détermination de cette fonction revient à la décomposer en ses différents éléments que sont :

- la **composante tendancielle** (ou *tendance* ou *trend*). Elle est relative au comportement de la série sur une longue période.
- la **composante cyclique** (ou **composante extra saisonnière**) qui décrit la succession des phases de croissance et de dépression.
- La **composante saisonnière** (effets saisonniers et de calendrier combinés). Elle est caractérisée par des mouvements de courte période, mouvements infra annuels se répétant régulièrement d'année en année.
- la **composante irrégulière** (ou l'irrégulier) se manifestant dans des variations aléatoires, qui sont des fluctuations imprévisibles causées par des événements indépendants des autres composantes.

Il existe des modèles économétriques pour analyser les séries temporelles.

## 5.2. Notion de désaisonnalisation et correction des variations saisonnières

Les séries chronologiques sont souvent caractérisées par des inflexions saisonnières. L'analyse d'une telle série exige que l'on procède à une " correction des variations saisonnières " afin d'évaluer la tendance sous-jacente.

### **Désaisonnalisation d'une série temporelle**

Désaisonnaliser une série temporelle consiste à déceler la composante saisonnière, à l'estimer et à l'éliminer de la série. On obtient une série qui ne contient plus que la tendance et la composante irrégulière. On l'appelle **série désaisonnalisée** ou **série corrigée des variations saisonnières**.

Les **représentations graphiques annuelles superposées** permettent de déceler la composante saisonnière d'une série.

La **désaisonnalisation** ou la **correction des variations saisonnières** consistent à éliminer l'effet des fluctuations saisonnières sur les données pour ne retenir que des composantes tendancielle et irrégulière.

Cette correction des variations saisonnières s'effectue à l'aide des **coefficients saisonniers** qui sont calculés en utilisant ce qu'on appelle les **moyennes mobiles** calculées sur la « **période-saison** », celle-ci pouvant être, selon la période d'analyse arrêtée, la **journée**, la **semaine**, le **mois**, le **trimestre** ou le **semestre**.

Une fois cette *période-saison* connue, on calcule les moyennes mobiles (MM) sur le nombre de *périodes-saison* couvrant l'année ; le phénomène de « saisonnalité » étant des fluctuations infra-annuelles se répétant régulièrement d'année en année. La *période-saison* plus utilisée dans l'analyse étant le mois, il est très fréquent de rencontrer les moyennes mobiles calculées douze (12) mois (MM12).

Comme leur nom l'indique, les moyennes mobiles sont des moyennes arithmétiques (comme vous les connaissez) mais calculées de façon glissante. C'est-à-dire que, pour calculer la première MM (que l'on doit faire correspondre au milieu de la période), on laisse les observations de la première moitié de la période.

## Annexes

## Annexe 1 : Enoncés de Travaux dirigés

**Exercice N°1 : Application de l'Equilibre lignes – colonnes**

On dispose des deux tableaux de la répartition en (pourcentage) de la population de 12 ans et plus par degré de pauvreté selon l'état matrimonial (élaborés à l'aide des résultats définitifs de l'EMEP 2001).

**Tableau N°A1.1 :** Répartition de la population de 12 ans et plus par degré de pauvreté selon l'état matrimonial

Degré de pauvreté	Moins pauvres	Vulnérables	Pauvres	Très pauvres	Ensemble
<b>Situation matrimoniale</b>					
<i>Célibataire</i>	44,8%	39,9%	37,3%	36,0%	38,4%
<i>Marié (e) monogame</i>	38,7%	39,6%	38,1%	38,4%	38,6%
<i>Marié(e) polygame</i>	9,2%	12,9%	17,9%	19,8%	16,5%
<i>Veuf(ve)</i>	5,8%	6,4%	5,4%	5,1%	5,5%
<i>Divorcé(e)</i>	1,6%	1,3%	1,2%	0,7%	1,0%
<b>Ensemble</b>	100,0%	100,0%	100,0%	100,0%	100,0%

**Tableau N°A1.2 :** Répartition de la population de 12 ans et plus par degré de pauvreté selon l'état matrimonial

Degré de pauvreté	Moins pauvres	Vulnérables	Pauvres	Très pauvres	Ensemble
<b>Situation matrimoniale</b>					
<i>Célibataire</i>	21,2%	16,9%	12,4%	49,5%	100,0%
<i>Marié (e) monogame</i>	18,2%	16,7%	12,6%	52,5%	100,0%
<i>Marié(e) polygame</i>	10,1%	12,7%	13,8%	63,4%	100,0%
<i>Veuf(ve)</i>	19,2%	18,9%	12,6%	49,3%	100,0%
<i>Divorcé(e)</i>	28,6%	20,5%	15,9%	35,0%	100,0%
<b>Ensemble</b>	18,2%	16,3%	12,7%	52,8%	100,0%

En utilisant les deux tableaux des pourcentages ci-dessus, et sachant que la population totale de 12 ans et plus s'élevait à **6 353 304** personnes, déterminer les effectifs croisés du tableau ci-dessous et le remplir.

**Tableau N°A1.3 :** Répartition de la population de 12 ans et plus par degré de pauvreté selon l'état matrimonial

Degré de pauvreté	Moins pauvres	Vulnérables	Pauvres	Très pauvres	Total
<b>Situation matrimoniale</b>					
<i>Célibataire</i>					
<i>Marié (e) monogame</i>					
<i>Marié(e) polygame</i>					
<i>Veuf(ve)</i>					
<i>Divorcé(e)</i>					
<b>Total</b>					<b>6 353 304</b>

Source : EMEP 2001, Résultats définitifs, DNSI novembre 2003

**Exercice N° 2 - Lecture de tableau**

**Repérer et lire** les données suivantes issues de l'EPAM 2007 relatives à la pluriactivité des actifs.

**Tableau N°A1.4** : Taux de pluriactivité en % suivant le sexe et l'âge.

<b>Age</b>	<b>Sexe</b>	<b>Homme</b>	<b>Femme</b>	<b>Ensemble Mali</b>
15 – 19 ans		12,42	14,80	13,84
20 – 24 ans		19,79	21,09	20,57
25 – 29 ans		29,31	21,86	25,13
30 – 34 ans		32,48	21,30	26,67
35 – 39 ans		32,20	18,45	25,94
40 – 44 ans		31,85	25,63	28,71
45 – 49 ans		37,74	23,25	32,11
50 – 54 ans		34,67	23,13	29,86
55 – 59 ans		31,57	19,53	26,91
60 – 64 ans		31,69	14,17	24,75
Ensemble		29,12	20,41	24,69
15 – 19 ans		12,42	14,80	13,84
<b>Total</b>		<b>29,12</b>	<b>20,41</b>	<b>24,69</b>

Source : EPAM 2007/DOEF/ANPE

**Annexe 2 : Correction****Chapitre 1****Exercice N° 1**

- a) la population étudiée est : Ensemble des demandes d'emploi enregistrées auprès des services d'intermédiation en 2008 ;
- b) le(s) caractère(s) étudié(s) est/sont : la qualification professionnelle des demandeurs d'emploi;
- c) le nombre **1 403** représente : Effectif des hommes ouvriers ayant enregistré une demande d'emploi auprès services d'intermédiation en 2008 ;
- d) le nombre **10,9%** représente : Pourcentage des femmes parmi les demandeurs d'emploi ayant un niveau de qualification employé ;
- e) le nombre **8 638** représente : Effectif des hommes toute qualification confondue ayant enregistré une demande d'emploi en 2008 auprès des services d'intermédiation en 2008 .....
- f) le nombre **10 758** représente : Effectif total des demandeurs d'emploi en 2008 auprès des services d'intermédiation en 2008.

**Exercice N° 2**

2°) Soit la répartition des pays de l'UEMOA selon le Produit intérieur brut (PIB). On demande de déterminer :

- a) la population étudiée : Ensemble des pays de l'UEMOA ;
- b) l'unité statistique : Pays ;
- c) le caractère étudié et sa nature : Produit intérieur brut (PIB).

**Exercice N° 3**

3°) On considère les données météorologiques relatives aux températures moyennes mensuelles du Mali au cours de l'année 2001.

Même questions que l'exercice N°2 ci-dessus.

- a) la population étudiée : Les 12 mois de l'année ;
- b) l'unité statistique : mois ;
- c) le caractère étudié et sa nature : Température moyenne.

**Chapitre 2****Exercice 1**

**Questions** : Déterminer :

- a) la population statistique étudiée : Ménages du Mali en 2001
- b) l'individu statistique : ménage
- c) le(s) caractère(s) étudié(s) : Poste de dépense, groupe de consommation
- d) les modalités du (des) caractère(s) étudiés.
- Poste de dépense : 8 modalités
  - Groupe de consommation : 4 modalités

**Exercice N° 2 - Exercez vous à lire les tableaux du second ordre**

**Repérer et lire** les données suivantes indiquées dans le **tableau N°07** :

- a) 1<sup>ère</sup> ligne, 4<sup>ème</sup> colonne : 58 898 –Dépense moyenne en alimentation des ménages « très pauvres » ;
- b) 5<sup>ème</sup> ligne, 2<sup>ème</sup> colonne : 4 222 -Dépense moyenne des ménages « vulnérables » en santé
- c) 7<sup>ème</sup> ligne, 1<sup>ère</sup> colonne : 5 136 –Dépense moyenne en loisirs et éducation des ménages « moins pauvres ».

**Exercice N° 3 – Calcul des effectifs à partir des fréquences (pourcentages)**

On donne ci-dessous le tableau représentant la répartition (en %) des chefs de ménages selon leur situation matrimoniale et leur sexe (données issues de l'ELIM 2003),



**Tableau N°08 bis** : Répartition des chefs de ménages selon leur situation matrimoniale

Etat matrimonial	Sexe	Homme	Femme	Total
Célibataire		9 901	3 646	13 534
Marié(e) monogame		733 773	12 501	746 808
Marié(e) polygame		333 333	7 422	340 800
Divorcé(e)/séparé(e)		4 400	10 417	14 764
Veuf(ve)		18 702	96 230	114 420
<b>Total</b>		<b>1 100 110</b>	<b>130 216</b>	<b>1 230 326</b>

**Tableau N°09** : Répartition des ménages par zone de résidence selon a) leur taille, b) le sexe de leur chef.

Caractéristique	Urbain		Rural		Ensemble Mali	
	Effectif	Pourcentage	Effectif	Pourcentage	Effectif	Pourcentage
<b>a) Taille</b>						
1 et 2 personnes	9 784	2,7	3 561	5,1	54 134	4,4
3 et 4 personnes	40 222	11,1	9 147	13,1	153 791	12,5
5 et 6 personnes	71 022	19,6	14 733	21,1	254 677	20,7
7 personnes et plus	241 330	66,6	42 315	60,6	767 723	62,4
<b>TOTAL</b>	<b>362 357</b>	<b>100</b>	<b>69 826</b>	<b>100</b>	<b>1 230 326</b>	<b>100</b>
Taille moyenne	<b>9,2</b>	-	<b>8,9</b>	-	<b>9</b>	-
Taille médiane	<b>8</b>	-	<b>8</b>	-	<b>8</b>	-
<b>b) Sexe du Chef de ménage</b>						
Homme	301 967	83,3	798 143	92,0	1 100 110	89,4
Femme	60 390	16,7	69 826	8	130 216	10,6
<b>TOTAL</b>	<b>362 357</b>	<b>100</b>	<b>867 969</b>	<b>100</b>	<b>1 230 326</b>	<b>100</b>

**Tableau N°10** : Demandes d'emploi enregistrées par qualification du demandeur et par sexe en 2008

Qualification du demandeur d'emploi	Homme		Femme		Total	
	Effectif	%	Effectif	%	Effectif	%
Cadre supérieur	2 160	25,0	541	25,5	2 700	25,1
Cadre moyen et agent de maîtrise	2 617	30,3	1 007	47,5	3 624	33,7
Ouvrier	1 399	16,2	210	9,9	1 609	15,0
Employé	1 382	16,0	231	10,9	1 613	15,0
Manœuvre et assimilé	1 071	12,4	131	6,2	1 203	11,2
<b>Total</b>	<b>8 638</b>	<b>100</b>	<b>2120</b>	<b>100</b>	<b>10 758<sup>1</sup></b>	<b>100</b>

**2°) Restitution des effectifs des modalités à partir des fréquences**

A l'aide du tableau N°11 ci-dessus, remplissez les cellules vides du tableau N°11bis suivant

**Tableau N°11** : Répartition en pourcentage des demandes d'emploi enregistrées en 2007 par groupe d'âge et par sexe.

	Homme	Femme	Total
< 15 ans	0,32%	0,45%	0,34%
15 à 19 ans	2,56%	2,20%	2,49%
20 à 24 ans	18,41%	21,97%	19,12%
25 à 29 ans	37,16%	43,54%	38,44%
30 à 39 ans	29,23%	27,78%	28,94%
40 à 59 ans	11,45%	3,54%	9,86%
60 ans et plus	0,43%	0,04%	0,35%
Non déclaré	0,45%	0,48%	0,46%
<b>Total</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>

Source : BSIMT 02 et 03 / DOEF/ANPE

<sup>1</sup> La reconstitution donne comme population totale 10 749 soit 9 de moins que le tableau original. Cela est dû au pourcentage de la colonne « Homme » dont la somme au tableau initial est 99.9 et non 100. Cela est dû aux arrondis. Il faudra par conséquent que l'on veuille reconstituer les tableaux.

**Tableau N°11bis** : Répartition des effectifs des demandes d'emploi enregistrées en 2007 par groupe d'âge et par sexe

	Homme	Femme	Total
< 15 ans	34	11	45
15 à 24 ans	2 242	649	2 891
25 à 39 ans	7 098	1 915	9 013
40 à 59 ans	1 224	95	1 319
60 ans et plus	46	1	47
Non déclaré	48	13	62
<b>Total</b>	<b>10 692</b>	<b>2 685</b>	<b>13 377</b>

Source : BSIMT 02 et 03 / DOEF/ANPE

**3°) Détermination d'un effectif à partir d'informations extérieures**

Le *tableau N°12* ci-dessous donne la répartition des actifs en emploi en 2007 à Bamako selon leurs revenus mensuels d'emploi et leur sexe.

**Tableau N°12** : Répartition en % des actifs en emploi selon les revenus mensuels d'emploi

Revenus	Sexe		Ensemble Bamako
	Homme	Femme	
Moins de 29 000 FCFA	18,5%	52,3%	32,5%
[ 29 000, 50 000 [	31,2%	17,9%	25,7%
[ 50 000, 75 000 [	14,6%	11,3%	13,2%
[ 75 000, 100 000 [	12,5%	7,1%	10,3%
[ 100 000, 200 000 [	14,4%	7,5%	11,5%
[ 200 000, 500 000 [	6,3%	2,3%	4,6%
500 000 FCF et plus	0,4%	0,0%	0,2%
Rien du tout	2,1%	1,7%	1,9%
<b>Total</b>	<b>100,0%</b>	<b>100,0%</b>	<b>100,0%</b>

Source : EPAM 2007/DOEF/

Par ailleurs, en cette même année 2007, on disposait des données suivantes :

- le nombre total d'actifs en emploi à Bamako était évalué à **4 108 779** dont **2 020 243** hommes.
- le **SMIG** était fixé à environ **29 000 F**.

**Question** : Calculer le nombre de femmes qui ont un revenu mensuel inférieur au SMIG.

- Nombre de femmes qui ont un revenu mensuel inférieur au SMIG (29 000 F CFA) : % de femmes avec un revenu de moins de 29 000 F CFA X nombre total de femmes actives en emploi soit 1 092 304
- Nombre de femmes actives en emploi : (nombre total d'actifs en emploi) – (nombre d'hommes actifs en emploi) soit 2 088 536

**Annexe 1 : Enoncés de Travaux dirigés (à supprimer !)****Exercice N°1** : Application de l'Equilibre lignes – colonnes**Tableau N°A1.3** : Répartition de la population de 12 ans et plus par degré de pauvreté selon l'état matrimonial

Degré de pauvreté	Moins pauvres	Vulnérables	Pauvres	Très pauvres	Total
<b>Situation matrimoniale</b>					
<b>Célibataire</b>	517 210	412 304	302 519	1 207 636	2 439 669
<b>Marié (e) monogame</b>	446 332	409 547	308 999	1 287 497	2 452 375
<b>Marié(e) polygame</b>	105 878	133 133	144 665	664 619	1 048 295
<b>Veuf(ve)</b>	67 091	66 043	44 028	172 270	349 432
<b>Divorcé(e)</b>	18 170	13 024	10 102	22 237	63 533
<b>Total</b>	<b>1 154 681</b>	<b>1 034 051</b>	<b>810 313</b>	<b>3 354 259</b>	<b>6 353 304</b>

Source : EMEP 2001, Résultats définitifs, DNSI novembre 2003

### **Annexe 3 : Références bibliographiques**

#### **Livres**

1°) Cours de Statistique descriptive, G. CALOT, Editions Dunod, Paris, 1965

2°) Statistique descriptive en science économique, J. DUBOIS, Editions Dunod, Paris, 1970

#### **Publications statistiques**

1°) Enquête Malienne sur l'Evaluation de la Pauvreté 2001 (EMEP 2001), Résultats définitifs, DNSI, novembre 2003

2°) Rapport d'Analyse Situationnelle Annuelle du Marché du Travail 2008 (RASAMT 2008), ANPE/OEF, septembre 2009

3°) Activité, Emploi et Chômage au Mali en 2004, Premiers résultats de l'Enquête Permanente Auprès des Ménages (EPAM) – Mali - 2004, ANPE/OEF, décembre 2004

4°) Bulletin semestriel d'information sur le marché du travail 04 & 05 / DOEF-ANPE

5°) Activité, Emploi et Chômage au Mali en 2007, Premiers résultats de l'Enquête Permanente Auprès des Ménages (EPAM) – Mali - 2007, ANPE/OEF, juin 2008

6°) Rapport d'Analyse Situationnelle Annuelle du Marché du Travail – Année 2007, (RASAMT 2008) ANPE/OEF, décembre 2008

7°) Situation Economique et Sociale du Mali en 2008 et perspectives pour 2009, DNPDI/INSAT

8°) Enquête Légère Intégrée auprès des Ménages ELIM 2006, vol 1 Indicateur de bien être et de gouvernance, DNSI, juillet 2007

9°) Enquête Démographique et de Santé Mali IV-2006 (EDSM-IV – 2006), CPS-MS/DNSI.