

Centre de Recherche sur l'Information Scientifique et Technique

Post graduation spécialisée en ligne

Option Information Scientifique et Technique

## **Module**

# **Recueil et Traitement Statistique des Données: Introduction Générale à la Statistique**

**M. NEKRI**

Responsable du module

Recueil et traitement statistique des données

CERIST, Septembre 2011

# Sommaire

Introduction.....	3
-------------------	---

## Chapitre1: Recueil de données

1. Définition de recueil de données.....	7
2. Les principales méthodes du recueil de données .....	7
2.1 L'interview .....	8
2.2 L'observation.....	8
2.3 L'étude de documents.....	8
2.4 Le questionnaire .....	9
3. Les étapes d'une enquête par questionnaire.....	9
4. Rédaction du questionnaire.....	10
4.1 Choix des questions .....	10
4.2 Mise en forme du questionnaire .....	13
Activité1.....	15

## Chapitre2: Statistique descriptive

1. Terminologie statistique.....	18
2. Tableau statistique.....	19
3. Les différentes représentations graphiques.....	24
4. Paramètres de position.....	27
5. Paramètres de dispersion.....	32
Activité1.....	33
Activité2.....	34

## Chapitre3: Modèle de Morse

1. Introduction.....	41
2. Définitions de base.....	41
3. Formulation du modèle.....	41
4. Mise en œuvre de la formulation du modèle de Morse.....	42
5. Applications du modèle de Morse.....	43
6. Activité1.....	52

## 1. Introduction

La statistique est la science qui procède à l'étude méthodique à partir de modélisations mathématiques, des modes d'utilisation et de traitement de données dans le but de conduire et d'étayer une réflexion ou de prendre une décision en situation concrète soumise aux aléas de l'incertain. Elle comprend un ensemble de techniques et de méthodes devant conduire à l'acquisition de connaissances générales sur la base de données incomplètes, à partir d'un système scientifique rigoureux guidant le recueil de données, leur organisation, leur analyse et leur interprétation, pour autant qu'on puisse leur donner une forme numérique. Le terme statistique désigne à la fois :

- L'ensemble des données numériques concernant une catégorie de faits (sens très ancien qu'on emploie souvent avec un 's'). Le mot a été introduit à l'origine pour caractériser les études méthodiques des faits sociaux par des procédés numériques destinés à renseigner et aider les gouvernants (classements, recensements, dénombrements, inventaires).
- L'ensemble des méthodes mathématiques permettant :
  - a) de résumer quantitativement l'information recueillie sur un ensemble d'éléments au moyen d'une investigation exhaustive.
  - b) de généraliser à de grands ensembles d'éléments les conclusions tirées des résultats obtenus avec des ensembles beaucoup plus restreints appelés échantillons.
- Une mesure calculée à partir de données provenant d'un échantillon (moyenne, variance, covariance etc.).

A partir de cette définition, nous pouvons dégager deux grandes familles de méthodes utilisées en statistique. Il s'agit de:

- La statistique descriptive, qui est un ensemble de méthodes permettant de décrire un ensemble relativement important de données. Ces méthodes permettent, d'une part, une représentation graphique des données (histogramme, diagramme en bâtons, diagramme circulaire, etc.) et d'autre part, d'obtenir un ensemble d'indicateurs qui, d'une certaine façon, résument les données de base (moyenne, écart-type, coefficient de corrélation, indice, etc.).
- La statistique mathématique, dont l'objet est de formuler des lois à partir de l'observation d'échantillons, intervient dans les enquêtes et les sondages. Elle s'appuie principalement sur la statistique descriptive.

La statistique est utilisée dans des domaines très variés comme les sciences humaines, les sciences sociales, les sciences économiques, les sciences de l'éducation, la psychologie, la médecine,...

En sciences de l'information, le champ d'application est très large nous citons à titre non exhaustif:

- Gestion des bibliothèques et systèmes d'information ;
- Etudes des lois bibliométriques ;
- L'analyse des citations et scientométrie ;
- Etude et analyse de la circulation de l'information dans les systèmes d'information ;
- Aspects théoriques en recherche d'information documentaire.

Généralement, la démarche suivie pour mener une étude statistique est composée des étapes suivantes:

- Recueil et collecte des données;
- Traitement des données collectées;
- l'interprétation des données qui s'appuie sur la théorie des sondages et la statistique mathématique.
- la présentation afin de rendre les données compréhensibles par tous

Ainsi à travers ce module qui vise à:

- Présenter et clarifier les fondements et les pratiques du recueil d'information quantitative et qualitative à des fins d'évaluation...etc.
- Savoir élaborer une stratégie d'enquête, élaborer un questionnaire, collecter les réponses et les exploiter statistiquement.
- Se familiariser avec les concepts propres à la Statistique descriptive
- Accroître son aptitude à prendre de meilleures décisions pour une gestion rationnelle des ressources de la bibliothèque

Nous allons traiter, dans une première partie, quelques aspects liés à la procédure de recueil de données à travers les points suivants:

- Définition de recueil de données
- Les principales méthodes du recueil de données
- Les étapes d'une enquête par questionnaire

- Les règles de rédaction d'un questionnaire

La deuxième partie sera consacrée aux traitements statistiques des données recueillées en abordant les points suivants:

- Concepts et notions de base en statistique
- Représentations graphiques et numériques des données statistiques
- Calcul des paramètres de position et de dispersion

# **Chapitre I**

## **Recueil de Données**

## **1. Définition de Recueil de Données**

« Le recueil d'informations est défini comme le processus organisé mis en œuvre pour obtenir des informations auprès de sources multiples en vue de passer d'un niveau de connaissance ou de représentation d'une situation donnée à un autre niveau de connaissance ou de représentation de la même situation, dans le cadre d'une action délibérée dont les objectifs ont été clairement définis, et qui donne des garanties suffisantes de validité »<sup>1</sup>

De façon générale, on est amené à rechercher de l'information ou recueillir des données lorsqu'on désire cerner une situation soit :

- Pour évaluer
- Pour analyser des besoins
- Pour prendre une décision
- Pour améliorer un fonctionnement, des performances
- Pour former
- Pour résoudre un problème
- Pour cerner un phénomène
- Pour tester des hypothèses scientifiques

### **1. Les principales méthodes du recueil de données**

Pour recueillir des données, il est nécessaire d'élaborer une stratégie qui va faire appel à des méthodes. Le choix d'une de ces méthodes doit être basé sur un ensemble de critères fixé au préalable à savoir: le rapport entre le degré de détail et de précision souhaité et les coûts que cela entraîne en termes d'énergie, ainsi qu'en termes de temps d'exécution, de traitement et de disponibilité des résultats finaux.

Les principales méthodes sont :

- Les méthodes face à face "La pratique d'interviews";
- L'observation ;
- Les enquêtes par questionnaires ;
- L'étude de documents.

---

<sup>1</sup> Jean-marie De Ketele, Xavier Roegiers. Méthodologie du recueil d'information-De Boeck, Bruxelles, 1993

## **1.1 L'interview**

L'interview est une méthode de recueil d'informations qui consiste en des entretiens oraux, individuels ou de groupes, avec plusieurs personnes sélectionnées, afin d'obtenir des informations sur des faits ou des représentations, dont on analyse le degré de pertinence, de validité et de fiabilité en regard des objectifs du recueil d'informations<sup>2</sup>.

Une interview est un entretien individuel en face à face avec des personnes issues du groupe cible. Elle peut être basée sur un questionnaire ou consister en une conversation ouverte, ou combiner les deux.<sup>3</sup>

## **1.2 L'observation**

« Observer est un processus incluant l'attention volontaire et l'intelligence, orienté par un objectif terminal ou organisateur et dirigé sur un objet pour en recueillir des informations »<sup>4</sup>

L'observation est un processus dont l'action première immédiate est de recueillir de l'information sur l'objet pris en considération en fonction de l'objectif de l'étude. Ce recueil suppose une activité de codage : l'information brute sélectionnée est traduite grâce à un code pour ensuite être transmise.

## **1.3 L'étude de documents**

L'étude de documents est une des méthodologies qualitatives utilisées dans les sciences sociales et humaines. Elle consiste à faire de la recherche en analysant des documents pour mieux comprendre le phénomène à étudier.

Elle dépend de :

- La nature des documents à analyser
- De la quantité des documents à analyser
- De l'objet et du but de l'investigation

## **1.4 La nature des documents :**

Différents types de documents peuvent être concernés :

- Documents scientifiques
- Documents officiels (textes de lois, programmes d'enseignement, normes...)

---

<sup>2</sup> Jean-marie De Ketele, Xavier Roegiers. Méthodologie du recueil d'information-De Boeck, Bruxelles, 1993

<sup>3</sup> Evaluer des actions de communication. COMM collection, octobre 2006.

<sup>4</sup> Jean-marie De Ketele, Xavier Roegiers. Méthodologie du recueil d'information-De Boeck, Bruxelles, 1993



- Documents à usage limité dans le temps
- Etc.

### **i. La quantité des documents à analyser :**

La quantité des documents à analyser déterminera le type d'analyse à effectuer : une analyse exhaustive, ou une analyse par échantillonnage ou par sélection.

### **ii. L'objet et le but d'investigation :**

Deux types d'analyse des documents sont distingués :

- Le type d'analyse de documents présentant un caractère exploratoire comme par exemple la recherche documentaire, dont l'objectif est l'exploration en vue d'élaborer une problématique (cadre théorique et hypothèse)
- Le type d'analyse présentant un caractère confirmatoire où le document sélectionné selon une stratégie est traité comme une donnée, au même titre que le discours recueilli par l'entretien ou un comportement recueilli par l'observation, et dont l'objectif est de vérifier une hypothèse.

### **iii. Le questionnaire**

Selon qu'il s'agisse d'une évaluation des performances de personnes, ou au contraire d'une évaluation d'un fonctionnement, de l'évaluation d'un système, d'une recherche descriptive ou expérimentale, le questionnaire prendra deux sens différents.

Sens 1 : Le questionnaire de contrôle de connaissances qui permettra de recueillir des données sur soit une activité demandée à un individu (donner une définition, construire...) soit un contenu sur lequel s'exerce cette activité (une formule, une démarche à suivre etc.)

Sens2 : Le questionnaire d'enquête concerne l'enquête au sens d'une étude d'un thème précis auprès d'une population dont on détermine un échantillon afin de préciser certains paramètres.

## **2. Les étapes d'une enquête par questionnaire**

Une enquête par questionnaires est un vrai projet : elle implique des objectifs clairs, une méthodologie et une organisation rigoureuse, une planification précise et, bien sûr, des investissements parfois importants en temps et en argent.

Les étapes essentielles de la réalisation d'un questionnaire d'après Claude Javeau<sup>5</sup> sont:

---

<sup>5</sup> Claude Javeau. L'enquête par questionnaire-Editions de l'université de Bruxelles, Belgique, 1982

- Définition de l'objet de l'enquête
- Inventaire des moyens matériels mis à la disposition des réalisateurs de l'enquête (problème de la durée de l'enquête, le budget alloué à l'enquête.)
- Recherches préalables
- Recherche de la documentation sur le sujet : autres enquêtes, rassemblement des données énumératives, etc.
- Détermination des objectifs et des hypothèses de travail (que veut-on montrer ?
- Détermination de la population ou univers de l'enquête (public ciblé)
- Construction de l'échantillon (taille de l'échantillon, procédé de l'échantillonnage, etc.)
- Rédaction du projet du questionnaire (choix de la méthode d'administration, choix des questions, mise en forme etc.)
- Mise à l'épreuve du questionnaire (pré-test ou enquête pilote)
- Rédaction du questionnaire définitif (mise en page, pré codage, etc.)
- Formation des enquêteurs
- Réalisation matérielle de l'enquête (collecte des questionnaires remplis, contrôle des questionnaires etc.)
- Codage des questionnaires
- Dépouillement des questionnaires
- Analyse des résultats

#### **4. Rédaction du questionnaire**

Après que les étapes préparatoires ont été achevées et la décision a été prise pour utiliser la méthode du questionnaire comme outil d'investigation, il convient de rédiger le questionnaire. Dans la première épreuve, il convient de rassembler toutes les questions que l'on voudrait poser et qui correspondent aux objectifs définis par les hypothèses de travail retenues.

Le questionnaire ne prendra sa forme définitive qu'après avoir défini la forme des questions, l'ordre des successions des questions, les items.

##### **4.1 Choix des questions**

###### **4.1.1 Questions fermées**

Dans ces questions, les réponses sont fixées à l'avance et le répondant doit choisir parmi l'éventail qui lui est présenté.

*Exemple :*

Lisez-vous des périodiques rédigés dans une autre langue que votre langue maternelle ?

Régulièrement

Occasionnellement

Rarement

Jamais

### ***Caractéristiques***

Ce type de question est celui qui se prête le mieux au dépouillement et à l'analyse statistique. En effet, on peut répartir les différents répondants selon la réponse qu'ils ont fourni sans passer par des étapes d'analyse intermédiaires.

- Ces questions se comprennent facilement et on peut répondre facilement.
- Elles garantissent un certain anonymat
- Elles peuvent servir de questions filtres, c'est-à-dire de discriminants servant de répartir les répondants entre plusieurs séries de réponses ultérieures.

Cependant, l'inconvénient de ces questions est de dicter la réponse au répondant n'autorisant aucune expression de nuance. Leur champ d'application se limite au recueil de caractéristiques objectives.

### ***Exemple de questions filtres***

Jugez-vous la mise en place d'un centre de documentation nécessaire :

Oui

Non

Si oui, pourquoi ?

Accès rapide à l'information

Disponibilité d'information actualisée

### **4.1.2 Questions ouvertes**

Dans les questions ouvertes, la réponse n'est pas prévue et le répondant est libre de s'exprimer comme il veut.

On réserve, dans le questionnaire, un emplacement suffisant pour que la réponse de la personne puisse être enregistrée en son entier.

***Exemple***

Quelle est, à votre avis, la meilleure formule d'examen possible dans l'enseignement universitaire ?

***Caractéristiques :***

- Si ces questions sont bien formulées, on pourra obtenir de bonnes informations sur n'importe quel sujet.
- Elles sont indispensables pour recueillir des renseignements sur des problèmes délicats.
- Leur utilisation s'impose lorsqu'on ne peut prévoir les réponses possibles, dans ce cas on essayera d'abord d'énoncer une question de type semi ouverte.
- Leur formulation est très délicate : elles doivent être aisément comprises et ne comporter aucune ambiguïté ou contresens.
- Leur dépouillement est plutôt difficile, il faut souvent ressortir les occurrences pour pouvoir faire l'analyse statistique.

**4.1.3 Questions semi ouvertes ou semi fermées**

Les principales réponses possibles sont prévues, comme dans une question fermée, mais on laisse la possibilité d'ajouter des réponses libres, en dehors de l'éventail proposé, comme dans une question ouverte.

***Exemple :***

Avez-vous déjà utilisé des produits documentaires ?

Bulletin signalétique ☐

Bulletin des sommaires ☐

Liste des acquisitions ☐

Catalogue imprimé de la bibliothèque ☐

Autres : .....

***Caractéristiques :***

Elles contribuent surtout à faciliter le dépouillement étant donné que le nombre de réponses est déjà prévu. Elles risquent d'influencer la réaction du répondant, par la suggestion de réponses qui peuvent paraître plus habituelles ou moins habituelles et la partie questions fermées doit être bien établie.

**4.2 Mise en forme du questionnaire**

Après avoir défini le contenu du questionnaire, c'est-à-dire les divers domaines que devront aborder les questions, il est nécessaire de procéder à sa mise en forme.

La présentation générale doit être claire et aérée. Le questionnaire doit disposer d'un titre motivant suivi des éléments ci-dessous :

A. L'introduction

- brève, dynamique
- se présenter
- indiquer les objectifs de l'enquête
- préciser le délai de réponse

B. Premières questions

- mettent l'interrogé en confiance
- questions faciles
- questions de mise en valeur de l'interrogé (si nécessaire)

C. Questions essentielles

- disposées selon un ordre soigneusement étudié
- présenter les questions par thème
- risque : la logique des questions peut amener une personne à répondre en veillant à ne pas se contredire. Il faut alors "casser" la cohérence des questions et "dissimuler" les questions importantes

D. Questions personnelles

- placées, généralement, à la fin pour éviter un réflexe prématuré de rejet
- le questionnaire doit être anonyme

- on peut toutefois vouloir connaître certaines caractéristiques de la personne (âge, catégorie socioprofessionnelle)

#### E. Remerciements

- directs et brefs
- peuvent être précédés par un espace destiné aux remarques

En résumé voici ci-dessous ce qu'il faut ou ne faut pas faire.

Ne pas commencer le questionnaire par des questions :

- Impliquant un engagement personnel du répondant ;
- Provoquant des efforts particuliers de réflexion ;
- Traitant de problèmes délicats ;

Ne pas utiliser de questions :

- Abordant directement des thèmes délicats ;
- Pouvant suggérer certaines réponses ;
- pouvant faire préférer la réponse positive ;
- comportant des noms de personnalités pouvant provoquer identification ou rejet ;
- rédigées dans un langage compliqué, peu accessible ;
- trop longues ;

Veiller soigneusement :

- à disperser les questions susceptibles de provoquer « l'effet de halo »<sup>6</sup>
- que le passage d'un thème à un autre ou d'une méthode d'interrogation à une autre ne provoque pas de retrait de la part de l'enquête soit :
- en assurant harmonieusement les transitions (questions neutres, questions progressives, etc.)
- en préparant convenablement l'enquête à répondre après que le passage aura été marqué de manière expresse ;
- que la préparation générale du répondant au questionnaire est bien assurée : texte d'introduction incorporé au questionnaire ou correctement récité par l'enquêteur ;
- que le questionnaire ne dépasse pas une longueur optimale : 45 à 60 minutes, 20 à 30 questions au maximum.

---

<sup>6</sup> Effet de halo : l'irritation ressentie à l'égard d'une question peut se prolonger sur d'autres questions, ou bien le répondant peut être tenté, par souci de logique, de fournir un ensemble de réponses cohérentes, s'alignant sur les premières réponses fournies. «Définition donnée par Claude Javeau dans son ouvrage L'enquête par questionnaire»

- Grouper, si possible, les questions relatives à un même sujet ;
- Regrouper les questions d'identification (sexe, âge, profession...) au début ou à la fin du questionnaire ;
- Commencer par des questions « faciles » susceptibles de mettre le répondant en confiance ;
- Intercaler entre les groupes de questions difficiles ou délicates des questions plus faciles ;
- Donner au questionnaire un aspect cohérent et logique, tout en tenant compte de la possibilité de perturbations qui peuvent être provoquées chez le répondant.

### **Activité 1 :**

L'entreprise dans laquelle vous exercez ayant vécu sans système documentaire organisé et formalisé a éprouvé le besoin de créer un service de documentation. Ce service doit fonctionner au bénéfice des divers secteurs de l'entreprise (recherche et développement, organisation, production, gestion, commercial et marketing...).

Avant la mise en place de ce service de documentation, l'entreprise a décidé de procéder à une enquête auprès de l'ensemble des cadres. Dans ce sens, l'entreprise a fait appel à vous pour procéder à une étude des besoins en effectuant une enquête par questionnaire auprès des utilisateurs, qui portera sur les points suivants :

- connaissance des utilisateurs potentiels, de leurs fonctions,
- de l'information recherchée (domaines, types , formes , délais , des sources d'information utilisées , ...) Ce travail se soldera par la rédaction d'un questionnaire.

### **Tâches à réaliser**

#### **1. Travail individuel :**

Proposez les thèmes du questionnaire et rédigez une série de questions.

Chaque apprenant lira le cours et consultera les ressources, - proposera les thèmes du questionnaire, - rédigera une série de questions.

## **2. Travail collaboratif par équipe:**

Par équipe de 3 apprenants

Les membres de l'équipe discuteront les différentes questions proposées et mettront en forme le questionnaire de l'équipe.

Les apprenants doivent déposer le questionnaire dans l'espace documents

## **3. Travail de tout le groupe**

Les équipes se consulteront pour le choix du questionnaire définitif du groupe.



## **Chapitre II**

# **Statistique Descriptive**

Ce chapitre présente le champ de la statistique descriptive ainsi que son vocabulaire de base.

## 1. Terminologie statistique

**Population:** C'est l'ensemble des unités ou individus sur lequel on effectue une analyse statistique. Elle est définie comme un ensemble d'unités statistiques (par exemple des humains, des objets...) de même nature sur lequel on recherche des informations quantifiables. Elle constitue l'univers de référence lors de l'étude d'un problème statistique.

### Echantillon

C'est un ensemble d'individus prélevés dans une population déterminée

**Individu :** un élément de la population est appelé individu (l'unité élémentaire constituante de la population).

**Caractère :** C'est la propriété ou l'aspect singulier que l'on se propose d'observer dans la population ou l'échantillon. Un caractère qui fait le sujet d'une étude porte aussi le nom de variable statistique. Un caractère est dit **quantitatif** s'il est mesurable. Un caractère est dit **qualitatif** s'il est repérable sans être mesurable.

Si la variable quantitative ne prend que quelques valeurs, pris parmi un ensemble limité, la variable est discrète. Si elle prend n'importe quelle valeur d'un intervalle, la variable est continue. Pour quantifier la variable continue, on utilise des intervalles d'amplitude égale.

**Effectif :** Le nombre d'individus associé à chaque valeur (classe ou modalité) de la variable est l'effectif. La somme de tous les effectifs est l'effectif total de la population. Les effectifs peuvent donner lieu à la présentation de colonnes d'effectifs cumulés (cumulés croissants notés ECC ou cumulés décroissants notés ECD)

**Fréquence :** En divisant l'effectif d'une valeur (classe ou modalité) par l'effectif total, on obtient sa fréquence. La fréquence, notée  $f$ , est toujours un nombre compris entre 0 et 1. La somme des toutes les fréquences est égale à 1. Très souvent, on exprime la fréquence en pourcentage.

### Exemple 1

Soit la liste suivante des étudiants suivis entre parenthèses d'une indication du nombre de livres lus dans l'année (A = peu, B = moyen, C = beaucoup, D = exceptionnel) :

Etudiant1 (C), Etudiant2 (C), Etudiant3 (A), Etudiant4 (B), Etudiant5 (A), Etudiant6(B), Etudiant7 (C), Etudiant8 (B), Etudiant9 (B), Etudiant10 (C), Etudiant11 (D), Etudiant12 (B), Etudiant13 (A), Etudiant14 (C), Etudiant15 (C), Etudiant16 (C), Etudiant17 (C), Etudiant18 (D), Etudiant19 (C), Etudiant20 (C).

La population étudiée est l'ensemble des étudiants de taille 20. Le caractère est le degré de lecture. C'est un caractère qualitatif qui prend ses valeurs dans l'ensemble: {A, B, C, D}.

### 2- Tableau statistique

Etant donnée une population P de  $n$  individus, sur laquelle est mesuré un caractère X. La valeur du caractère X prise par l'individu i est  $X(i)$ , que l'on note tout court  $X_i$ . Le tableau statistique de la série statistique  $X_1, X_2, X_3, \dots, X_n$  de cette population est un tableau à plusieurs colonnes comportant une partie ou toute les données relatives à la population telles que : l'effectif, la fréquence,...

### Exemple 2: Cas d'une variable statistique discrète

Considérons les 16 observations suivantes qui représentent le nombre de documents empruntés par les étudiants d'une bibliothèque pendant une période donnée. Les étudiants sont notés par  $E_1, E_2, \dots$

$E_1$	$E_2$	$E_3$	$E_4$	$E_5$	$E_6$	$E_7$	$E_8$	$E_9$	$E_{10}$	$E_{11}$	$E_{12}$	$E_{13}$	$E_{14}$	$E_{15}$	$E_{16}$
11	8	7	8	11	11	5	4	11	7	9	5	5	6	6	7

**Tableau 1**

Le tableau statistique correspondant à cette série statistique est le suivant :

Nombre d'emprunts de documents (Caractère)	Nombre d'Etudiant (L'effectif)
4	1
5	3
6	2
7	3
8	2
9	1
11	4
	N= 16

**Tableau 2**

- Dans cet exemple la variable est discrète et prend les 7 valeurs :  
 $X_1 = 4, X_2 = 5, X_3 = 6, X_4 = 7, X_5 = 8, X_6 = 9, X_7 = 11$
- A chacune de ces valeurs correspond un effectif :  
 $n_1 = 1, n_2 = 3, n_3 = 2, n_4 = 3, n_5 = 2, n_6 = 1, n_7 = 4$
- La série statistique étudiée est la série  $(X_i, n_i)$ , l'indice  $i$  prend les valeurs 1, 2, 3, 4, 5, 6 et 7.
- La somme des effectifs  $n_i$  est l'effectif total  $N$  de la population étudiée.
- $N = n_1 + n_2 + n_3 + n_4 + n_5 + n_6 + n_7$  que l'on note  $N = \sum_{i=1}^{i=7} n_i$  et qui se lit :  
« Somme de  $i = 1$  à  $i = 7$  des indices  $n_i$  »
- Si on divise l'effectif d'une valeur par l'effectif total, on obtient sa fréquence

Nombre d'emprunts de documents (Caractère)	Nombre d'Etudiant (L'effectif)	Fréquence	La fréquence (en %)
4	1	0.065	6.25
5	3	0.187	18.75
6	2	0.125	12.5
7	3	0.187	18.75
8	2	0.125	12.5
9	1	0.065	6.25
11	4	0.25	25
	N= 16		

**Tableau 3**

- On remarquera que la somme des fréquences est égale à 1 et la somme des pourcentages est égale à 100 %.
- A partir des effectifs on peut calculer les effectifs cumulés (cumulés croissants ou cumulés décroissants) permettant de répondre aux deux types de questions suivantes :
  1. Quel est le nombre d'étudiants ayant emprunté moins de 7 documents
  2. Quel est le nombre d'étudiants ayant emprunté au moins 7 documents (7 documents et plus)

Nombre d'emprunts de documents (Caractère)	Nombre d'Etudiant (L'effectif)	Effectifs cumulés croissants	Effectifs cumulés décroissants
4	1	1	16
5	3	4	15
6	2	6	12
7	3	9	10
8	2	11	7
9	1	12	5
11	4	16	4

**Tableau 4**

Détermination des effectifs cumulés croissants :

1<sup>er</sup> cumul croissant: 1

2<sup>eme</sup> cumul croissant:  $1 + 3 = 4$

3<sup>eme</sup> cumul croissant :  $4 + 2 = 6$  etc...

Détermination des effectifs cumulés décroissants :

1<sup>er</sup> cumul décroissant: 16

2<sup>eme</sup> cumul décroissant:  $16 - 1 = 15$

3<sup>eme</sup> cumul décroissant :  $15 - 3 = 12$  etc...

En utilisant le tableau on déduit que :

- Le nombre d'étudiants ayant emprunté moins de 7 documents est égal à 6 (la lecture à partir de la colonne des effectifs cumulés croissants)
- Le nombre d'étudiants ayant emprunté plus de 7 documents est égal à 10 (la lecture à partir de la colonne des effectifs cumulés décroissants)

**Exemple 3 : Cas d'une variable statistique continue**

Distribution statistique, à une certaine date, des adhérents d'une bibliothèque suivant leur âge.

Age	Effectifs ( $n_i$ )
[20 ; 25 [	9
[25 ; 30 [	27
[30 ; 35 [	36
[35 ; 40 [	45
[40 ; 45 [	18
[45 ; 50 [	9
[50 ; 55 [	3
[55 ; 60 [	3
	N = 150

*Tableau 5*

- Dans cet exemple, la variable est continue. Elle prendra un intervalle d'amplitude 5 ans
- A chacun tranche d'âge correspond un effectif :  
 $n_1 = 9, n_2 = 27, n_3 = 36, n_4 = 45, n_5 = 18, n_6 = 9, n_7 = 3, n_8 = 3$
- Les valeurs 22,5 ; 27,5 ; 32,5 ; 37,5 ; 42,5 ; 47,5 ; 52,5 ; 57,5 sont appelées les centres des intervalles.
- Les effectifs peuvent donner lieu à la présentation de colonnes d'effectifs cumulés (cumulés croissants ou cumulés décroissants) permettant de répondre aux deux types de questions suivantes :

1. Quel est le nombre d'adhérents ayant moins de 35 ans ?
2. Quel est le nombre d'adhérents ayant au moins 35 ans (35 ans et plus)

Age	Effectifs ( $n_i$ )	Fréquence f	f %	ECC	ECC (%)	ECD	ECD (%)
[ 20 ; 25 [	9	0,06	6%	9	6%	150	100%
[ 25 ; 30 [	27	0,18	18%	36	24%	141	94%
[30 ; 35 [	36	0,24	24%	72	48%	114	76%
[ 35 ; 40 [	45	0,30	30%	117	78%	78	52%
[ 40 ; 45 [	18	0,12	12%	135	90%	33	22%
[ 45 ; 50 [	9	0,06	6%	144	96%	15	10%
[ 50 ; 55 [	3	0,02	2%	147	98%	6	4%
[ 55 ; 60 [	3	0,02	2%	150	100%	3	2%
	N = 150	1	100%				

**Tableau 6**

Détermination des effectifs cumulés croissants :

1<sup>er</sup> cumul croissant: 9

2<sup>ème</sup> cumul croissant:  $9 + 27 = 36$

3<sup>ème</sup> cumul croissant :  $9 + 27 + 36 = 72$  etc...

Détermination des effectifs cumulés décroissants :

1<sup>er</sup> cumul décroissant: 150

2<sup>ème</sup> cumul décroissant:  $150 - 9 = 141$

3<sup>ème</sup> cumul décroissant :  $150 - 9 - 27 = 114$  etc...

En utilisant le tableau ci-dessus, on déduit que :

- Le nombre d'adhérents ayant moins de 35 ans est égal à 72 soit 48 % de la population.
- Le nombre de personnes ayant au moins 35 ans (i.e 35 ans et plus) est égal à 78, soit 52 % de la population.

### 3. Les différentes représentations graphiques

La représentation graphique de l'ensemble des valeurs numériques obtenues peut se faire sous la forme de polygones, d'histogrammes et de courbes. Elles permettent une visualisation des variations des variables étudiées.

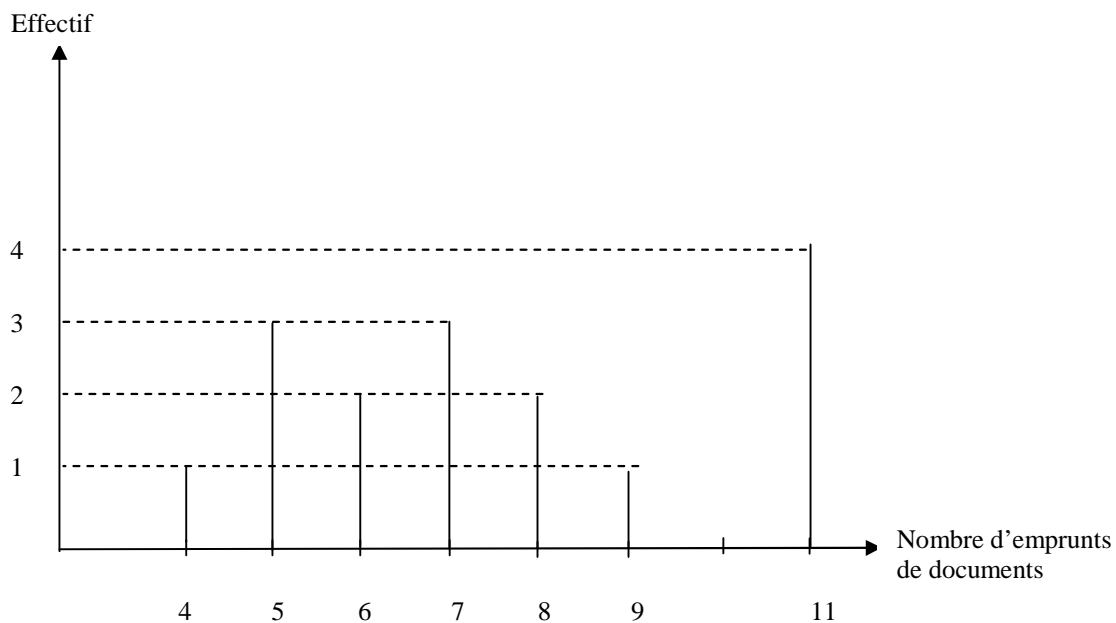
Les graphiques utilisés dépendent de la nature de la variable. Nous utiliserons, pour représenter les distributions d'effectifs (ou de fréquences), les diagrammes en bâtons, les histogrammes et le polygone des effectifs. Pour les distributions cumulées, nous utiliserons les polygones des effectifs (ou des fréquences) cumulés croissants et décroissants.

#### 3.1 Diagramme en bâtons

On appelle diagramme en bâtons un graphique qui à chaque modalité d'une variable quantitative discrète associe un segment (bâton) dont la hauteur est proportionnelle à l'effectif (ou à la fréquence).

**Exemple 4:** Le diagramme en bâtons ci-dessous représente les données de la série statistique de l'exemple 2.



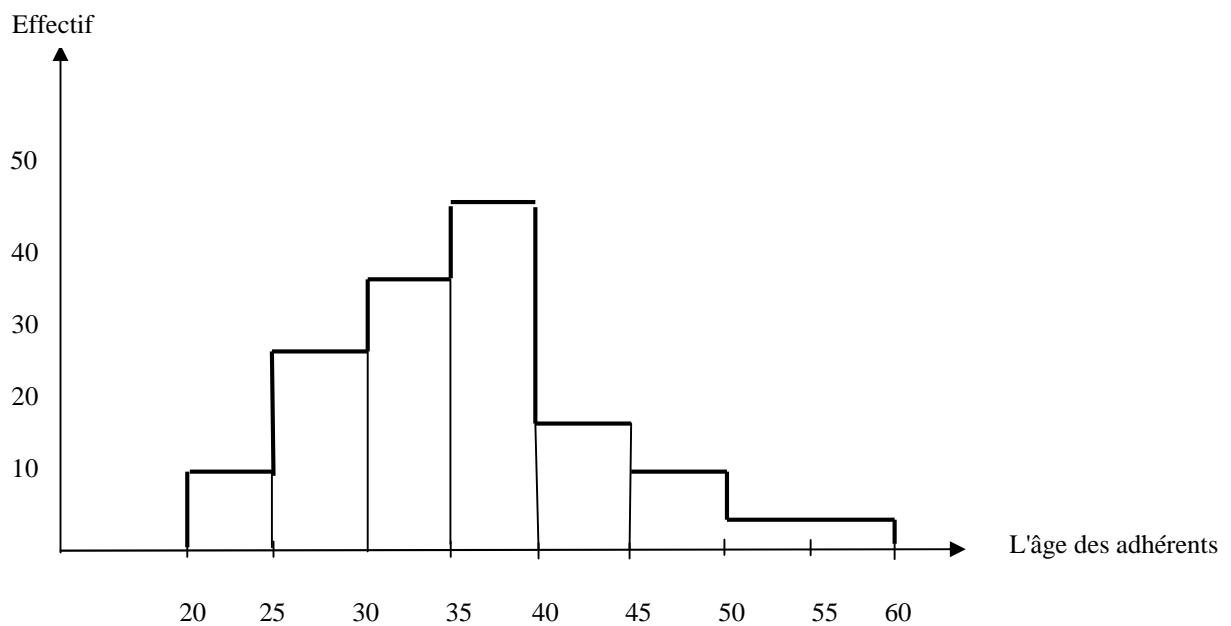


*Schéma 1*

### 3.2 Histogramme

Un histogramme est un diagramme composé de rectangles contigus dont les aires sont proportionnelles aux effectifs (ou aux fréquences) et dont les bases sont déterminées par les intervalles de classes. Il est utilisé pour représenter les données d'une série statistique à variable continue.

**Exemple 5:** l'histogramme ci-dessous représente les données de la série statistique de l'exemple 3



*Schéma 2*

### 3.3 Polygone des effectifs cumulés croissants

Il est utilisé pour les deux types de variables discrète et continue. Il montre quel est le pourcentage des cas qui tombent au-dessus ou au-dessous d'une valeur donnée.

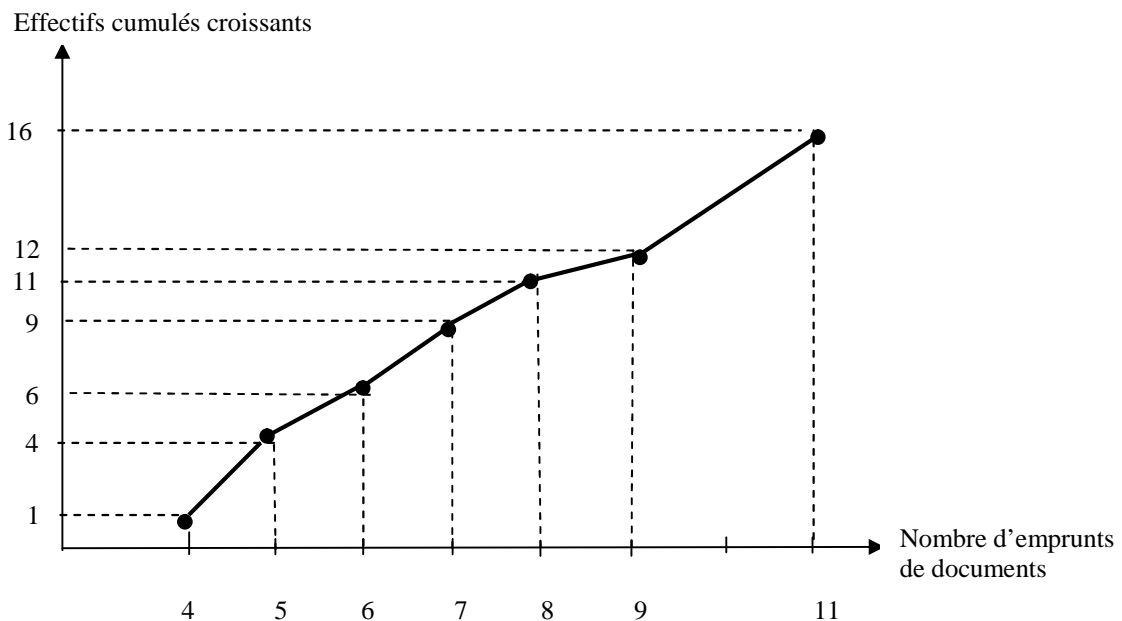
Dans le cas discret, on porte sur l'axe des abscisses les valeurs de la variable et sur l'axe des ordonnées les effectifs cumulés croissants. Ensuite on relie les points obtenus par des segments de droite.

Dans le cas continu, les classes sont représentées par leur limite supérieure. Le polygone est obtenu de la même façon que dans le cas discret i.e en reliant les points (limite supérieure de la classe, son effectif) par des segments de droite et en ajoutant un point fictif (limite inférieure de la première classe, 0).

#### Exemple:

##### Cas discret

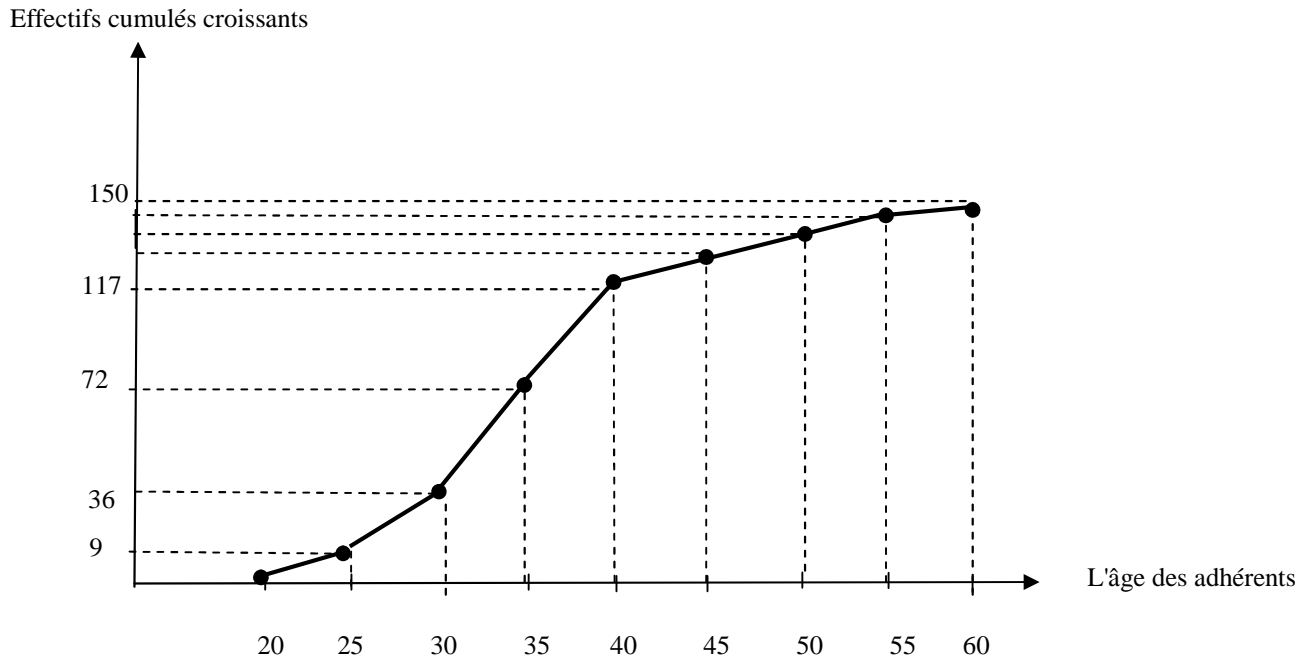
A partir des données de tableau 4, on obtient le polygone des effectifs cumulés croissants qui relie les points suivants: (4, 1), (5, 4), (6, 6), (7, 9), (8, 11), (9, 12), (11, 16)



*Schéma 3*

### Cas continu

A partir du tableau 6, le polygone des effectifs cumulés croissants reliant les points (0, 20), (25, 9), (30, 36), (35, 72), (40, 117), (45, 135), (50, 144), (55, 147), (60, 150) se présente comme suit:



*Schéma 4*

### Remarque:

On peut représenter la variation de la fréquence de la même façon que l'effectif. Il suffit de remplacer les valeurs des effectifs par celles des fréquences. Et à ce moment on peut tracer le diagramme en bâtons ou l'histogramme des fréquences, polygone des fréquences cumulées croissantes ou décroissantes.

## 4. Paramètres de position

Les paramètres de position (ou de tendance centrale) permettent de savoir autour de quelles valeurs se situent les valeurs d'une variable statistique. Ces paramètres sont le mode, la médiane et la moyenne.

### Le mode:

Le mode est la valeur distincte correspondant à l'effectif le plus élevé. Pour le cas d'une variable quantitative continue on parle d'une classe modale (i.e. classe correspondante à l'effectif le plus élevé).

## La médiane

La médiane, notée  $Me$ , est un paramètre de position, qui permet de couper la population étudiée en deux groupes contenant le même nombre d'individus.

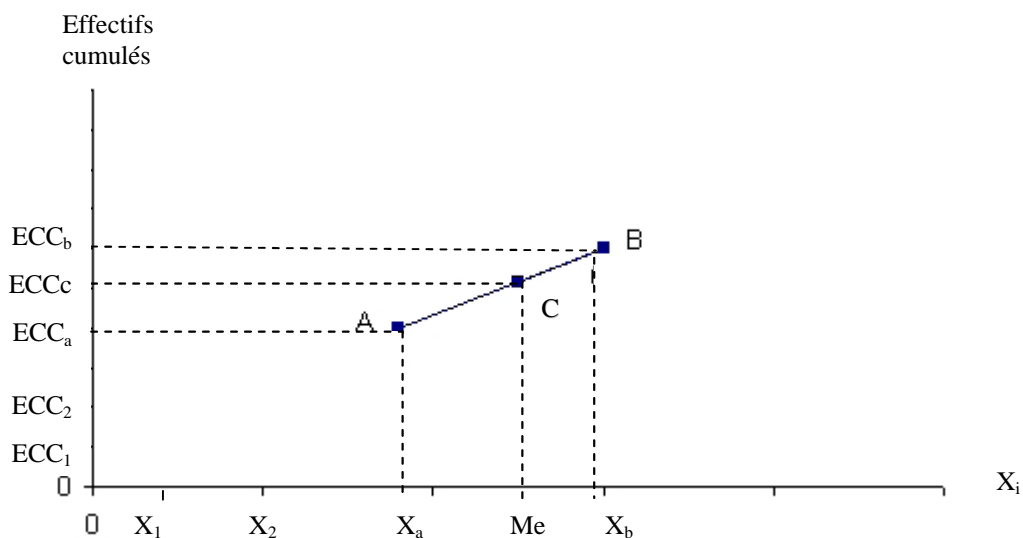
Ce paramètre est utile pour donner la répartition du caractère étudié, car 50 % environ de la population étudiée a une modalité inférieure à la médiane et 50 % une modalité supérieure à la médiane.

Pour le cas d'une variable quantitative discrète le calcul de la médiane se fait comme suit: soit  $n$  la taille de la population

1. Si  $n$  est impair, la médiane est la  $\left(\frac{n+1}{2}\right)^{ième}$  observation.
2. Si  $n$  est pair, la médiane est définie comme étant la moyenne de la  $\left(\frac{n}{2}\right)^{ième}$  observation et la  $\left(\frac{n}{2}+1\right)^{ième}$  observation.

Pour une variable quantitative continue:

1. On détermine la classe médiane  $[a, b[$  de même façon que pour une variable discrète en utilisant les effectifs cumulés.
2. La médiane appartient à l'intervalle  $[a, b[$  et elle est déterminée par interpolation linéaire comme suit:



Avec

$X_i$  : les extrémités supérieures des classes

$X_a$ : l'extrémité supérieure de la classe se trouvant avant la classe médiane

$X_b$  : l'extrémité supérieure de la classe médiane

$ECC_c = n/2$  (50% de l'effectif total)

$$\frac{Me - X_a}{ECC_c - ECC_a} = \frac{X_b - X_a}{ECC_b - ECC_a}$$

## Moyenne

La moyenne arithmétique, notée  $m$ , d'une série statistique est la somme des valeurs divisée par le nombre total des valeurs.

$$m = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{n_1 + n_2 + \dots + n_k} \quad \text{où}$$

$n_1 + n_2 + \dots + n_k = n = \text{effectif total}$

$$x_i = \begin{cases} \text{la valeur de la variable } X_i & \text{si la variable est discrète} \\ \text{le centre de la classe } [a_i, b_i[ = \frac{a_i + b_i}{2} & \text{si la variable est continue} \end{cases}$$

## Exemple

### Cas d'une variable quantitative discrète

Dans un centre de documentation, 20 périodiques, notés de A à T, ont été analysés durant un an. On a comptabilisé pour chacun d'entre eux le nombre d'articles pertinents. Les résultats obtenus sont représentés dans le tableau suivant :

$X_i$	$n_i$
2	2
3	2
4	3
5	2
6	2
7	3
8	3
9	3

1. L'effectif maximal de cette série est 3. Nous avons 4 valeurs ayant un effectif égal à 3 et donc cette série possède 4 modes : 4, 7, 8 et 9
2. La taille de la population est 20 (nombre pair) et donc la médiane est égale à la moyenne de la 10<sup>ème</sup> et 11<sup>ème</sup> observation =  $(6+6)/2 = 6$

N° de l'observation	Sa valeur	
1 <sup>ère</sup> observation	2	50% de la population
2 <sup>ème</sup> observation	2	
3 <sup>ème</sup> observation	3	
4 <sup>ème</sup> observation	3	
5 <sup>ème</sup> observation	4	
6 <sup>ème</sup> observation	4	
7 <sup>ème</sup> observation	4	
8 <sup>ème</sup> observation	5	
9 <sup>ème</sup> observation	5	50% de la population
10 <sup>ème</sup> observation	6	
11 <sup>ème</sup> observation	6	
12 <sup>ème</sup> observation	7	
13 <sup>ème</sup> observation	7	
14 <sup>ème</sup> observation	7	
15 <sup>ème</sup> observation	8	
16 <sup>ème</sup> observation	8	
17 <sup>ème</sup> observation	8	
18 <sup>ème</sup> observation	9	
19 <sup>ème</sup> observation	9	
20 <sup>ème</sup> observation	9	

3. La moyenne

$$m = (2 \times 2 + 2 \times 3 + 3 \times 4 + 2 \times 5 + 2 \times 6 + 3 \times 7 + 3 \times 8 + 3 \times 9) / 20 = 5.8$$

### Cas d'une variable quantitative continue

Soit la distribution des 200 employés d'une entreprise selon leur salaire mensuel :

salaire mensuel (10000 DA)	effectifs (n <sub>i</sub> )
[50,60[	20
[60,70[	60
[70,90[	50
[90,100[	40
[100,130[	30

1. [60, 70 [ est la classe modale car son effectif = 60 (la plus grande valeur des n<sub>i</sub> ).

2. La médiane

La taille de la population est 200 donc 50% de l'effectif total est 100.

La médiane correspond à l'effectif cumulé 100, d'après la colonne des ECC on a

- 80 employés ont moins de 70\*10000DA
- 130 employés ont moins de 90\*10000DA

Donc la médiane se trouve dans l'intervalle [70,90[. Par interpolation linéaire on obtient :

$(Me-70)/(100-80) = (90-70)/(130-80)$  où

- 100 représente les 50% de la taille de la population
- 80 est ECC qui vient juste avant ECC de la classe médiane (i.e ECC de la classe [60, 70]
- 130 est ECC de la classe médiane
- 70 et 90 sont les extrémités supérieures de la classe avant la classe médiane et la classe médiane.

$$(Me-70)/20 = 20/50 \quad \text{donc} \quad 50(Me-70)=400$$

$$(Me-70) = 40/5 \quad \text{donc} \quad Me = 8 + 70$$

$$Me = 78$$

Donc le salaire médian est d'environ 78\*10000DA

3. la moyenne

$$m = ((55*20) + (65*60) + (80*50) + (40*95) + (115*30)) / 200$$

$$m = 81.25$$

Donc le salaire moyen est 81.25\*10000DA

## 5. Paramètres de dispersion

Les paramètres de dispersion nous renseignent sur la dispersion des valeurs autour de la moyenne de la série statistique.

### L'étendue

L'étendue d'une série statistique est la différence entre la plus grande modalité du caractère et la plus petite modalité pour le cas discret (une modalité est la valeur que prend une variable  $X_i$ ). Pour calculer l'étendue pour une variable quantitative continue, il suffit de soustraire la borne supérieure de la dernière classe à la borne inférieure de la 1<sup>ère</sup> classe.

### Variance

La variance est un indicateur de la dispersion d'une série par rapport à sa moyenne. La définition de la variance d'une série statistique est donnée par la formule :

$$V = \frac{n_1(x_1-m)^2 + n_2(x_2-m)^2 + \dots + n_k(x_k-m)^2}{n_1 + n_2 + \dots + n_k}$$

Où:

$n_i$  est l'effectif de la variable  $X_i$

$n_1 + n_2 + \dots + n_k = N$  = effectif total

$m$  est la moyenne de la série statistique

### L'écart-type

La définition de l'écart-type d'une série est donnée par la formule :  $\sigma = \sqrt{V}$ .

Si l'écart-type est faible, cela signifie que les valeurs sont assez concentrées autour de la moyenne et si l'écart-type est élevé, cela veut dire au contraire que les valeurs sont plus dispersées autour de la moyenne.

**Exemple:** on reprend les données des deux exemples ci-dessus.

#### Cas discret

1. L'étendue de la série statistique à variable discrète:  $9-2 = 7$
2.  $V = (2 * (2 - 5.8)^2 + 2 * (3 - 5.8)^2 + 3 * (4 - 5.8)^2 + 2 * (5 - 5.8)^2 + 2 * (6 - 5.8)^2 + 3 * (7 - 5.8)^2 + 3 * (8 - 5.8)^2 + 3 * (9 - 5.8)^2) / 20 = 5.26$
3.  $\sigma = \sqrt{V} = 2.29$



### Cas continu

1. L'étendue de la série statistique à variable continue:  $130 \cdot 10^4 - 50 \cdot 10^4 = 80 \cdot 10^4$
2.  $V = (20 \cdot (55 - 81.25)^2 + 60 \cdot (65 - 81.25)^2 + 50 \cdot (80 - 81.25)^2 + 40 \cdot (95 - 81.25)^2 + 30 \cdot (115 - 81.25)^2) / 200 = 357.185$
3.  $= \sqrt{V} = 18.89$

### Activité 1:

**Exercice 1** Donnez la nature des variables suivantes (quantitative ou qualitative, discrète ou continue)

- Année de naissance
- La couleur d'un objet
- Nombre des étudiants
- Profession
- Niveau de scolarité : primaire, secondaire, universitaire
- La distance entre les villes
- Age
- Appréciation : Jamais, rarement, à l'occasion, fréquemment, toujours

**Exercice 2.** Les 17 ingénieurs d'un laboratoire, notés de A à Q, ont effectué respectivement, pendant une semaine, les nombres suivants de connexions à leur messagerie électronique.

Ingénieurs	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Connexions	6	9	10	14	16	17	14	16	14	10	9	14	14	16	9	17	16

1. Déterminer la population étudiée ainsi que sa taille.
2. Quel est le caractère
3. Construisez le tableau représentatif de cette distribution contenant le caractère, l'effectif, l'effectif cumulé croissant et la fréquence
4. Tracer la courbe des effectifs ainsi que celle des fréquences

**Exercice 3 :** Les gestionnaires d'une bibliothèque décident de mener une étude statistique afin de connaître le temps que passe un adhérent à la salle de lecture. Les données collectées sont présentées dans le tableau suivant:

Le temps en minutes	effectif
[0, 20[	8
[20, 40[	5
[40, 60[	10
[60, 80[	3
[80, 100[	5
[100, 120[	9
[120, 140[	4
[140, 160[	6
[160, 180[	10

1. Quelle est la population,
2. Quelle est la variable étudiée,
3. Quel est le type de la variable : quantitative ou qualitative ? Continue ou discrète
4. Construisez le tableau représentatif de cette distribution.
5. Tracer la courbe des effectifs et des effectifs cumulés croissants.

## Activité 2

### Exercice 1.

Calculer la médiane, la moyenne, l'écart type et la variance à partir des données de l'exercice2/activité 1

### Exercice 2:

A partir des données de l'exercice3 / activité1

1. Calculer le temps médian passé à la salle de lecture.
2. Calculer le temps moyen.
3. Calculer la variance et l'écart type

## Correction des activités:

### Activité 1:

#### Exercice1

Variable	nature
Année de naissance	Quantitative, discrète
La couleur d'un objet	Qualitative
Nombre des étudiants	Quantitative, discrète
Profession	Qualitative
Niveau de scolarité : primaire, secondaire, universitaire	Qualitative
La distance entre les villes	Quantitative, continue
Age	Quantitative, continue
Appréciation : Jamais, rarement, à l'occasion, fréquemment, toujours	Qualitative

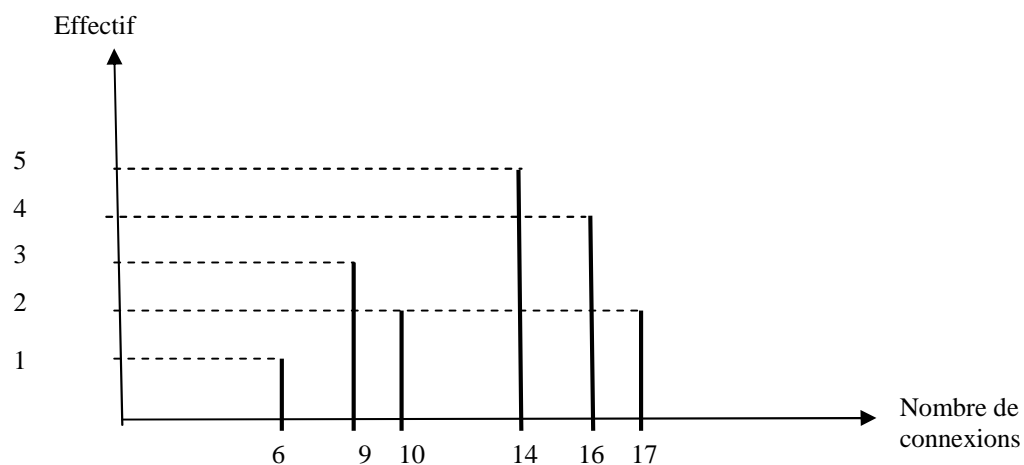
#### Exercice 2

- La population étudiée est l'ensemble des ingénieurs d'un laboratoire. Sa taille est égale à 17
- Le caractère est le nombre de connexions à la messagerie électronique

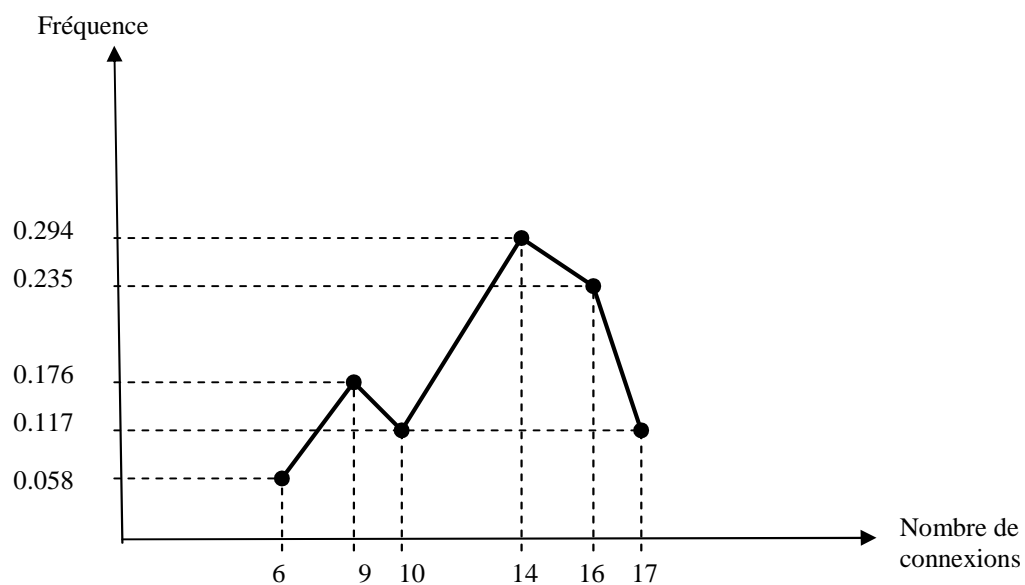
Tableau statistique

Caractère	effectif	ECC	fréquence
6	1	1	0.058
9	3	4	0.176
10	2	6	0.117
14	5	11	0.294
16	4	15	0.235
17	2	17	0.117

### La courbe des effectifs



### La courbe des fréquences



### Exercice 3:

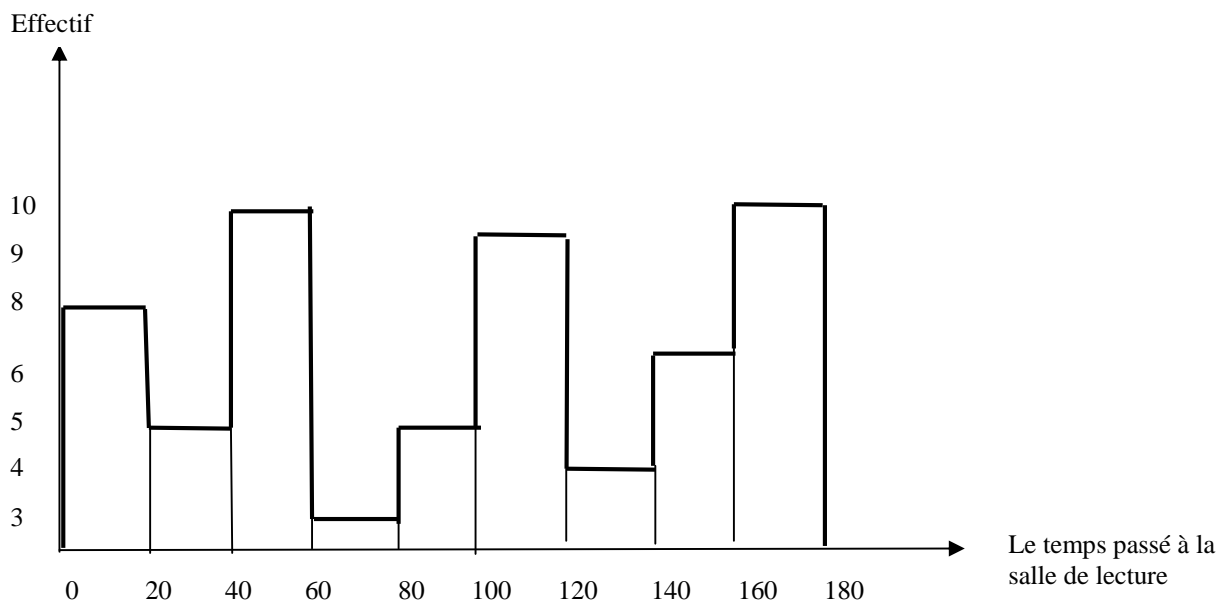
1. La population est l'ensemble des adhérents d'une bibliothèque
2. La variable étudiée est le temps que passe un adhérent dans une salle de lecture
3. La variable est quantitative continue
4. Le tableau représentatif de cette distribution.

Le temps en minutes	effectif	ECC	ECD	Fréquence
[0, 20[	8	8	60	0.133
[20, 40[	5	13	52	0.083
[40, 60[	10	23	47	0.166

[60, 80[	3	26	37	0.05
[80, 100[	5	31	34	0.083
[100, 120[	9	40	29	0.15
[120, 140[	4	44	20	0.066
[140, 160[	6	50	16	0.1
[160, 180[	10	60	10	0.166

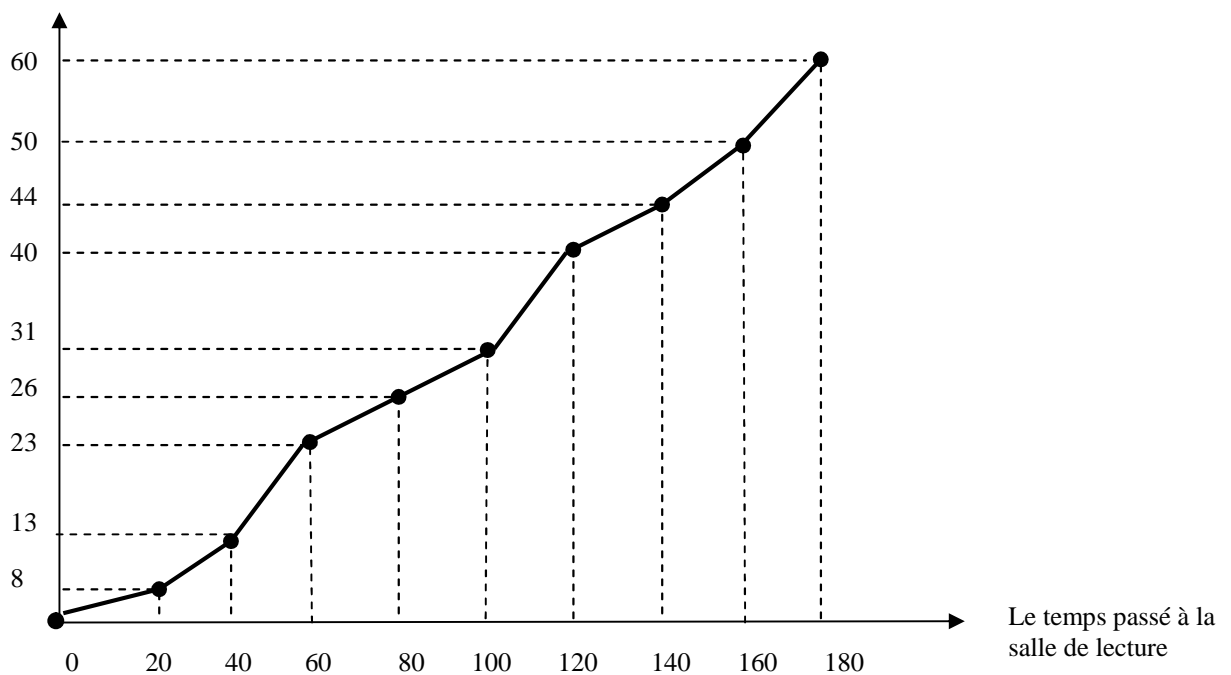
5. Tracer la courbe des effectifs et des effectifs cumulés croissants.

- La courbe des effectifs



- La courbe des effectifs cumulés croissants

Effectifs cumulés croissants



## Activité 2

### Exercice 1:

La taille de la population est 17 et donc la médiane correspond à la valeur de la 9<sup>ème</sup> observation. Ainsi  $M = 14$ .

- Calcul de la moyenne

$$m = \frac{n_1x_1 + n_2x_2 + \dots + n_6x_6}{n_1 + n_2 + \dots + n_6}$$

$$m = \frac{1 \times 6 + 3 \times 9 + 2 \times 10 + 5 \times 14 + 4 \times 16 + 2 \times 17}{1 + 3 + 2 + 5 + 4 + 2} = 13$$

- Calcul de la variance

$$V = \frac{n_1(x_1 - m)^2 + n_2(x_2 - m)^2 + \dots + n_6(x_6 - m)^2}{n_1 + n_2 + \dots + n_6}$$

$$V = \frac{1 \times (6 - 13)^2 + 3 \times (9 - 13)^2 + 2 \times (10 - 13)^2 + 5 \times (14 - 13)^2 + 4 \times (16 - 13)^2 + 2 \times (17 - 13)^2}{1 + 3 + 2 + 5 + 4 + 2} = 8.94$$

- Calcul de l'écart type

$$\sigma = \sqrt{V} = 2.98$$

### Exercice 2

- Calcul du temps médian passé à la salle de lecture

La taille de la population est 60 donc 50% de l'effectif total est 30.

La médiane correspond à l'effectif cumulé 30, d'après la colonne des ECC on a

- 26 adhérents passent moins de 80 minutes à la salle de lecture
- 31 adhérents passent moins de 100 minutes à la salle de lecture

Donc la médiane se trouve dans l'intervalle  $[80, 100[$ . Par interpolation linéaire on obtient :

$$(Me - 80) / (31 - 30) = (100 - 80) / (31 - 26) \quad \text{où}$$

- 30 représente les 50% de la taille de la population
- 26 est ECC qui vient juste avant ECC de la classe médiane (i.e ECC de la classe  $[60, 80]$ )
- 31 est ECC de la classe médiane
- 80 et 100 sont les extrémités supérieures de la classe avant la classe médiane et la classe médiane.

$$(Me - 80) / 1 = 20 / 5 \quad \text{donc} \quad Me - 80 = 4$$

Ainsi  $Me = 84$  minutes.

- Calcul du temps moyen

$$m = \frac{8 \times 10 + 5 \times 30 + 10 \times 50 + 3 \times 70 + 5 \times 90 + 9 \times 110 + 4 \times 130 + 6 \times 150 + 10 \times 170}{8 + 5 + 10 + 3 + 5 + 9 + 4 + 6 + 10} = 91.66$$

- Calcul de la variance

V =

$$\frac{8 \times (10 - m)^2 + 5 \times (30 - m)^2 + 10 \times (50 - m)^2 + 3 \times (70 - m)^2 + 5 \times (90 - m)^2 + 9 \times (110 - m)^2 + 4 \times (130 - m)^2 + 6 \times (150 - m)^2 + 10 \times (170 - m)^2}{8 + 5 + 10 + 3 + 5 + 9 + 4 + 6 + 10}$$

V = 3030.55 et ainsi  $\sigma = \sqrt{V} = 55.05$ .

## **Chapitre III**

### **Le Modèle de Morse**



## 1 - Introduction

Le modèle de Morse est une méthode de micro-évaluation qui s'inscrit dans le cadre de l'approche systémique. Cette approche mesure la qualité globale d'un système en analysant la contribution de chacun de ses sous systèmes à l'efficacité de l'organisation entière<sup>7</sup>.

La méthode de Morse mesure l'activité de la partie la plus importante du sous système de diffusion, à savoir le prêt d'ouvrages à l'extérieur; ce prêt concerne uniquement les monographies. Cette méthode donne, en outre des indications sur l'ensemble de la circulation des ouvrages par catégorie étudiée et permet d'évaluer l'efficacité de la politique d'acquisition.

Les principales applications du modèle sont la gestion des duplications et la mise en réserve des ouvrages inactifs.

## 2- Définitions

Nous désignons par circulation ou fréquence de circulation le nombre d'usages (prêts, photocopies, consultations sur place) pendant une période donnée d'un document ou d'un corpus de documents.

## 3 - Formulation du modèle

Le modèle de Morse tient compte à la fois de l'**obsolescence** qui affecte les collections et des phénomènes **de regain de popularité** qui peuvent toucher certains livres. Il permet de prévoir la circulation moyenne des volumes d'une catégorie, en fonction de leur circulation l'année précédente, en appliquant l'équation suivante :

$$R(t+1) = \alpha + \beta R(t)$$

$R(t)$  : Circulation moyenne des volumes (ou taux de rotation) durant l'année écoulée.

$R(t+1)$  : Circulation moyenne prévisible pour la prochaine année,

---

<sup>7</sup> Cane, Simon, « TRAFIC », *BBF*, 1987, n° 1, p. 26-40  
[en ligne] <<http://bbf.enssib.fr/>> Consulté le 18 octobre 2011

$\alpha$ : Paramètre indiquant la circulation moyenne des plus anciens volumes.

$\beta$ : Paramètre indiquant l'obsolescence des volumes d'année en année.

#### **4- Mise en œuvre de la formulation de Morse**

La mise en œuvre de cette équation suppose que l'on distingue le comportement des nouveautés, les volumes acquis depuis moins de deux ans, et celui des volumes disponibles depuis plus de deux ans, les ouvrages « anciens ».

En ce qui concerne les nouveautés, on effectue les mesures suivantes :

- Nombre et pourcentage de volumes n'ayant pas circulé pendant l'année écoulée, les volumes « inactifs » :  $P_0(t_N)$

- Nombre et pourcentage des volumes ayant été empruntés au moins une fois l'année écoulée, les « ouvrages actifs » :  $C(t_N) = 100 \% - P_0(t_N)$

- Circulation moyenne des volumes « actifs » pendant l'année écoulée :  $R_a(t_N)$  et circulation moyenne de l'ensemble des nouveautés pendant l'année écoulée :  $R(t_N)$

Ces données constituent des indicateurs de l'efficacité de la politique d'acquisition dans chaque domaine étudié.

En ce qui concerne les volumes anciens, on effectue les mêmes mesures afin de connaître le pourcentage des volumes actifs et le taux de rotation pour l'ensemble des volumes de chaque catégorie et de pouvoir comparer ces indicateurs à ceux des nouveautés.

Les données sur l'ensemble des volumes permettront d'effectuer des prévisions sur le comportement de la collection étudiée.

##### **4.1 - Détermination des paramètres $\alpha$ et $\beta$**

Le paramètre  $\alpha$  indique la circulation moyenne des plus anciens ouvrages; le paramètre  $\beta$  mesure l'obsolescence des ouvrages. Le paramètre  $\beta$  doit être inférieur à l'unité et il varie ordinairement entre 0,2 et 0,8 contrairement au paramètre  $\alpha$  qui peut être plus grand que l'unité, mais il est souvent compris entre les valeurs 0,3 et 0,7.

Le calcul des  $R_0$  et  $R_1$  s'effectue uniquement à partir des mesures du comportement des anciens volumes. Il se fait en déterminant d'abord les paramètres suivants:

$R_0(t_A)$  : Circulation moyenne des volumes sortis 0 fois en l'an (t -1)

$R_1(t_A)$  : Circulation moyenne des volumes sortis 1 fois en l'an (t -1)

$R_2(t_A)$  : Circulation moyenne des volumes sortis 2 fois en l'an (t -1).

$R_3(t_A)$  : Circulation moyenne des volumes sortis 3 fois en l'an (t -1)

$R_4(t_A)$  : Circulation moyenne des volumes sortis 4 fois en l'an (t -1)

Et ainsi:

$$= R_0(t_A);$$

$$= \frac{R_1(t_A) + R_2(t_A) + R_3(t_A) + R_4(t_A) - 4R_0(t_A)}{10}$$

## 5 - Applications du modèle de Morse

### 5.1 - Gestion des duplications

Afin de définir une meilleure politique d'acquisition, le modèle de Morse permet de connaître le nombre de prêts générés pendant les dix prochaines années par l'achat d'un volume supplémentaire d'un ouvrage prêté  $R(1)$  fois l'année écoulée en calculant  $R_d$  comme suit:

$$R_d = \frac{1}{\mu} \left[ \left( \frac{\partial}{1-\beta} \right)^2 \left( 10 - \frac{1+\beta}{1-\beta} \right) + \left( \frac{\partial}{1-\beta} + \beta R(1) \right)^2 \left( \frac{1}{1-\beta^2} \right) \right]$$

Où

$R(1)$  : nombre de prêts du volume dont on envisage d'acheter un exemplaire supplémentaire.

$\frac{1}{\mu}$  : Fraction moyenne de l'année pendant laquelle un volume d'une catégorie donnée est

absent des rayons à chaque emprunt. Par exemple, si les volumes d'informatique sont restitués

en moyenne après trois semaines  $\frac{1}{\mu} = \frac{3}{52} = 0.057$  (52 est le nombre de semaines

d'une année).

## 5.2 - Mise en réserve des ouvrages inactifs :

Morse préconise la mise en réserve des ouvrages qui n'ont pas circulé pendant l'année écoulée. Son équation permet d'étudier ce qu'il adviendra alors du nombre de prêts. Il semble en effet que les paramètres  $\alpha$  et  $\beta$  soient à peu près semblables pour la partie active de la collection et pour sa partie inactive.

Pour étudier ce qu'il adviendrait de la collection laissée en l'état, on calcule la circulation moyenne de ses volumes en utilisant la formule suivante:

$$R(t+1) = \alpha + \beta R(t).$$

On calcule ensuite le nombre de prêts qui seront effectués en multipliant  $R(t+1)$  par le nombre de volumes de la collection. Puis on analyse le comportement des volumes laissés en rayon si les inactifs de l'année précédente sont mis en réserve. Cela revient à effectuer l'opération sur les seuls volumes actifs :

$$R'(t+1) = \alpha + \beta R_a(t).$$

On calcule le nombre de prêts en multipliant  $R'(t+1)$  par le nombre de volumes actifs l'année écoulée.

Il reste à calculer le nombre de prêts effectués par les volumes inactifs l'année précédente. Si ces volumes étaient restés en rayon, ce nombre aurait été égal au nombre de volumes inactifs multiplié par  $\alpha$ .

Des recherches ont montré qu'en reléguant une collection en réserve, on réduit son  $\alpha$  au deux tiers de sa valeur initiale. Les volumes mis en réserve généreront donc des prêts en nombre égal à leur nombre total multiplié par  $\frac{2}{3}\alpha$ . On peut ainsi comparer le nombre de prêts qui seraient effectués, si la collection restait en l'état, au nombre total de prêts générés par les volumes restés en rayon et, si la collection restait en l'état, au nombre total de prêts générés par les volumes restés en rayons et par les volumes mis en réserve si l'on décidait de choisir la solution envisagée par Morse.

### Exemple:

I. Une bibliothèque possède 600 ouvrages disponibles depuis plus de deux ans. Le nombre de prêts de ces ouvrages est donné dans le tableau suivant:

Nombre de sorties	Nombre de volumes	Nombre de prêts
-------------------	-------------------	-----------------

Volumes prêtés 0 fois en 2010	<b>186</b>	
Volumes prêtés 1 fois en 2010	<b>113</b>	
Volumes prêtés 2 fois en 2010	<b>105</b>	
Volumes prêtés 3 fois en 2010	<b>86</b>	
Volumes prêtés 4 fois en 2010	<b>80</b>	
Volumes prêtés 5 fois en 2010	<b>30</b>	

- Remplir la colonne relative aux nombres de prêts.

- Calculer  $P_0(10_A)$  (Pourcentage de volumes inactifs parmi les anciens ouvrages durant l'année 2010)

- Calculer  $C(10_A)$  (Pourcentage de volumes actifs parmi les anciens ouvrages durant l'année 2010)

- Calculer  $R_a(10_A)$  (Circulation moyenne des volumes actifs parmi les anciens ouvrages durant l'année 2010)

- Calculer  $R(10_A)$  (Circulation moyenne des anciens ouvrages durant l'année 2010)

II. Dans une deuxième étape, on a collecté les données relatives aux nombres de prêts des 50 nouveaux ouvrages.

Nombre de sorties	Nombre de volumes	Nombre de prêts
Volumes prêtés 0 fois en 2010	<b>05</b>	
Volumes prêtés 1 fois en 2010	<b>09</b>	
Volumes prêtés 2 fois en 2010	<b>08</b>	
Volumes prêtés 3 fois en 2010	<b>06</b>	
Volumes prêtés 4 fois en 2010	<b>07</b>	
Volumes prêtés 5 fois en 2010	<b>06</b>	
Volumes prêtés 6 fois en 2010	<b>05</b>	
Volumes prêtés 7 fois en 2010	<b>04</b>	

- Remplir la colonne relative aux nombres de prêts.

- Calculer  $P_0(10_N)$  (Pourcentage de volumes inactifs parmi les nouveaux ouvrages durant l'année 2010)

- Calculer  $C(10_N)$  (Pourcentage de volumes actifs parmi les nouveaux ouvrages durant l'année 2010)

- Calculer  $R_a(10_N)$  (Circulation moyenne des volumes actifs parmi les nouveaux ouvrages durant l'année 2010)
- Calculer  $R(10_N)$  (Circulation moyenne des nouveaux ouvrages durant l'année 2010)

III. En fusionnant les anciens et les nouveaux ouvrages, complétez le tableau suivant:

Nombre total de volumes	
Nombre total de volumes actifs	
Nombre total de volumes inactifs	
Nombre total de prêts	

- Calculer  $P_0(10)$  (Pourcentage de volumes inactifs durant l'année 2010)
- Calculer  $C(10)$  (Pourcentage de volumes actifs durant l'année 2010)
- Calculer  $R_a(10)$  (Circulation moyenne des volumes actifs durant l'année 2010)
- Calculer  $R(10)$  (Circulation moyenne de l'ensemble des volumes durant l'année 2010)

IV. Pour calculer  $R_0(10_A)$  et  $R_1(10_A)$ , on s'est intéressé aux ouvrages anciens qui sont sortis respectivement 0, 1, 2, 3 et 4 fois en 2009, et pour chaque volume on a noté le nombre de prêts réalisés en 2010.

Nombre de fois de sortie des volumes en 2009	Nombre de volumes	Nombre de prêts réalisés en 2010
0 fois	<b>199</b>	<b>65</b>
1 fois	<b>90</b>	<b>72</b>
2 fois	<b>141</b>	<b>239</b>
3 fois	<b>72</b>	<b>187</b>
4 fois	<b>63</b>	<b>208</b>

- Calculer les circulations moyennes suivantes :

$R_0(10_A)$  : Circulation moyenne en 2010 des volumes sortis 0 fois en 2009.

$R_1(10_A)$  : Circulation moyenne en 2010 des volumes sortis 1 fois en 2009.

$R_2(10_A)$  : Circulation moyenne en 2010 des volumes sortis 2 fois en 2009.

$R_3(10_A)$  : Circulation moyenne en 2010 des volumes sortis 3 fois en 2009.

$R_4(10_A)$  : Circulation moyenne en 2010 des volumes sortis 4 fois en 2009.

- Calculer les paramètres et

On rappelle que :

$$= R_0(10_A) \text{ et } = \frac{R_1(t_A) + R_2(t_A) + R_3(t_A) + R_4(t_A) - 4R_0(t_A)}{10}$$

V. Calcul de l'évolution des prêts au cas où la collection étudiée resterait inchangée

- Calculer la circulation moyenne prévisible pour l'année 2011 des ouvrages de l'échantillon étudié (on utilisera l'équation  $R(11) = + .R(10)$  )

- Calculer le nombre de prêts que devrait réaliser en 2011 l'échantillon étudié

**Corrigé:**

Nombre de sorties	Nombre de volumes	Nombre de prêts
Volumes prêtés 0 fois en 2010	<b>186</b>	$186*0 = 0$
Volumes prêtés 1 fois en 2010	<b>113</b>	$113*1 = 113$
Volumes prêtés 2 fois en 2010	<b>105</b>	$105*2 = 210$
Volumes prêtés 3 fois en 2010	<b>86</b>	$86*3 = 258$
Volumes prêtés 4 fois en 2010	<b>80</b>	$80*4 = 320$
Volumes prêtés 5 fois en 2010	<b>30</b>	$30*5 = 150$
	<b>600</b>	1051

2. Calculer  $P_0(10_A)$  (pourcentage des volumes inactifs parmi les anciens ouvrages durant l'année 2010)

$$P_0(10_A) = \frac{186}{600} \times 100 = 31\%$$

3. Calculer  $C(10_A)$  (pourcentage des volumes actifs parmi les anciens ouvrages durant l'année 2010)

$$C(10_A) = 100\% - 31\% = 69\%$$

4. Calculer  $R_a(10_A)$  (circulation moyenne des volumes actifs parmi les anciens ouvrages durant l'année 2010)

Il y'a  $600 - 186 = 414$  volumes actifs qui ont réalisé 1051 prêts durant l'année 2010

$$R_a(10_A) = \frac{\text{Nb.re.de.prêts.réalisés.par.les.volumes.actifs}}{\text{Nb.re.de.volumes.actifs}} = \frac{1051}{414} = 2.53$$

5. Calculer  $R(10_A)$  (circulation moyenne des anciens ouvrages durant l'année 2010)

$$R(10_A) = \frac{\text{Nb.re.de.prêts.réalisés.par.les.anciens.ouvrages}}{\text{Nb.re.total.d'ouvrages}} = \frac{1051}{600} = 1.75$$

## II. Ouvrages nouveaux (disponibles depuis moins de deux ans) :

Nombre de volumes étudiés : 50

1 – Remplir la colonne relative aux nombres de prêts

Nombre de sorties	Nombre de volumes	Nombre de prêts
Volumes prêtés 0 fois en 2010	<b>05</b>	<b>0</b>
Volumes prêtés 1 fois en 2010	<b>09</b>	<b>9</b>
Volumes prêtés 2 fois en 2010	<b>08</b>	<b>16</b>
Volumes prêtés 3 fois en 2010	<b>06</b>	<b>18</b>
Volumes prêtés 4 fois en 2010	<b>07</b>	<b>28</b>
Volumes prêtés 5 fois en 2010	<b>06</b>	<b>30</b>
Volumes prêtés 6 fois en 2010	<b>05</b>	<b>30</b>
Volumes prêtés 7 fois en 2010	<b>04</b>	<b>28</b>
	<b>50</b>	<b>159</b>

2. Calculer  $P_0(10_N)$  (pourcentage des volumes inactifs parmi les nouveaux ouvrages durant l'année 2010)

$$P_0(10_N) = \frac{5}{50} \times 100 = 10\%$$



3. Calculer  $C(10_N)$  (pourcentage des volumes actifs parmi les nouveaux ouvrages durant l'année 2010)

$$C(10_N) = 100 - 10 = 90\%$$

4. Calculer  $R_a(10_N)$  (circulation moyenne des volumes actifs parmi les nouveaux ouvrages durant l'année 2010)

Il y'a  $50 - 5 = 45$  volumes actifs qui ont réalisé 159 prêts durant l'année 2010

$$R_a(10_N) = \frac{\text{Nb.re.de.prêts.réalisés.par.les.volumes.actifs}}{\text{Nb.re.de.volumes.actifs}} = \frac{159}{45} = 3.53$$

5. Calculer  $R(10_N)$  (circulation moyenne des nouveaux ouvrages durant l'année 2010)

$$R(10_N) = \frac{\text{Nb.re.de.prêts.réalisés.par.les.nouveaux.ouvrages}}{\text{Nb.re.total.d'ouvrages}} = \frac{159}{50} = 3.18$$

### III. Comportement d'ensemble de la classe étudiée

1. Remplir la 2<sup>ème</sup> colonne

Nombre total de volumes	$600+50 = 650$
Nombre total de volumes actifs	$414+45 = 459$
Nombre total de volumes inactifs	$186+5 = 191$
Nombre total de prêts	$1051+159 = 1210$

2. Calculer  $P_0(10)$  (pourcentage des volumes inactifs durant l'année 2010)

$$P_0(10) = \frac{191}{650} \times 100 = 29\%$$

3. Calculer  $C(10)$  (pourcentage des volumes actifs durant l'année 2010)

$$C(10_N) = 100 - 29 = 71\%$$

4. Calculer  $R_a(10)$  (circulation moyenne des volumes actifs durant l'année 2010)

$$R_a(10) = \frac{\text{Nb.re.de.pr\^ets.r\^ealis\^es.par.les.volumes.actifs}}{\text{Nb.re.de.volumes.actifs}} = \frac{1210}{459} = 2.63$$

5. Calculer  $R(10)$  (circulation moyenne de l'ensemble des volumes durant l'année 2010)

$$R(10) = \frac{\text{Nb.re.de.pr\^ets.r\^ealis\^es.par.l'ensemble.des.ouvrages}}{\text{Nb.re.total.d'ouvrages}} = \frac{1210}{650} = 1.86$$

#### IV. Calcul des paramètres $\alpha$ et $\beta$ :

$\alpha$  : indicateur de la circulation moyenne des ouvrages anciens

$\beta$  : indicateur de l'obsolescence des ouvrages

Pour calculer  $\alpha$  et  $\beta$ , On s'est intéressé aux ouvrages anciens qui sont sortis respectivement 0, 1, 2, 3 et 4 fois en 2009, et pour chaque volume on a noté le nombre de prêts réalisés en 2010.

Nombre de fois de sortie des volumes en 2009	Nbre de volumes	Nbre de prêts réalisés en 2010
0 fois	<b>199</b>	<b>65</b>
1 fois	<b>90</b>	<b>72</b>
2 fois	<b>141</b>	<b>239</b>
3 fois	<b>72</b>	<b>187</b>
4 fois	<b>63</b>	<b>208</b>

- calculer les circulations moyennes suivantes :  $R_0(10_A)$ ,  $R_1(10_A)$ ,  $R_2(10_A)$ ,  $R_3(10_A)$ ,  $R_4(10_A)$

$R_0(10_A)$  : Circulation moyenne en 2010 des volumes sortis 0 fois en 2009.

$$R_0(10_A) = \frac{65}{199} = 0.32$$

$R_1(10_A)$  : Circulation moyenne en 2010 des volumes sortis 1 fois en 2009.

$$R_1(10_A) = \frac{72}{90} = 0.80$$

$R_2(10_A)$  : Circulation moyenne en 2010 des volumes sortis 2 fois en 2009.

$$R_2(10_A) = \frac{239}{141} = 1.69$$

$R_3(10_A)$  : Circulation moyenne en 2010 des volumes sortis 3 fois en 2009.

$$R_3(10_A) = \frac{187}{72} = 2.59 \quad R_4(10_A) : \text{Circulation moyenne en 2010 des volumes sortis 4 fois en 2009.}$$

$$R_4(10_A) = \frac{208}{63} = 3.30$$

2. Calcul des paramètres  $\alpha$  et  $\beta$

On rappelle que

$$\alpha = R_0(10_A)$$

$$\beta = \frac{R_1(10_A) + R_2(10_A) + R_3(10_A) + R_4(10_A) - 4R_0(10_A)}{10}$$

$$\alpha = 0.32$$

$$\beta = \frac{0.80 + 1.69 + 2.59 + 3.30 - 4 \times 0.32}{10} = 0.72$$

## V. Prévision de l'évolution des prêts :

1. Calcul de l'évolution des prêts au cas où la collection étudiée resterait inchangée

Calculer la circulation moyenne prévisible pour l'année 2011 des ouvrages de l'échantillon étudié (on utilisera l'équation  $R(11) = \alpha + \beta \cdot R(10)$ )

En appliquant la formule de Morse  $R(11) = \alpha + \beta \cdot R(10)$

$$R(11) = 0.32 + 0.72 \cdot 1.86 = 1.65$$

2. Calculer le nombre de prêts que devrait réaliser en 2011 l'échantillon étudié

Le nombre de prêts que devrait réaliser en 2011 l'échantillon étudié est :

$$N = 650 * 1.65 \cong 1073 \text{ prêts}$$

Soit une diminution de  $1210 - 1073 = 137$  prêts par rapport à l'année 2009

$$\text{C'est-à-dire une diminution de } \frac{137}{1073} \times 100 = 12.76\%$$

### **Conclusion :**

C'est une diminution qui reste quand même assez problématique, le gestionnaire de la bibliothèque devra à cet effet procéder à des correctifs sur l'un, au moins, des trois axes suivants:

- 1- la politique d'acquisition
- 2- la gestion des duplications
- 3- la mise en réserve des ouvrages inactifs

### **Activité1**

Afin de les prévisions en terme de prêt de l'année 2010, une bibliothèque décide de mener une étude statistique en se basant sur les données collectées pendant l'année en cours.

Nombre de sorties des ouvrages en 2010	Nombre de volumes	Nombre de prêts réalisés en 2011
0 fois	20	6
1 fois	50	40
2 fois	60	30
3 fois	80	80
4 fois	100	120

1. Calculer les paramètres  $\alpha$  et  $\beta$
2. Calculer la circulation moyenne prévisible pour l'année 2012.
3. Calculer le nombre de prêts que devrait réaliser en 2012 l'échantillon étudié.
4. Que peut-on conclure?

### Corrigé

1. Avant de calculer  $\alpha$  et  $\beta$ , on doit d'abord calculer les  $R_i(11)$  correspondant à la circulation moyenne en 2011 des volumes sortis  $i$  fois en 2010 avec  $i = 0, 1, 2, 3$  et  $4$

$$R_0(11) = 6/20 = 0.3$$

$$R_1(11) = 40/50 = 0.8$$

$$R_2(11) = 30/60 = 0.5$$

$$R_3(11) = 80/80 = 1$$

$$R_4(11) = 120/100 = 1.2$$

Ainsi

$$\alpha = R_0(11) = 0.3 \text{ et}$$

$$\beta = \frac{0.80 + 0.5 + 1 + 1.2 - 4 \times 0.3}{10} = 0.23$$

2. En appliquant la formule de Morse  $R(12) = \alpha + \beta.R(11)$ . On doit calculer  $R(11)$  (la circulation moyenne des ouvrages pendant l'année 2011)

$R(11)$  = le nombre total des prêts/ le nombre total des ouvrages

$$R(11) = (6+40+30+80+120)/(20+50+60+80+100) = 276/310 = 0.89$$

$$\text{Ainsi, } R(12) = 0.3 + 0.23 \times 0.89 = 0.5$$

3. le nombre de prêt que devrait réaliser le même échantillon en 2012 est égal à:

$R(12)$  = le nombre total des prêts en 2012/ le nombre total des ouvrages et donc

$$\text{Le nombre total des prêts en 2012} = R(12) \times \text{le nombre total des ouvrages} = 0.5 \times 310 = 155$$

4. Le nombre de prêt des 310 ouvrages va baisser d'une façon considérable en passant de 276 à 155. Donc la bibliothèque est sous l'obligation d'enrichir son fonds documentaire par d'autres titres.

## Références bibliographiques

1. Jean-marie De Ketele, Xavier Roegiers. Méthodologie du recueil d'information. Les éditions De Boeck et les Editions de l'Université de Bruxelles, 1993.
2. Claude Javeau. L'enquête par questionnaire-Editions de l'université de Bruxelles, Belgique, 1982
3. Evaluer des actions de communication. COMM collection, octobre 2006.
4. Thierry Lafouge, Yves-François Le Coadic, Christine Michel. Eléments de Statistique et de mathématique de l'information. Cours avec exemples et exercices corrigés. Presses de l'enssib, 2002.
5. E. BRESSOUD, J. KAHANE. Statistique descriptive avec Excel et la calculatrice. Synthex - Synthèse de cours et exercices corrigés, 1ère édition 2008.
6. A. HAMON, N. JEGOU. Statistiques descriptives : cours et exercices corrigés. Pratique de la statistique, 1ère édition 2008

## **Annexe**

### **Epreuve de Recueil et Traitement Statistique des Données Année 2011**

#### **Exercice 1**

Pour chaque question, une seule réponse est correcte

L'interview est une méthode de collecte

- ☐ Directe à sens unique
- ☐ Indirecte à sens unique
- ☐ Directe à double sens
- ☐ Indirecte à double sens

L'étude de document est une méthode

- ☐ Quantitative
- ☐ Qualitative
- ☐ Quantitative et qualitative

Le mode d'une série statistique est un caractère de

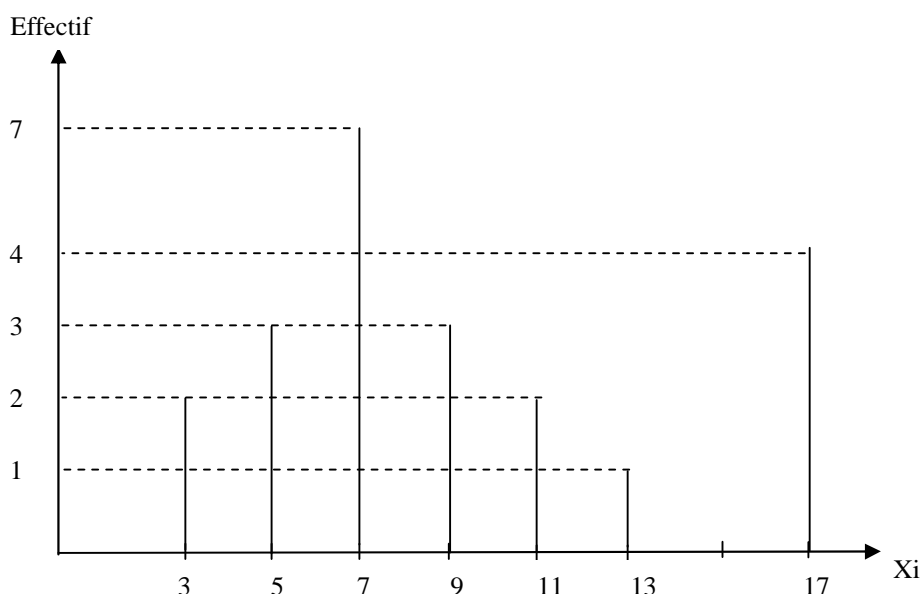
- ☐ Position
- ☐ Dispersion

L'étendu d'une série statistique est :

- ☐ L'écart entre la valeur minimale et la valeur maximale du caractère
- ☐ La plus grande valeur que peut prendre un caractère
- ☐ La moyenne des modalités

## Exercice 2

A partir du diagramme suivant:



1. Déterminer la nature de la variable: discrète ou continue.
2. Construire le tableau représentatif de cette distribution contenant le caractère, l'effectif, l'effectif cumulé décroissant et la fréquence.
3. Calculer la médiane.

## Exercice 3

Un centre de documentation vient d'acquérir 50 livres. On a établi une répartition de ces livres d'après leurs prix.

Prix d'un livre ( 100 DA )	Nombre de livres
[ 5 ; 10 [	10
[ 10 ; 15 [	20
[ 15 ; 20 [	7
[ 20 ; 25 [	8
[ 25 ; 30 [	5

1. Déterminer le prix moyen
2. Calculer la variance et l'écart type.

## Exercice 4

Une bibliothèque a collecté un ensemble de données relatives à la circulation de 400 ouvrages. Ces données ont permis de calculer les paramètres suivants :

$$R_0(11) = 0.75 \quad R_1(11) = 1.8 \quad R_2(11) = 2.5 \quad R_3(11) = 3.6 \quad R_4(11) = 3.5$$

Où  $R_i(11)$  est la circulation moyenne des volumes sortis  $i$  fois en l'an 2011 ( $i = 0, 1, 2, 3$  et 4).

Pour l'année 2012, la bibliothèque voudrait doubler le nombre de prêts pour ce même échantillon. En appliquant le modèle de Morse, est ce que son objectif sera atteint sachant que ces 400 ouvrages ont enregistré 1500 prêts durant l'année 2011?



## Corrigé

### Exercice 1

Pour chaque question, une seule réponse est correcte

L'interview est une méthode de collecte

- ☐ Directe à sens unique
- ☐ Indirecte à sens unique
- ☒ Directe à double sens
- ☐ Indirecte à double sens

L'étude de document est une méthode

- ☐ Quantitative
- ☒ Qualitative
- ☐ Quantitative et qualitative

Le mode d'une série statistique est un caractère de

- ☒ Position
- ☐ Dispersion

L'étendu d'une série statistique est :

- ☒ L'écart entre la valeur minimale et la valeur maximale du caractère
- ☐ La plus grande valeur que peut prendre un caractère
- ☐ La moyenne des modalités

### Exercice 2:

1. La variable étudiée est une variable quantitative discrète car elle est représentée par un diagramme en bâtons.
2. Le tableau représentatif de cette distribution contenant le caractère, l'effectif, l'effectif cumulé décroissant et la fréquence.

caractère	effectif	ECD	fréquence
3	2	22	$2/22 = 0.09$
5	3	20	$3/22 = 0.13$
7	7	17	$7/22 = 0.31$
9	3	10	$3/22 = 0.13$
11	2	7	$2/22 = 0.09$
13	1	5	$1/22 = 0.04$

17	4	4	4/22 = 0.18
----	---	---	-------------

3. Calculer la médiane.

Etant donné que la taille de population est paire (n =22), la médiane est la moyenne de la 11<sup>ème</sup> et 12<sup>ème</sup> observation.

$$Me = (7+7)/2 = 7$$

### Exercice 3:

Prix d'un livre ( 100 DA )	Nombre de livres	Centre des classes
[5 ; 10 [	10	7.5
[10 ; 15 [	20	12.5
[15 ; 20 [	7	17.5
[20 ; 25 [	8	22.5
[25 ; 30 [	5	27.5

1. Le prix moyen

$$m = \frac{10 \times 7.5 + 20 \times 12.5 + 7 \times 17.5 + 8 \times 22.5 + 5 \times 27.5}{10 + 20 + 7 + 8 + 5} = 15.3$$

Le prix moyen est égal à 15.3 x 100 DA

2. Calcul de la variance

$$V = \frac{10 \times (7.5 - 15.3)^2 + 20 \times (12.5 - 15.3)^2 + 7 \times (17.5 - 15.3)^2 + 8 \times (22.5 - 15.3)^2 + 5 \times (27.5 - 15.3)^2}{10 + 20 + 7 + 8 + 5}$$

$$V = 39.16$$

Et donc la variance de cette série statistique est 39.16 x 100 DA

$$\sigma = \sqrt{V} = 6.25$$

L'écart type est égal à 6.25 x 100 DA.

### Exercice 4

Pour pouvoir répondre à la question, on doit d'abord calculer la circulation moyenne des années 2011 et 2012

1. La circulation moyenne de l'année 2011

$$R(11) = \text{le nombre de prêts en 2011} / \text{le nombre total des ouvrages} = 1500/400 = 3.75$$

2. La circulation moyenne prévue pour l'année 2011

$$R(12) = \alpha + \beta * R(11)$$

Sachant que

$$\alpha = R_0(11) = 0.75$$

$$\beta = (R_1(11) + R_2(11) + R_3(11) + R_4(11) - 4 * R_0(11)) / 10 = 0.75$$

$$\beta = (1.8 + 2.5 + 3.6 + 3.5 - 4 * 0.75) / 10 = 0.84$$

Et donc

$$R(12) = 0.75 + 0.84 * 3.75 = 3.9$$

Et puisque  $R(12)$  = nombre de prêts de l'année 2012/ le nombre total des ouvrages alors

$$\text{Le nombre de prêts de l'année 2012} = R(12) * \text{le nombre total des ouvrages} = 3.9 * 400 = 1560$$

La bibliothèque pourra augmenter le nombre de prêts durant l'année 2012 mais il sera pas doublé.