

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

جامعة 8 ماي 1945 قالمة

Université 8 Mai 1945 Guelma

Faculté des Sciences de la Nature et de la Vie, Sciences de la terre et de l'Univers



## Mémoire En Vue de l'Obtention du Diplôme de Master

Domaine: Sciences de la Nature et de la Vie  
Spécialité/Option: Biologie moléculaire des procaryotes  
Département: Biologie

### Thème

## Logiciels informatique en biologie

Présenté par : Maizi Marwa

Mehallel Souhaila

Messaadia Fatima El Zahra

Devant la commission composée de :

Khallef M	Présidente	Université de Guelma
Athamnia M	Encadreur	Université de Guelma
Khenaka K	Examinatrice	Université de Guelma
Grara N	Membre	Université de Guelma
Drif F	Membre	Université de Guelma
Benbelkacem S	Membre	Université de Guelma

Juin 2017

## *Remerciement*

*Ce travail a été effectué à l'université de Guelma Algérie*

*Il est le fruit de tant d'années d'étude. Il nous donne l'occasion de  
remercier tous les enseignants de notre département*

*De Biologie ayant participé d'une manière ou d'une autre à notre  
formation.*

*Nos remerciements vont particulièrement à notre encadreur  
Mr. ATHAMNIA Mohammed (M.A.B) qu'a pu aménager son temps  
pour nous guider dans la réalisation de ce projet.*

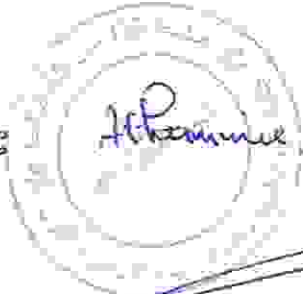
*Nous remercions également Mme KHALLEF M (M.C. B)  
à l'Université de 8 Mai 1945, Guelma ; pour avoir accepté  
d'évaluer ce travail et de présider le jury*

*Nous remercierons également Mme KHANAKA K (M.A.A) de nous  
avoir fait l'honneur de juger et évaluer ce travail en qualité  
d'examinatrice*

*Ces mêmes remerciements vont également à l'endroit des parents,  
amis et camarades étudiants qui n'ont cessé de nous soutenir dans  
toutes nos études. Que la paix du seigneur soit avec eux.*

*Encadré par :*

*Athamnia Mohammed*



## Dédicace

*Au nom du dieu clément et miséricordieux et que le salut de dieu*

*Soit sur son prophète mohamed*

*Je dédie ce modeste travail;*

*Aux deux être le plus chers au monde, qui ont souffert nuit et jour*

*pour nous couvrir de leur amour, mes parents.*

*A mon père abed el hak pour sa patience avec moi et son*

*encouragement ;*

*A ma source de bonheur, la prunelle de mes yeux, ma mère Warda.*

*Que le bon dieu vous garde en bonne santé;*

*A mon frère: Hamza.*

*A mes sœurs : Hadjer, Narjes*

*A celles qui je passe avec elles les bonnes heures et je porte avec*

*Elles que les bonnes souvenirs, à mes amies intimes :*

*Malika, aicha, imen.*

*En fin, mes remerciements vont à toutes nos amies, et en*

*Particulièrement celle de notre promotion de BMP (2017).*

*Marwa*

## Dédicace

*A mes chers parents source de tendresse, de volonté et de patience.  
Mes Yeux à travers lesquels j'ai vu et je vois ce monde : Abed el aziz  
et Hemama.*

*Je vous remercie d'être toujours à mes côtés, de me soutenir, m'aimer,  
me protéger et pour tous ce que vous avez fait pour moi.*

*A mon frère: Youcef.*

*A mes sœurs :Soltana, Wafa, Meriem*

*A mes proches : Sana, Mouna, Fatima.*

*A celles qui je passe avec elles les bonnes heures et je porte avec*

*Elles que les bonnes souvenirs, à mes amies intimes :*

*Nawel, Luiza, Hana.*

*En fin, mes remerciements vont à toutes nos amies, et en  
Particulièrement celle de notre promotion de BMP (2017).*

*Fatima.Z*

## Dédicace

*Je dédie se travaille à:*

*Mes chers parent: Akila et Lemnaour que nulle dédicace ne puisse  
exprimer mes sincers sentiments pour leurs passiences illimitées leurs  
encouragement contenu leur aide en témoignage de mon profond  
amour et respect pour leur grand sacrifices*

*A mon frère: Salah*

*A mes sœurs : Soumia, Samia, Amina, chahra zad  
Pour leur grand amour et leur soutien qu'ils trouvent ici  
L'expression de ma haute gratitude*

*En fin, mes remerciements vont à tous nos professeurs*

*Souhaila*

Produced with

## Abstract

Our work has interested about the use of the computing software applied in biology, among these software one in based on the statistical and cartographic software, they allow to know the analysis statistics and the realization of maps in some domains.

The statistics is very useful in all the disciplinary fields, it is taught in all the university sectors, from the economy to the biology.

The geographical information system handled(treated) in our study can be used by the students to make modern techniques of remote detection (MAPINFO 8 software, applied to the satellite images and the aerial photos) represent the distribution of the current real vegetation on the ground according to their ecological affinities and consecutive to the fires(lights) of degradation.

We also studied the contributions of the biocomputing in particular by the various databases and the biocomputing tools which she allowed to create these last years.

**Keywords:** statistics, cartography, bionformatics, geographical information system.

Produced with Scantopdf

## Résumé

Notre travail a porté sur l'utilisation des logiciels informatiques appliqués en biologie, parmi ces logiciels on a basé sur les logiciels statistiques et cartographiques, elles permettent de connaître l'analyse statistique et la réalisation des cartes géographiques dans quelques domaines.

La statistique est très utile dans tous les champs disciplinaires, elle est enseignée dans toutes les filières universitaires, de l'économie à la biologie.

Le système d'informations géographiques traitées dans notre étude peut être employé par les étudiants pour faire des techniques modernes de télédétection (logiciel Mapinfo 8.... Appliqué sur les images satellitaires et les photos aériennes) représente la répartition de la végétation réelle actuelle sur le terrain en fonction de leurs affinités écologiques et consécutives aux feux de dégradation.

On a étudié aussi les apports de la bioinformatique notamment par les différentes bases de données et outils bioinformatiques qu'elle a permis de créer ces dernières années.

**Mots clés :** statistiques, cartographie, bioinformatique, système d'information géographique.

Produced with ScanTopdf

## المخلص

عملنا يتعلّق باستخدام برنامج الحوسبة في علم الاحياء، ومن بين هذه البرامج ركزنا على برامج الاحصاءات ورسم الخرائط، هذه البرامج تسمح لنا معرفة التحليل الاحصائي ورسم الخرائط في بعض المجالات.

ان الاحصائيات مفيدة جدا في جميع المجالات التعليمية لجميع القطاعات، من الاقتصاد الى البيولوجيا.

نظم المعلومات الجغرافية المدروسة يمكن استخدامها من قبل الطلاب على التقنيات الحديثة للكشف عن بعد (من خلال موقع 8 MAPINFO الذي يطبق على صور الأقمار الصناعية، الصور الجوية) و التي تمثل توزيع النباتات الحقيقية الحالية على الارض حسب الانتماءات الايكولوجية على التوالي.

كما اننا درسنا مساهمات البيومعلوماتية وخاصة قواعد البيانات المختلفة وادوات البيومعلوماتية التي انشأت في السنوات الاخيرة.

الكلمات المفتاحية: الاحصاءات، رسم الخرائط، الحوسبة الحيوية، نظام المعلومات الجغرافية.



## Sommaire

Titre		Page
Résumé		
Listes des figures.		
Liste de tableaux.		
Liste des abréviations.		
Introduction		1
<b>CHAPITRE I : Informatique et statistiques appliqués à la biologie</b>		
1	Objectifs du chapitre	3
2	Les statistiques	3
2.1	Bref historique	3
2.2	Définition de la statistique	3
2.3	Définition de la statistique	4
3	Démarche générale en statistique	4
3.1	L'identification du problème	4
3.2	Le recueil des données	5
3.3	L'analyse et l'interprétation des données	5
4	Notions de base et terminologie	6
4.1	Ensemble / Population / Echantillon / Elément / Individu	6
4.2	Caractère / Modalité / Variable	7
4.2.1.	Le caractère	7
4.2.2	Modalité / Mesure	8
4.2.3	Tableau élémentaire	8
4.3	Les différents types de variable statistique	9
4.3.1	Les variables qualitatives	9
4.3.1.1	Les variables qualitatives nominales	9
4.3.1.2	Les variables qualitatives ordinales	9
4.3.2	Les variables quantitatives	9
4.3.2.1	Les variables quantitatives discrètes	9
4.3.2.2	Les variables quantitatives continues	9
5	Représentation des données	10
5.1	Tableaux statistiques	10
5.2	Représentation graphiques et statistique descriptive	10
5.2.1	Statistique univariée	10
5.2.1.1	Graphiques de dispersion d'un vecteur	11
5.2.1.2	Graphiques de dispersion de deux vecteurs	11
5.2.1.3	Boîtes à moustaches - Box Plot	13
5.2.1.4	La réduction des données à une dimension	14
5.2.2	Statistique bivariée	14
5.2.2.1	Nuages de points	14
5.2.2.2	La réduction des données à deux	15
5.2.3	Statistique multivariée	16
5.2.3.1	L'Analyse en Composantes Principales (ACP : exemple)	16
5.2.3.2	La réduction des données à deux dimensions	17
6	Analyse inferentielle des résultats	17

6.1	Un peu de théorie	17
6.1.1	Lois de probabilité discontinues généralités	17
6.1.2	Lois de probabilité continues – généralités	18
7	Test d'hypothèse	18
7.1	Hypothèse statistique	18
7.2	Test d'hypothèse	18
7.3	Principe d'un test d'hypothèses	18
7.4	Les différents types de test statistique	18
8	liste de quelque logiciel statistique	19
9	Model des logiciels	19
9.1	Logiciel SPSS	19
9.1.1	Définition	19
9.1.2	Les fonctions	19
9.1.3	Raisons de l'utilisation de ce logiciel	20
9.1.4	Les différentes fenêtres de SPSS	20
9.1.4.1	Les trois premiers types de fenêtres	20
9.1.5	Ouvrir un fichier de données avec SPSS	23
9.2	Logiciel R	25
9.2.1	Définition	25
9.2.2	Pourquoi utiliser R	25
9.2.3	R et les statistiques	26
9.2.4	R et les graphiques	26
9.2.5	L'interface R	27
10	Quelques statisticiens	27
<b>Chapitre II : Le système d'information géographique appliqué en biologie</b>		
1	Objectif	29
2	Bref historique : Les premiers SIG	29
3	Définition d'un Système d'Information Géographique (SIG)	29
4	Eléments et fonctions d'un SIG	31
5	Structure d'un SIG	31
6	Fonctionnalités de base des SIG	32
6.1	La saisie de données	32
6.2	La Transformation	32
6.3	Stockage de données	32
6.4	L'analyse spatiale	33
6.4.1	Fonctions de mesure	33
6.4.2	La requête spatiale	33
6.4.3	Fonctions de superposition	33
6.4.4	Fonctions de proximité	33
6.4.5	Fonctions de réseau	34
6.4.5	Classification	34
6.5	Visualisation	34

<b>5.3</b>	L'industrie agro-alimentaire	<b>56</b>
<b>5.4</b>	L'écologie	<b>56</b>
<b>5.5</b>	Les biopuces	<b>56</b>
<b>5.6</b>	La guerre bactériologique ou chimique	<b>57</b>
<b>5.7</b>	Les molécules darwiniennes	<b>57</b>
<b>5.8</b>	Le domaine anténatal	<b>58</b>
<b>Conclusion</b>		<b>59</b>
<b>Références bibliographiques</b>		<b>60</b>

Produced with ScanTOPDF

6.5.1	Définition de la cartographie	34
7	Les domaines d'application des S.I.G	36
8	L'intérêt des S.I.G	36
9	Liste de principaux logiciels de (SIG)	37
10	Usages du SIG en écologie	37
10.1	Les outils utilisés pour cartographier les stations forestières	38
11	logiciel Mapinfo	40
11.1	Définition	40
11.2	Environnement de MapInfo	41
11.2.1	La barre d'outils Général	41
11.2.2	La structure des données au format MapInfo	41
11.2.3	Le contrôle des tables (ou des couches)	43
11.2.4	Interrogation de la base à partir de la fenêtre carte	44
11.2.5	Calculs simples à partir de la fenêtre carte	45
11.2.6	Application en biologie (Model)	46
<b>Chapitre III : La bioinformatique</b>		
1	Définition	49
2	Les banques de données	50
2.1	Les différents types de base de données	50
2.1.1	Les banques généralistes	50
2.1.1.1	Séquences nucléiques	50
2.1.1.2	Séquences protéiques	51
2.1.2	Les banques spécialisées	51
2.2	Diffusion et utilisation des banques de données	52
2.2.1	Diffusion	52
2.2.2	Interrogation des banques de données	52
3	Les méthodes utilisées par la bioinformatique	53
3.1	Méthode comparative	53
3.2	Méthode statistique	53
3.3	Approche par modélisation	53
4	Les domaines de la bioinformatique	53
4.1	La génomique	53
4.2	La protéomique	54
4.3	Les autres « omes » et « omiques »	55
5	Les applications de la bioinformatique	55
5.1	La santé	55
5.2	L'industrie pharmaceutique	56

## Liste des abréviations

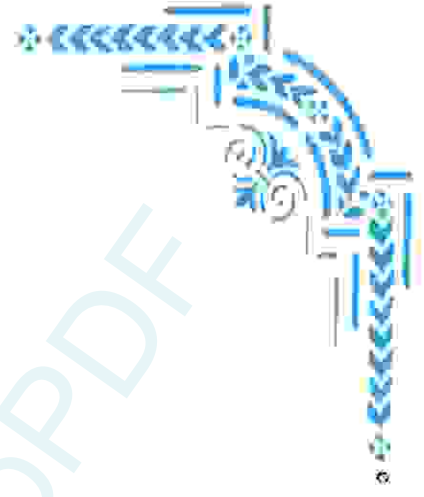
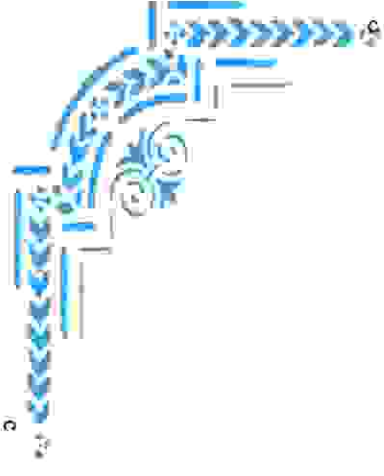
- % : pourcent
- °C: Degrés Celsius
- 2 D: deux Dimensions
- 3D: Trois Dimensions
- ACM : L'Analyse en Composantes Multiples
- ACP : L'Analyse en Composantes Principales
- ADN : Acide Désoxyribonucléique
- AFC : L'Analyse Factorielle des Correspondances
- EBI : European Bioinformatics Institute
- EMBO : European Molecular Biology Organization
- ExPASy : Expert Protein Analysis System
- GDB: Genome Data Base
- GLM : Modèle Linéaire Généralisé
- IBM: Inertational Business Machines
- IMD : Information Matrix Database
- IUT : Instituts Universitaires de Technologie
- JIPID : Japan International Protein Information Database
- KEGG: Kyoto Encyclopedia of Genes and Genomes
- MGD: Mouse Genome Database
- MGED: Microarray Gene Expression Data
- MIA : Molecular Information Agent
- MIM : Mendelian Inheritance in Man
- MIPS : Martinsried Institute for Protein Sequences, Munich, Allemagne
- NBRF : National Biomedical Research Foundation
- NCBI: National Center for Biotechnology Information
- NIG : National Institute of Genetics
- PDB : Protein Data Bank
- SIG : système d'information géographique
- SPSS: Statistical Package for Social Sciences
- STID : STatistique et Informatique Décisionnelle
- TFD : Transcription Factor Database
- YPD: Yeast Proteome Database

30	Méthode de dessin de la limite de la zone d'étude de la Mahouna	44
31	Localisation de la zone d'étude dans la carte topographique de Guelma (feuille nj-32-II-5 Est) (Echelle 1/50 000 ème)	45
32	La bioinformatique multidisciplinaire	48
33	Illustration du résultat de séquençage	54
34	Electrophorèse en 2D pour l'analyse protéique	54

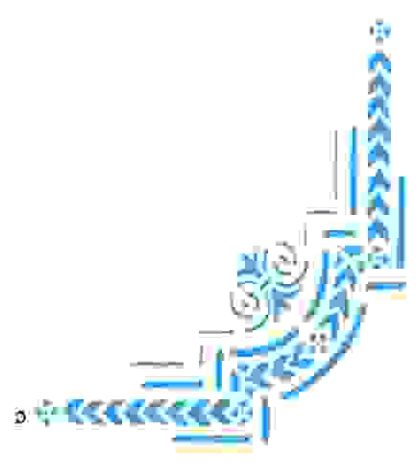
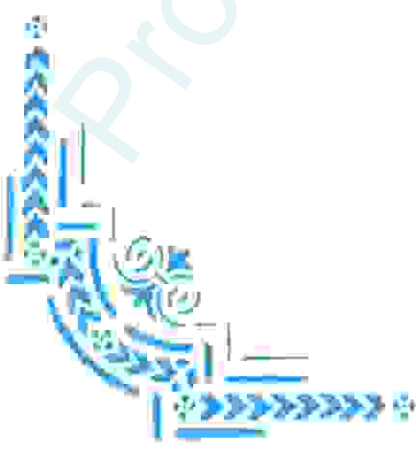
Produced with ScanTOPDF

### *Liste des figures*

<b>Figure</b>	<b>Titre</b>	<b>Page</b>
01	Exemple de présentation des données	9
02	Illustration de graphiques de dispersion	11
03	Illustration des histogrammes	11
04	Illustration des Box-plots	12
05	Illustration des nuages de points	13
06	Illustration de droite de régression	14
07	Illustration d'un ACP	15
08	La fenêtre des données	19
09	La fenêtre des commandes	20
10	La fenêtre de résultats d'un SPSS	20
11	Fenêtre d'ouvrir un fichier de données avec SPSS.	22
12	Fenêtre d'affichage de données avec SPSS.	22
13	fenêtre d'affichage des variables avec SPSS.	22
14	Quelques possibilités graphiques offertes par R	26
15	Interface graphique R	26
16	Modes de représentation de l'information géographique dans un SIG	27
17	Les disciplines contribuant au S.I.G	27
18	Structure d'un SIG	29
19	Procédé de superposition de couches	31
20	Répartition spatiale de deux espèces d'odonate dans le bassin versant de guelma	36
21	Présentation du logo du logiciel MapInfo	38
22	La barre d'outils Général de mapinfo	39
23	La barre d'outils Dessin de mapinfo	39
24	Menu Fichier ouvrir	41
25	L'option contrôle de couches	41
26	La boîte de dialogue Infos	41
27	L'outil somme de MapInfo.	42
28	L'outil statistique de MapInfo	43
29	Calage de la carte topographique par le MapInfo 8.0	44



# *Introduction*





**Introduction :**

L'informatique est l'art, la technique ou la science qui consiste à manipuler des informations à l'aide d'un outil, l'ordinateur. L'informatique a pour objet de définir des algorithmes qui permettent de modifier la vision que l'on a d'un problème, ou d'extraire d'une grande quantité d'informations mal structurées, de nouvelles connaissances plus utiles [17].

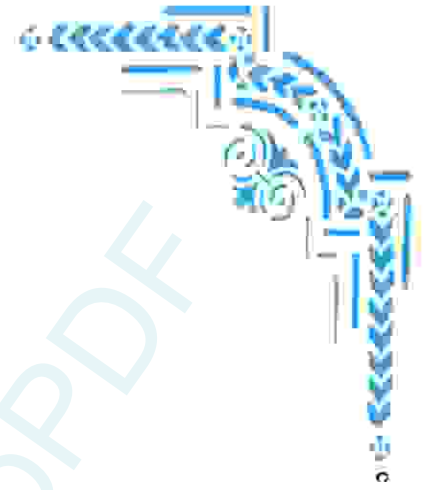
L'informatique, qu'elle soit considérée comme une science ou une technologie, tient une place croissante dans le développement des recherches en biologie. Il suffit pour s'en convaincre de considérer le grand nombre de revues, de conférences et plus généralement de publications à la frontière de l'informatique et de la biologie depuis une quinzaine d'années. Au delà de ce simple constat, peut-être est-il intéressant de comprendre et d'expliquer pourquoi. Pour ce faire, il convient de se plonger dans l'histoire de ces deux disciplines scientifiques et d'en trouver les points de liaison caractéristiques. Du point de vue de la biologie, l'on considère assez généralement que les premières utilisations remarquables de l'informatique dans le domaine se sont produites à partir des années 1970, au moment des premiers pas faits parallèlement par Sanger et Gilbert vers le séquençage de génomes à A.D.N. (Maxam, Gilbert, 1977; Sanger, Air et *al.*, 1977), qui ont soulevé la nécessité de développer des méthodes informatiques pour traiter la quantité de données émanant des génomes entiers. Sans développer plus avant les méthodes en question dès maintenant, malgré les recherches importantes menées en informatique dans ce contexte, cette vision d'un tel apport de l'informatique à la biologie se positionne principalement au travers de la dimension technologique et appliquée de la discipline informatique. Or, l'histoire montre que les premières intersections de ces deux disciplines sont plus anciennes et mettent en jeu la science informatique dans ses aspects fondamentaux. La première étude notable ayant ainsi combiné les connaissances accumulées dans ces deux disciplines date en réalité des années 1940 et a été menée par McCulloch et Pitts (McCulloch, Pitts, 1943). En proposant une première modélisation des interactions neuronales sous forme de réseaux, elle s'est construite autour de la volonté de comprendre les capacités de calcul induites par ces réseaux en considérant les neurones comme les unités logiques de base des calculs.

Aujourd'hui tout projet de biologie comporte une étape d'analyse informatique des données. Un biologiste passe environ 20-30 % de son temps à utiliser des outils informatiques. (Gilbert D et Manolo G, 2015)

Il faut développer et appliquer de nouvelles méthodes d'analyse informatique qui permettent d'extraire l'information utile cachée dans les données biologiques à grande échelle issues des progrès de l'expérimentation et des technologies de l'automatique. (Gilbert D et Manolo G, 2015)

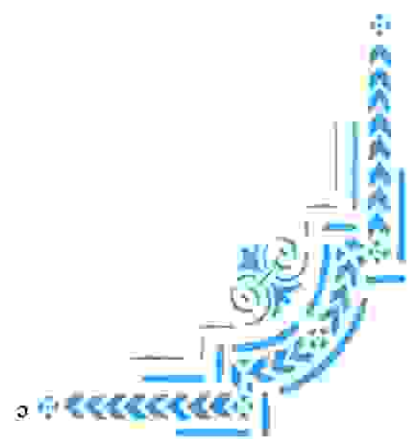
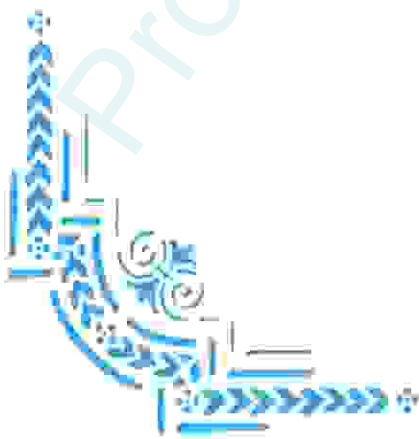
Le présent document résume notre travail et se compose de trois chapitres essentiels :

- Le premier chapitre comprend les statistiques appliquées à la biologie.
- Le second chapitre comprend le système d'information géographique appliquée en biologie.
- Le dernier chapitre rassemble toutes les informations relatives à la bioinformatique.



# *Chapitre: I*

## *Informatique et Statistiques Appliqués à la biologie*



## 1. Objectifs du chapitre :

Dans ce chapitre doit introduire toutes les analyses statistiques appliquées à des résultats obtenus à partir des expérimentations faites (*in vivo*, *in vitro*), il comprend deux parties essentielles :

- Connaissances des principes de base des calculs statistiques par l'utilisation des logiciels convenables,
- Application des différents tests statistiques pour valoriser toute recherche scientifique.

## 2. Les statistiques :

**2.1. Bref historique :** De façon un peu grossière, on peut distinguer trois phases essentielles dans l'évolution de la statistique.

Depuis l'antiquité et jusqu'à la fin du 19<sup>ème</sup> siècle, la statistique est restée principalement un ensemble de techniques de dénombrement : comptage d'une population (ou recensement), des effectifs d'une armée etc. Les techniques étaient très rudimentaires et leur mise en œuvre restait l'apanage du pouvoir politique.

Entre la fin du 19<sup>ème</sup> siècle et les années 1960, s'est construit, notamment à la suite de l'école anglaise, la statistique mathématique (ou statistique inférentielle).

Le développement de la statistique au cours de cette période, en fait, suivi le mouvement général de développement des sciences, notamment des mathématiques, de la physique et de la théorie des probabilités.

Depuis les années 1960, avec le développement et la banalisation des outils informatiques et graphiques, la statistique, et surtout la statistique descriptive multidimensionnelle, a connu une expansion considérable [1].

## 2.2. Définition de la statistique :

Le terme statistique désigne à la fois un ensemble de données d'observations, et l'activité qui consiste en leur recueil, leur traitement et leur interprétation. Les termes *statistique*, ou *statistiques* (au pluriel) englobent ainsi plusieurs notions distinctes :

\* D'une part le recensement de grandeurs d'intérêt comme le nombre d'habitants d'un pays, le revenu moyen par habitant, le nombre de séropositifs dans une population. Nous voyons que

la notion fondamentale qui se dégage de cette énumération est celle de *Population*. Une population est un ensemble d'objets, d'êtres vivants ou d'objets abstraits de même nature.

\* La statistique en tant que science s'intéresse aux propriétés des populations naturelles. Plus précisément elle traite de nombres obtenus en comptant ou en mesurant les propriétés d'une population. Cette population d'objets doit en outre être soumise à une variabilité, qui est due à de très nombreux facteurs inconnus (pour les populations d'objets biologiques qui nous intéressent ces facteurs sont les facteurs génétiques et les facteurs environnementaux) (Golmard *et al.* 2007).

### 2.3. L'utilisation de la statistique :

- Economie : assurance, finance: prévisions économétriques, analyse de la consommation des ménages, fixation des primes d'assurance et franchises, études quantitatives de marchés, gestion de portefeuille, évaluation d'actifs financiers, ...
- Biologie, médecine: essais thérapeutiques, épidémiologie, dynamique des populations, analyse du génome, ...
- Sciences de la terre: prévisions météorologiques, exploration pétrolière, ...
- Sciences humaines: enquêtes d'opinion, sondages, études de populations, ...
- Sciences de l'ingénieur: contrôle de qualité, maîtrise statistique des procédés (méthode "six-sigma"), sûreté de fonctionnement (fiabilité, disponibilité, sécurité...), maîtrise des risques industriels, évaluation des performances des systèmes complexes,...
- Sciences de l'information et de la communication: traitement des images et des signaux, reconnaissance des formes et de la parole, analyse exploratoire des grandes bases de données, analyse des réseaux de communication, ...
- Physique: mécanique statistique, théorie cinétique des gaz,...(Dagnelie, 2007)

### 3. Démarche générale en statistique :

Toute étude statistique peut être décomposée en deux phases au moins : le recueil ou la collecte des données statistiques, et leur analyse ou leur interprétation.

#### 3.1. L'identification du problème

La phase préliminaire à toute approche statistique vise à déterminer et identifier le problème par un ensemble de questionnements qui permettront de délimiter les investigations et les différentes approches :

- Quels sont les objectifs ?
- Quelle est la population ou l'échantillon à étudier ?
- Quels sont les caractéristiques et les variables ?
- Que pourra apporter une étude statistique ?

### 3.2. Le recueil des données :

Nous appellerons données les valeurs obtenus et référencés suite à une investigation ou une étude réalisée (mesures, observations, enquêtes,...).

Le recueil des données peut être réalisé soit par la simple observation des phénomènes, soit par l'expérimentation, c'est-à-dire en provoquant volontairement l'apparition de certains phénomènes contrôlés (Bertin, 1977).

**Exemple :** le rôle de quelques substances (N, P, K) dans la production de biomasse chez les végétaux.

Lorsque les données sont très nombreuses, ou particulièrement difficiles à obtenir, il sera nécessaire pour la mise en oeuvre rationnelle du recueil de définir des méthodes appropriées de collecte. Il s'agira de plans d'échantillonnage ou de plans d'expérience dont la mise en oeuvre sera fonction du type de problème que l'on est amené à résoudre (Bertin, 1977)..

- **Exemple :** la numération des mammifères d'une aire protégée : inventaire et recensement.

Il existe de ce fait plusieurs méthodes de collecte des données.

### 3.3. L'analyse et l'interprétation des données

L'analyse statistique se subdivise en deux étapes :

\* **La statistique déductive ou descriptive :** elle a pour but de résumer et de présenter les données observées sous la forme la plus accessible (simplification et réduction des données, à la fois visuelle et conceptuelle).

\* **L'analyse inductive ou inférence statistique** est l'ensemble des méthodes permettant de formuler en termes probabilistes un jugement sur une population, à partir des résultats observés sur un échantillon extrait au hasard de cette population. Les méthodes statistiques les plus classiques sont celles de l'estimation (estimation par domaine de confiance) et celles de

l'épreuve d'hypothèse. Leurs conceptions de base sont dues essentiellement à R.A. Fisher (1890 - 1962). Elle permet d'étendre ou de généraliser, dans certaines conditions, les conclusions obtenues par la statistique descriptive à partir de la fraction des individus (échantillon) que l'on a observé ou étudié expérimentalement, à l'ensemble des individus constituant la population. Les conditions (de validité) sont liées aux hypothèses faites sur la population contenant les individus et sur la façon dont ont été prises les mesures. Cette phase inductive comporte des risques d'erreur qu'il convient d'apprécier (Bertin, 1977).

Ces deux étapes sont interdépendantes. En particulier, l'observation et l'expérimentation doivent être organisées (protocole) de manière à satisfaire les conditions d'applications des méthodes de l'inférence. L'objectif de la statistique inférentielle est de fournir des résultats relatifs à une population à partir de mesures statistiques réalisées sur des échantillons (prévision, décision...) (Bertin, 1977).

#### 4. Notions de base et terminologie :

##### 4.1. Ensemble / Population / Echantillon / Élément / Individu :

\* **L'ensemble** en statistique, est la collection (finie ou infinie) d'unités, ou d'éléments, sur laquelle porte l'observation. Pour que cet ensemble soit correctement défini, il faut lui donner une définition précise de façon à ce que deux personnes différentes aboutissent toujours à la même liste d'éléments. L'ensemble des éléments observés sera appelé **E**.

\* **Les éléments** sont les objets constitutifs de l'ensemble. Ce sont des objets déterminés dont l'appartenance à tel ou tel ensemble **E** est sans ambiguïté. Les éléments peuvent être désignés par leur position dans le tableau de données : **1** pour le premier, **i** pour un élément quelconque, **n** pour le dernier élément, **N** pour la somme des éléments constituant l'ensemble.

*Exemple : Élément : membre d'une population statistique (spécimen, prélèvement d'eau, individu...) (Golmard et al. 2007)*

\* **La population** correspond à l'ensemble des **individus** sur lequel porte l'étude ou la prévision, (il est généralement difficile de l'étudier dans sa totalité), et l'**échantillon** représente la fraction de cette population qui est réellement observée ou étudiée :

- **Population-cible** : ensemble des éléments visés, en principe, par l'échantillonnage.

- **Population statistique** : ensemble des éléments effectivement représentés par l'échantillonnage. Les éléments qui la composent se caractérisent par au moins une caractéristique commune et exclusive qui permet de les distinguer sans ambiguïté.

- **Population biologique**: ensemble des individus d'une même espèce habitant un lieu donné à un moment donné. Notion qui relève davantage de la biologie que de la statistique. (Golmard *et al.* 2007)

\* **Communauté** : ensemble des individus de diverses espèces retrouvés dans un espace et un temps donnés. Notion qui relève davantage de la biologie que de la statistique.

#### Exemples généraux:

- Pour les instituts de sondage, la population étudiée sera un ensemble d'hommes et de femmes occupant une portion définie de l'espace (pays, région, commune) et l'échantillon "représentatif" sera un nombre limité mais représentatif des catégories pertinentes en fonction du problème posé (âge, sexe, catégories socio-professionnelles, origine géographique, etc.) (Golmard *et al.* 2007)

- Toute l'eau qui s'écoule d'une rivière à un moment donné constitue la population. Les 20 prélèvements de 10 cm<sup>3</sup> que l'on va analyser constituent l'échantillon.

- Le sang d'une personne peut être considéré comme une population, une prise de sang comme un prélèvement (individu, observation) et l'ensemble des prélèvements sera considéré comme un échantillon (Golmard *et al.* 2007).

\* **La notion d'individu** est très large : les éléments d'un échantillon ou d'une population sont appelés généralement des individus, cependant cette notion peut être remplacé par plusieurs dénominations: unité statistique, sujet, objet, élément, observation, mesure, doses, ... toutefois, dès que la dénomination est choisi aucune ambiguïté ne doit persister.

## 4.2. Caractère / Modalité / Variable :

4.2.1. **Le caractère** : les éléments d'un ensemble sont décrits par un caractère. Cela revient à établir une correspondance entre chaque élément  $i$  de l'ensemble  $E$  et l'ensemble  $X$  des modalités ou des valeurs du caractère. La fonction  $f: E \rightarrow X, i \rightarrow x_i$  est une application au sens mathématique : chaque élément de  $E$  a une modalité (caractère qualitatif) ou une valeur (caractère quantitatif) et une seule dans  $X$ . Ainsi le caractère peut être défini comme une des caractéristiques ou des attributs d'un individu,



**4.2.2. Modalité / Mesure :** la modalité (respectivement la mesure) est l'une des formes particulière d'un caractère. Les différentes situations où les éléments de E peuvent se trouver à l'égard d'un caractère qualitatif considéré, sont les différentes **modalités** du caractère qualitatif X.

Dans le cas où le caractère X est quantitatif, les différentes situations où les éléments de E peuvent se trouver sont des **mesures**. Ces *modalités* ou ces *mesures* doivent être à la fois **incompatibles** (un élément de E ne peut prendre qu'une seule modalité) et **exhaustive** (à chaque élément de E doit pouvoir correspondre une modalité de X) de sorte que chaque élément de E ait une modalité et une seule dans X (Golmard *et al.* 2007).

En statistique, chaque individu peut être défini par un ensemble d'attributs qui le caractérise dans le contexte d'un problème étudié.

- La couleur du pelage est un caractère
- Les variantes de la couleur du pelage sont des modalités : noir, gris,...
- Le sexe est un caractère, ses modalités sont de 2 types : soit male soit femelle
- La taille peut prendre plusieurs modalités : 1m ; 1,1m ; 1,2 m etc...

**Autres exemple de caractères :**

Le taux de glycémie, la vitesse de coagulation ; la production laitière ;

**4.2.3. Tableau élémentaire :** c'est un tableau à simple entrée où les lignes correspondent aux éléments de l'ensemble étudié et les colonnes aux caractères (ou variables) décrivant ces éléments (Tableau 1) (Golmard *et al.* 2007).

Observations	Variables			
	Variable 1	Variable 2	Variable...	Variable p
Observation 1				
Observation 2				
Observation...				
Observation n				

### 4.3. Les différents types de variable statistique:

Une variable est dite aléatoire lorsque chacune des valeurs qu'elle peut prendre est associée à une probabilité d'apparition lorsqu'un individu est choisi au hasard dans la population (Hervé, 2014).

#### 4.3.1. Les variables qualitatives:

Elles sont en général non numériques (mais pas toujours) et sont appelées facteurs. Leur valeur est appelée classe ou modalité. Elles peuvent être :

##### 4.3.1.1. Les variables qualitatives nominales:

La variable est dite qualitative nominale quand les modalités ne peuvent pas être ordonnées. Par exemple, le sexe a 2 modalités possibles : féminin ou masculin (Hervé, 2014).

##### 4.3.1.2. Les variables qualitatives ordinales:

La variable est dite qualitative ordinale quand les modalités peuvent être ordonnées. Par exemple, Par exemple : dans les catégories socioprofessionnelles, on admet d'ordonner les modalités : 'ouvriers', 'employés', 'cadres'. Si on ajoute les modalités 'sans profession', 'enseignant', 'artisan', l'ordre devient beaucoup plus discutable (Hervé, 2014).

#### 4.3.2. Les variables quantitatives:

Elles ont en général une infinité de valeurs numériques possibles. Elles peuvent être :

##### 4.3.2.1. Les variables quantitatives discrètes:

Une variable est dite discrète, si l'ensemble des valeurs possibles est dénombrable. Par exemple, le nombre de personnes dans le ménage, le nombre de cellulaires ou bien le nombre de présence au centre commercial par mois sont autant de possibilités pour des variables quantitatives discrètes (Hervé, 2014).

##### 4.3.2.2 Les variables quantitatives continues:

Une variable est dite continue, si l'ensemble des valeurs possibles est continu, (exemple: masse, temps, distance, volume) (Hervé, 2014).

## 5. Représentation des données : (Analyse des données)

Il existe plusieurs niveaux de description statistique : la présentation brute des données, des présentations par tableaux numériques, des représentations graphiques et des résumés numériques fournis par un petit nombre de paramètres caractéristiques.

### 5.1. Tableaux statistiques

En général une série statistique à caractère discret se présente sous la forme :

Valeurs  $x_1, x_2, \dots, x_p$

Effectifs  $n_1, n_2, \dots, n_p$

Fréquences  $f_1, f_2, \dots, f_p$

Plutôt que réécrire ce tableau on écrira souvent : la série  $(x_i, n_i)$ . (On n'indique pas le nombre de valeurs lorsqu'il n'y a pas d'ambiguïté). Souvent on notera  $N$  l'effectif total de cette série donc  $N = n_1 + n_2 + \dots + n_p$ . (Triola & Triola, 2012).

Table 1 The effects of bromoform and chloroform on mitotic index and mitotic phases in the root tips of *A. cepa* after 24 h

Treatments	Doses ( $\mu\text{g}/\text{mL}$ )	CCN	Mitotic index $\pm$ SD	Mitotic phases (%) $\pm$ standard deviation (SD)*			
				Prophase	Metaphase	Anaphase	Telophase
Control	-	5,109	23,35 $\pm$ 1,1a	77,8 $\pm$ 4,8a	9,8 $\pm$ 4,43ab	6,03 $\pm$ 1,14a	6,36 $\pm$ 1,91a
MMS	10	5,194	7,04 $\pm$ 0,75b	57,09 $\pm$ 6,78b	14,08 $\pm$ 3,2b	9,39 $\pm$ 3,52bc	19,42 $\pm$ 3,96b
Bromoform	25	5,235	23,02 $\pm$ 2,05a	46,58 $\pm$ 5,66c	7,84 $\pm$ 3,62a	5,11 $\pm$ 0,55a	40,46 $\pm$ 4,49c
	50	5,165	13,62 $\pm$ 0,91c	58,37 $\pm$ 3,83b	13,87 $\pm$ 3,86b	11,94 $\pm$ 1,82 cd	15,8 $\pm$ 2,38b
	75	5,138	13,11 $\pm$ 0,89c	47,01 $\pm$ 3,01c	19,48 $\pm$ 1,9c	13,68 $\pm$ 1,53d	19,82 $\pm$ 2,96b
	100	5,259	12,63 $\pm$ 0,39c	49,28 $\pm$ 6,62c	9,37 $\pm$ 2,62b	6,92 $\pm$ 0,93ab	34,41 $\pm$ 4,19d
Control	-	5,109	23,35 $\pm$ 1,1a	77,8 $\pm$ 4,8a	9,8 $\pm$ 4,43ab	6,03 $\pm$ 1,14a	6,36 $\pm$ 1,91a
MMS	10	5,194	7,04 $\pm$ 0,75b	57,09 $\pm$ 6,78b	14,08 $\pm$ 3,2b	9,39 $\pm$ 3,52b	19,42 $\pm$ 3,96b
Chloroform	25	5,231	15,5 $\pm$ 0,62c	43,47 $\pm$ 6,58c	5,31 $\pm$ 0,61b	6,17 $\pm$ 2h	45,04 $\pm$ 4,74c
	50	5,461	9,6 $\pm$ 1,15d	47,8 $\pm$ 7,16c	5,75 $\pm$ 2,54h	2,5 $\pm$ 1,92c	44,94 $\pm$ 4,48c
	100	5,066	10,37 $\pm$ 1,14d	50,9 $\pm$ 3,05bc	8,66 $\pm$ 2,89ab	7,41 $\pm$ 0,57ab	32,81 $\pm$ 1,32d
	200	5,092	16,55 $\pm$ 1,2c	49,55 $\pm$ 4,61bc	9,02 $\pm$ 1,67ab	5,46 $\pm$ 0,5a	35,96 $\pm$ 4,48d

MMS methyl methan sulfonate; SD standard deviation; CCN counting cell numbers  
\* Means with the same letter do not differ statistically at the level of 0.05

Figure 1 : Exemple de présentation des données (Khallel et al., 2015).

## 5.2. Représentation graphiques et statistique descriptive :

### 5.2.1. Statistique univariée :

Les représentations graphiques sont très importantes en statistique descriptive. Elles ont l'avantage de renseigner immédiatement sur l'allure générale de la distribution. Elles facilitent

l'interprétation des données recueillies. La représentation graphique des données montre la forme générale de la distribution et donne une image de la grandeur des nombres qui constituent les données. D'autres statistiques simples sont utilisées pour représenter le centre de la distribution et les mesures liées à la dispersion des observations autour de cette tendance centrale (Tillé, 2010).

#### 5.2.1.1. Graphiques de dispersion :

Le graphique tracé représente toutes les données individuelles d'un vecteur, d'une matrice ou d'un tableau. Il permet donc d'avoir un aperçu de la variabilité des données et d'identifier les observations aberrantes.

- Pour représenter un vecteur.
- Pour représenter plusieurs vecteurs.
- Pour donner un nom aux vecteurs sur le graphe, ajouter l'argument.
- Pour représenter des données en fonction d'un facteur où les deux objets sont des vecteurs contenant la valeur de chaque individu (dans le même ordre).
- Pour représenter les données verticalement, ajouter l'argument.
- Pour que les valeurs identiques ne se superposent pas, ajouter l'argument (Tillé, 2010)

#### 5.2.1.2. Histogrammes :

Le graphique tracé divise les données contenues dans un vecteur en classes, et représente chaque classe en effectif ou densité. Il permet donc d'avoir un aperçu de la distribution suivie par les données.

- Pour représenter les classes en effectifs.
- Pour représenter les classes en densités
- Pour ajouter une courbe de densité
- Pour ajouter une courbe de distribution

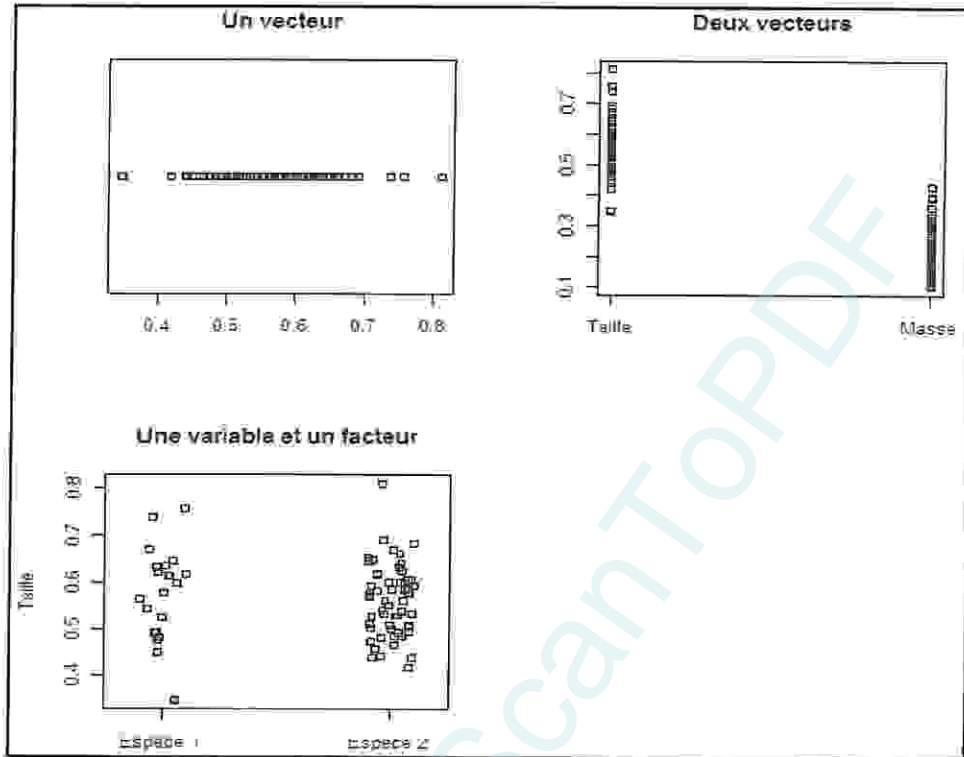


Figure 2 : Illustration de graphiques de dispersion (Hervé, 2014).

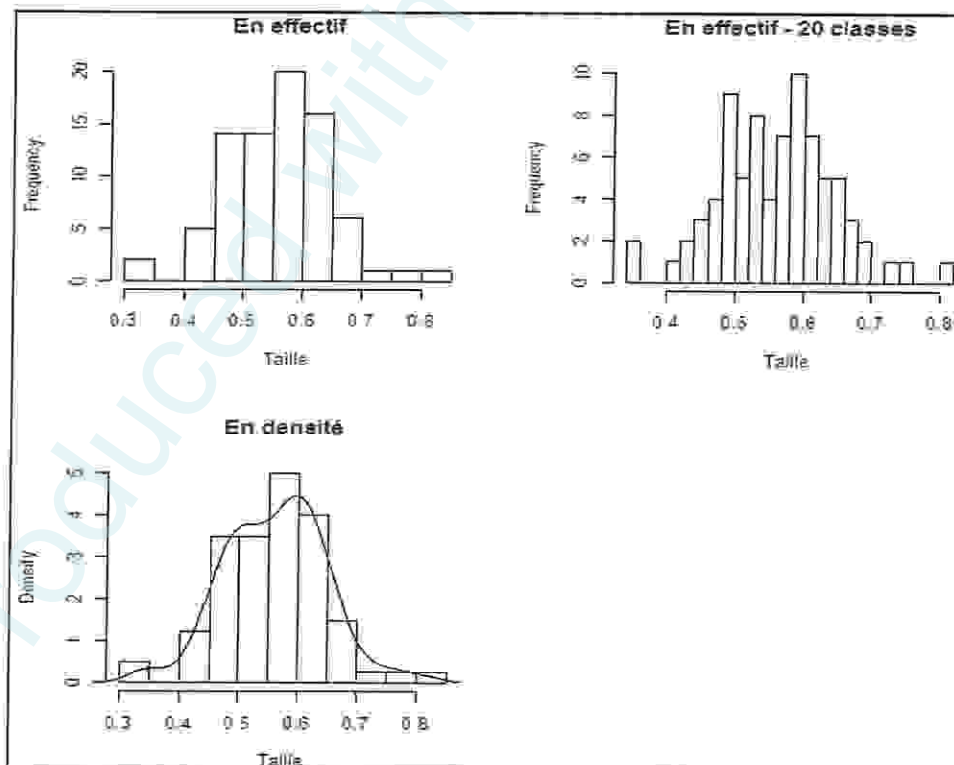


Figure 3 : Illustration des histogrammes (Hervé, 2014).

### 5.2.1.3. Boîtes à moustaches - Box Plot :

Le graphique tracé représente de façon simplifiée la dispersion des données contenues dans un vecteur. Il permet donc d'avoir un aperçu de la variabilité des données et d'identifier les observations aberrantes.

- Pour représenter un vecteur. Le trait épais représente la médiane, la boîte est formée par les valeurs des 1er et 3ème quartiles, et les moustaches mesurent maximum 1,5 fois la longueur de l'interquartile (3ème - 1er). Les valeurs au-delà des moustaches sont représentées individuellement.
- Pour représenter deux vecteurs
- Pour donner un nom aux boîtes, ajouter l'argument
- Pour représenter des données en fonction d'un facteur où les deux objets sont des vecteurs contenant la valeur de chaque individu (dans le même ordre).

Pour représenter les boîtes horizontalement, ajouter l'argument.

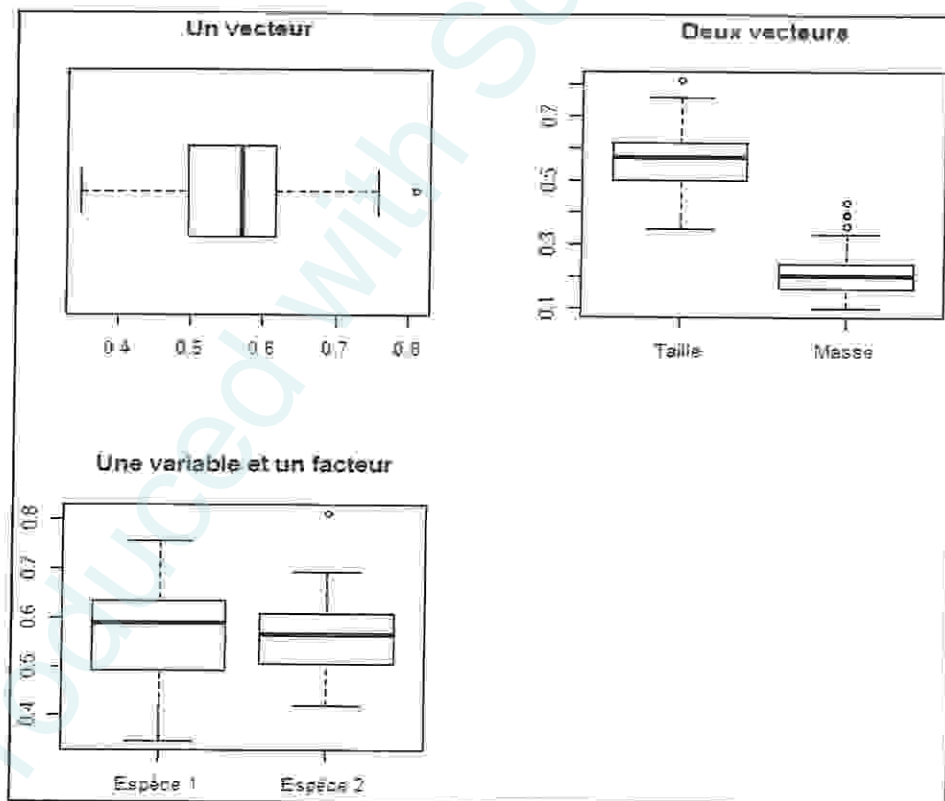


Figure 4 : Illustration des Box-plots (Hervé, 2014).

#### 5.2.1.4. La réduction des données à une dimension :

Les paramètres suivants permettent de réduire une série de données à quelques valeurs apportant une information générale.

- **Paramètres de position**

Ils permettent de donner l'ordre de grandeur des observations.

- Moyenne : La moyenne est la somme de valeurs observées divisé par leur nombre,
- Médiane : La médiane, est une valeur centrale de la série statistique.
- Mode : Le mode est la valeur distincte correspondant à l'effectif le plus élevé (Tillé, 2010).

- **Paramètres de dispersion**

Ils permettent d'estimer la variabilité des données autour des paramètres de position.

- Variance : La *variance* est la somme des carrés des écarts à la moyenne divisée par le nombre d'observations.
- Ecart-type : (*standard deviation*) L'*ecart-type* est la racine carré de la variance.
- Coefficient de variation : Le coefficient est par défaut exprimé en valeur absolue et en pourcentage (Tillé, 2010).

#### 5.2.2. Statistique bivariée :

##### 5.2.2.1. Nuages de points :

La fonction de Nuages de points permet de représenter deux caractères pour chaque individu dans un graphe du type  $y = f(x)$ , et ainsi d'avoir un aperçu de la liaison qu'il peut y avoir entre ces caractères. Elle peut être utilisée avec deux vecteurs, une matrice à deux colonnes ou un tableau.

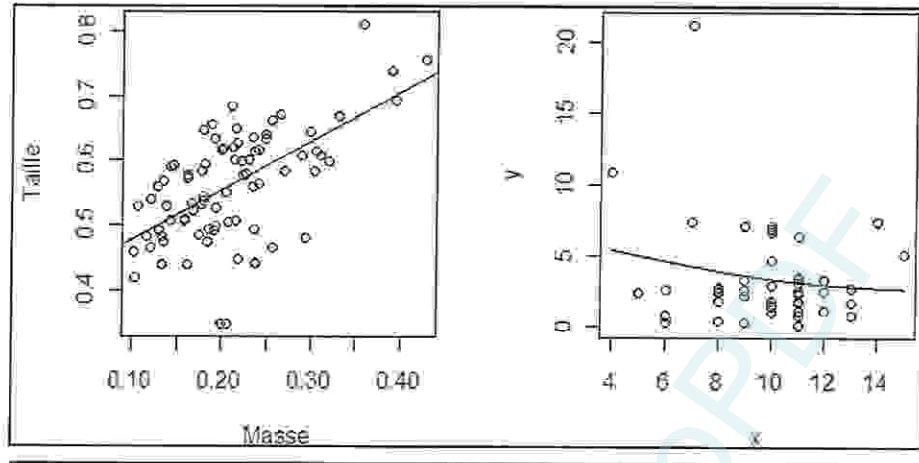


Figure 5 : Illustration des nuages de points (Hervé, 2014).

5.2.3.2. La réduction des données à deux dimensions

Les paramètres suivants permettent de réduire deux séries de données à quelques valeurs apportant une information générale sur la liaison qui peut les unir.

Pour deux vecteurs  $x$  et  $y$  contenant les valeurs de chaque individu pour chaque série (dans le même ordre) :

- Covariance.
- Coefficient de corrélation linéaire de Pearson : Le coefficient de corrélation est la covariance divisée par les deux écart-types marginaux.
- Droite de régression : La droite de régression est la droite qui ajuste au mieux un nuage de points au sens des moindres carrés.

On considère que la variable  $X$  est explicative et que la variable  $Y$  est d' dépendante:

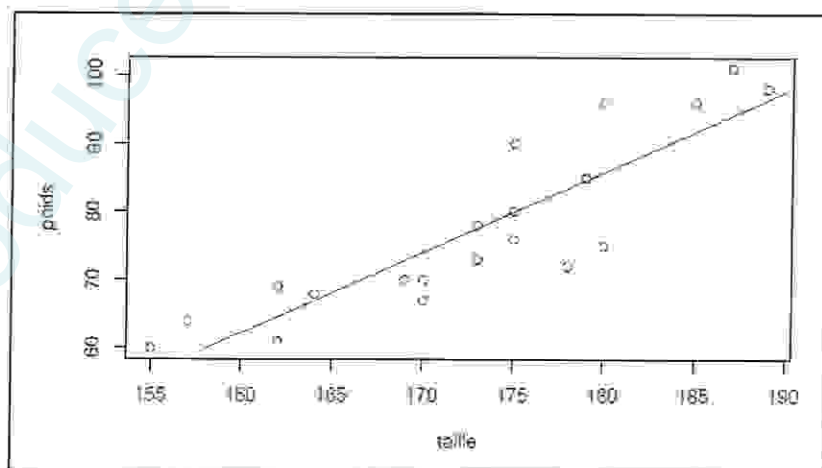


Figure 6 : Illustration de droite de régression (Hervé, 2014).



### 5.2.3. Statistique multivariée :

#### 5.2.3.1. L'Analyse en Composantes Principales (ACP exemple) :

L'analyse en composantes principales (ACP) est une méthode mathématique d'analyse graphique de données qui consiste à rechercher les directions de l'espace qui représentent le mieux les corrélations entre  $n$  variables aléatoires (relation linéaire entre elles).

Simplement dit, une ACP permet de trouver des similitudes de comportement d'achat entre les classes des données observées.

Même si l'ACP est majoritairement utilisée pour visualiser des données, il ne faut pas oublier que c'est aussi un moyen (Golmard *et al.* 2007).

- **Conditions** : les variables doivent être quantitatives ou ordinales, et les données doivent être contenues dans un tableau avec en colonnes les variables (avec un titre) et en ligne les individus. Pour visualiser la matrice des corrélations linéaires entre les variables. (Golmard *et al.* 2007).

Pour visualiser la corrélation entre chaque variable et une composante principale.

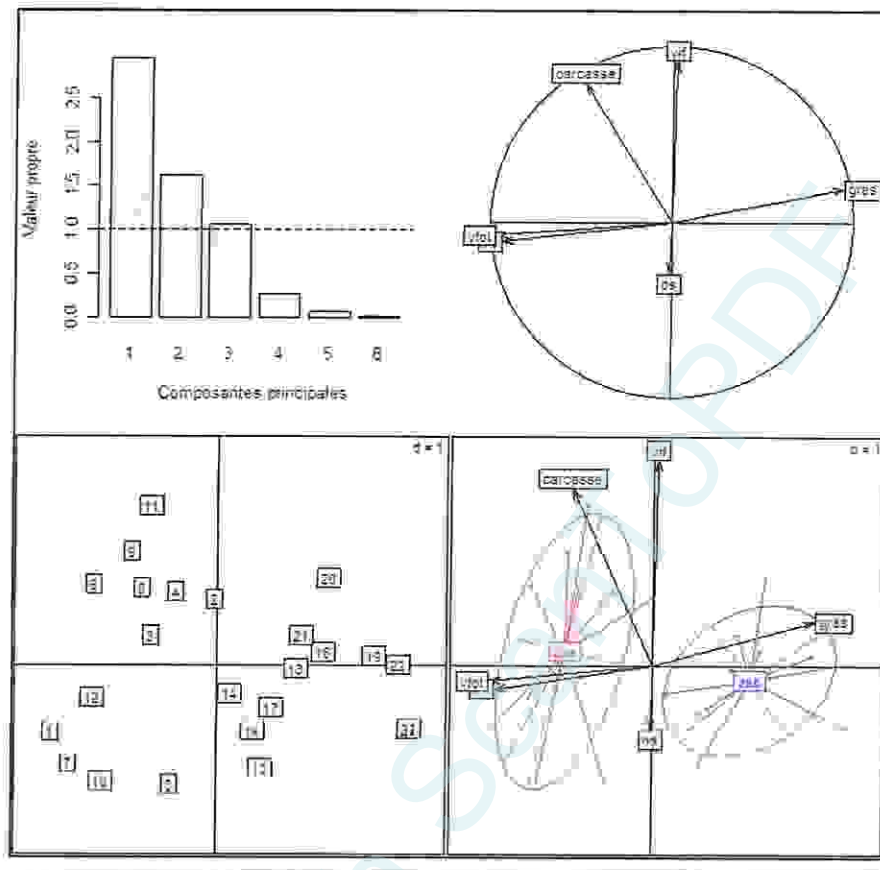


Figure 7 : Illustration d'un ACP (Hervé, 2014).

## 6. Analyse inferentielle des résultats :

### 6.1. Un peu de théorie :

La théorie (ou le calcul) des probabilités est une branche des mathématiques qui permet de modéliser les phénomènes où le hasard intervient (initialement développée à propos des jeux de hasard, puis progressivement étendue à l'ensemble des sciences expérimentales, dont la physique et la biologie) (Golmard *et al.*, 2007).

#### 6.1.1. Loïs de probabilité discontinues – généralités

Ces lois s'appliquent à des variables quantitatives discrètes. Elles sont : la loi binomiale, la loi de Poisson | loi multinomiale la loi binomiale négative la loi de Pascal La loi géométrique la loi hypergéométrique.

#### 6.1.2. Loïs de probabilité continues – généralités :

Ces lois s'appliquent à des variables quantitatives continues. Elles sont : La loi normale, la loi exponentielle, la loi gamma, la loi de *Fishe –Snedecor* et la loi de *Student*.

## 7. Tests d'hypothèses :

### 7.1. Hypothèse statistique :

Une hypothèse statistique est un énoncé (une affirmation) concernant les caractéristiques (valeurs des paramètres, forme de la distribution des observations) d'une population (Millot, 2009).

### 7.2. Test d'hypothèse:

Un test d'hypothèse (ou test statistique) est une démarche qui a pour but de fournir une règle de décision permettant, sur la base de résultats d'échantillon, de faire un choix entre deux hypothèses statistiques (Millot, 2009).

### 7.3. Principe d'un test d'hypothèses:

Les tests d'hypothèse constituent un autre aspect important de l'inférence statistique. Le principe général d'un test d'hypothèse peut s'énoncer comme suit :

- \* On étudie une population dont les éléments possèdent un caractère (mesurable ou qualitatif) et dont la valeur du paramètre relative au caractère étudié est inconnue.
- \* Une hypothèse est formulée sur la valeur du paramètre : cette formulation résulte de considérations théoriques, pratiques ou encore elle est simplement basée sur un pressentiment.
- \* On veut porter un jugement sur la base des résultats d'un échantillon prélevé de cette population (Millot, 2009).

### 7.4. Les différents types de test statistique :

Il existe quatre grandes familles de tests statistiques, qui se différencient par leur objectif :

- ✓ les tests d'homogénéité : ils permettent de tester l'égalité de paramètres de plusieurs populations (ex : égalité des moyennes entre  $n$  populations)
- ✓ les tests de conformité : ils permettent de tester l'égalité de paramètres de population(s) à une valeur théorique (ex : sex-ratio équilibré 1 : 1)
- ✓ les tests d'indépendance : ils permettent de tester si deux variables mesurées sont indépendantes l'une par rapport à l'autre

- ✓ les tests d'ajustement : ils permettent de tester si une distribution de valeurs observée est conforme à une distribution théorique ou si deux distributions observées sont identiques (Millot, 2009).

#### 8. liste de quelque logiciel statistique :

- **AreYouNet.com** : logiciel en ligne de création d'enquêtes et d'analyse statistique.
- **Minitab** : statistiques générales.
- **SPSS** : un des logiciels les plus compréhensibles techniquement.
- **Statistica** : statistiques générales, data mining et informatique décisionnelle.
- **SurveyMiner** : logiciel de traitement de données statistiques et de rapports intégré dans Microsoft Excel.
- **EduStat** : logiciel d'analyse statistique utilisé dans le domaine de l'éducation.
- **R** : une implémentation du langage S de programmation pour les statistiques [2].

#### 9. Model des logiciels :

##### 9.1. Logiciel SPSS :

###### 9.1.1 Définition :

Ce logiciel est un logiciel généraliste, bien qu'à l'origine orienté vers la statistique en sciences sociales (Statistical Package for Social Sciences). Aux États-Unis et en Grande-Bretagne, son utilisation est largement répandue dans le monde professionnel et ce dans les secteurs les plus variés. Il rencontre un succès moins large en France mais occupe cependant une place importante. Il couvre tous les champs de la statistique et il est basé sur l'utilisation de menus déroulants ; sa prise en main est rapide (Grégoire, 2012).

###### 9.1.2. Les fonction :

Ses fonctions diversifiées permettent de développer plusieurs types d'analyses :

- L'analyse « basique » (Fréquences, Moyennes, Tableaux croisés...)
- L'analyse économétrique (Régressions linéaires, multiples...)
- L'analyse de données (Analyse factorielle, ...)
- Et bien d'autres encore...

Il permet également de travailler directement sur des variables. Ex : recoder automatiquement une variable chaîne en numérique, calculer une variable qui est la somme des valeurs d'autres variables...

Quelques options relatives à des mises en forme de tableaux ou de graphiques peuvent également permettre à l'utilisateur d'imprimer directement des résultats lisibles et propres.

L'utilisation de SPSS est très variée selon les secteurs d'activité. En Banque, on l'utilise pour du *scoring*, de la segmentation de clientèle..., en cabinet d'études, on l'utilise plutôt pour des analyses basiques et des modélisations...

SPSS peut également être un support efficace de base de données [6].

### 9.1.3 Raisons de l'utilisation de ce logiciel :

Ce logiciel est enseigné pour trois raisons essentielles :

- SPSS est un logiciel professionnel (récemment racheté par IBM) largement utilisé à l'étranger (Etats-Unis, Grande-Bretagne, etc.) et qui se diffuse en France. Il a de grandes capacités de calcul et il est donc intéressant qu'il figure dans la liste des logiciels étudiés pour la recherche de stage et, à terme, pour la recherche d'emploi des étudiants. (Grégoire, 2012).
- Du point de vue pédagogique, ce logiciel offre l'avantage d'une utilisation facile, sans apprentissage informatique préalable (barre des tâches et menus déroulants avec des procédures prédéfinies) ce qui en autorise l'utilisation dès le début de la formation après une prise en main rapide. Il permet de mettre en application les méthodes statistiques et de servir de support à leur interprétation sur des données concrètes. Pour l'automatisation des tâches, ou la réalisation de celles qui ne figurent pas dans le pack de base, il peut être utilisé en programmant grâce à un éditeur de syntaxe (langage ressemblant à celui de SAS avec procédures et sous-commandes associées) (Grégoire, 2012).
- Les sorties graphiques sont de très bonne qualité et l'éditeur de graphiques permet des mises en forme personnalisées très appréciées pour les restitutions audiovisuelles ou imprimées (Grégoire, 2012).

### 9.1.4. Les différentes fenêtres de SPSS :

#### 9.1.4.1. Les trois premiers types de fenêtres :

- La fenêtre des données (DATA) (Fig.8). Cette fenêtre permet d'entrer des données, de les modifier ou de les effacer. Il est rare que l'on aille taper les données manuellement

dans SPSS car il y a trop d'erreurs de saisie possibles. On va plutôt ouvrir un fichier déjà existant (souvent généré par les instruments de mesures lors d'une expérience).

- La fenêtre des commandes (SYNTAX) (Fig.9). Cette fenêtre permet d'écrire les commandes d'analyses statistiques. Elle fonctionne comme un traitement de texte simple.
- La fenêtre des résultats (OUTPUT) (Fig.10) contient les résultats des commandes de calculs ou de manipulation des fichiers; elle fonctionne comme un éditeur (Grégoire, 2012).

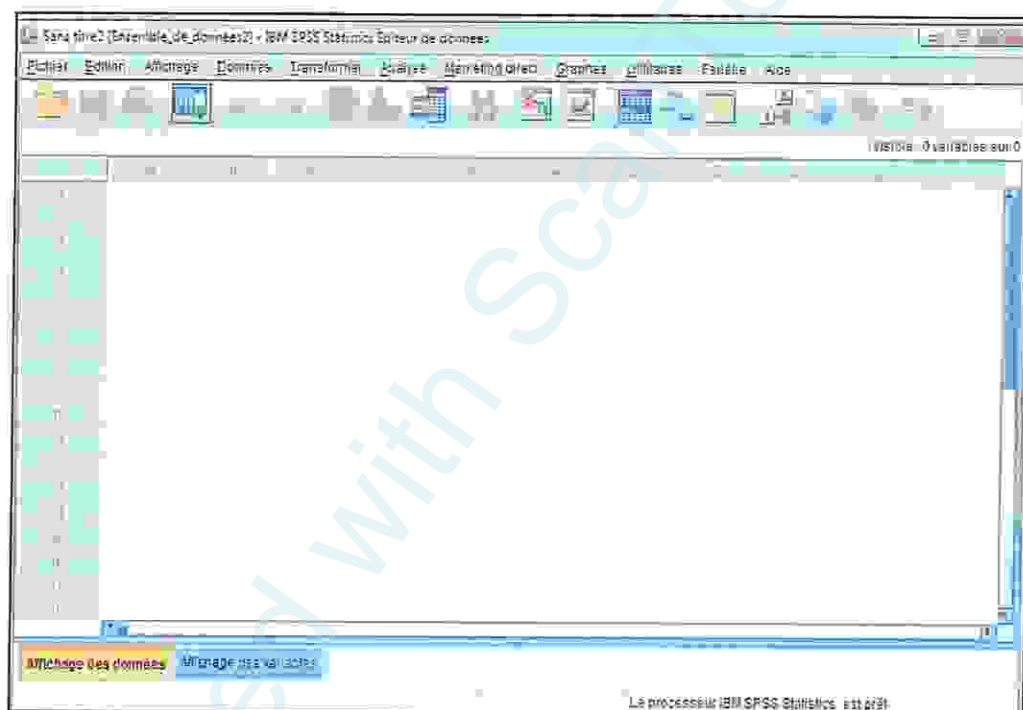


Figure 8 : La fenêtre des données (DATA).

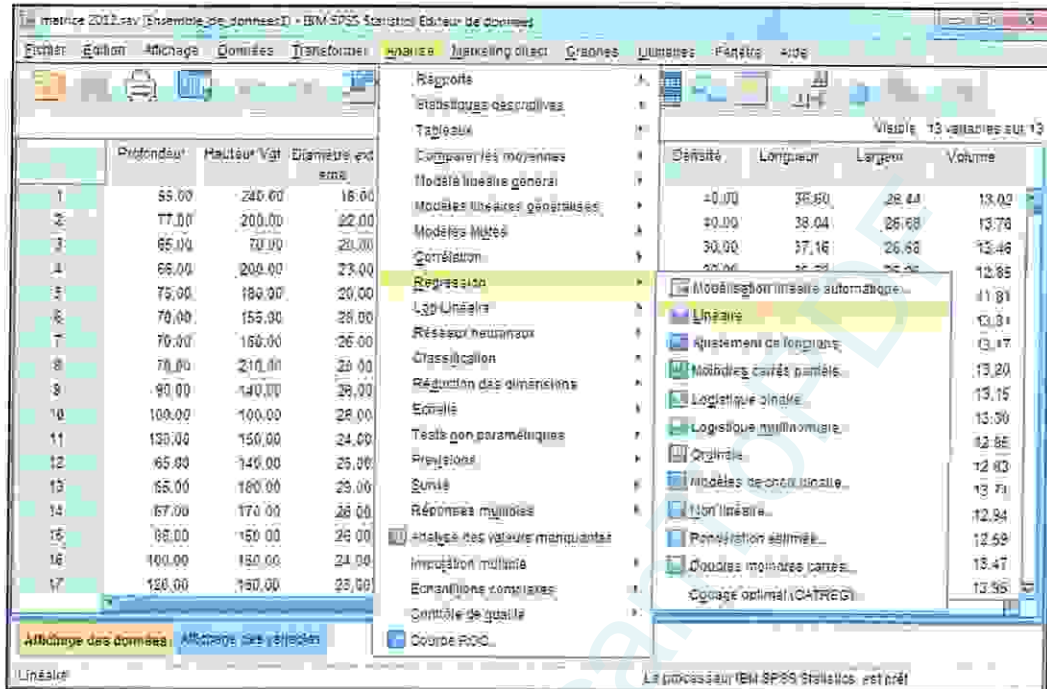


Figure 9 - 1 la fenêtre des commandes (SYNTAX).

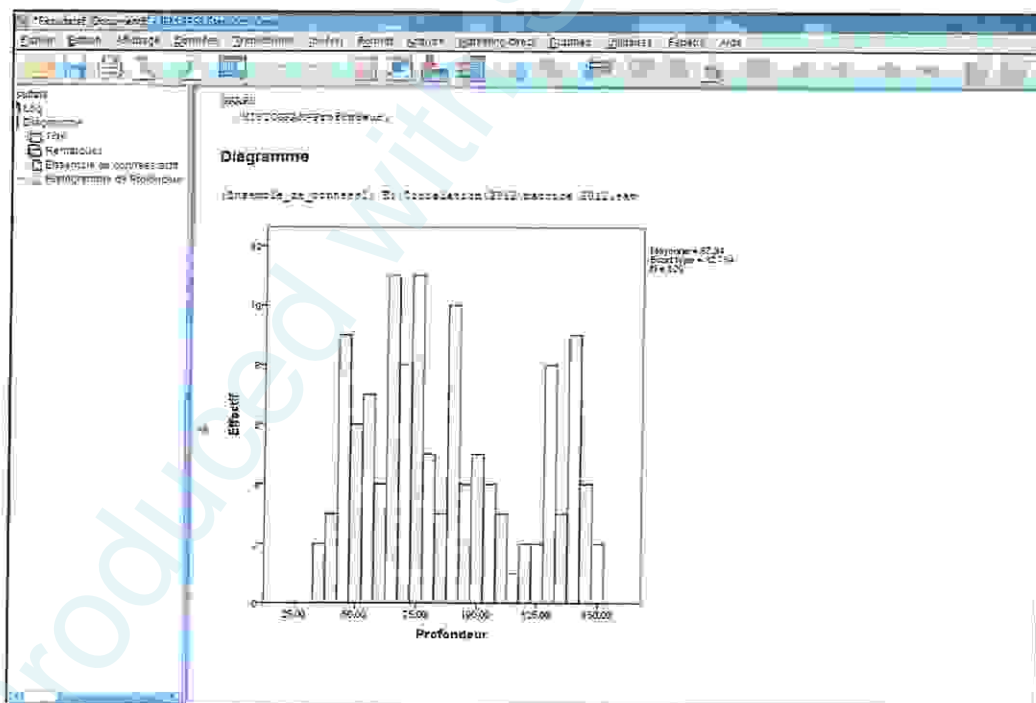
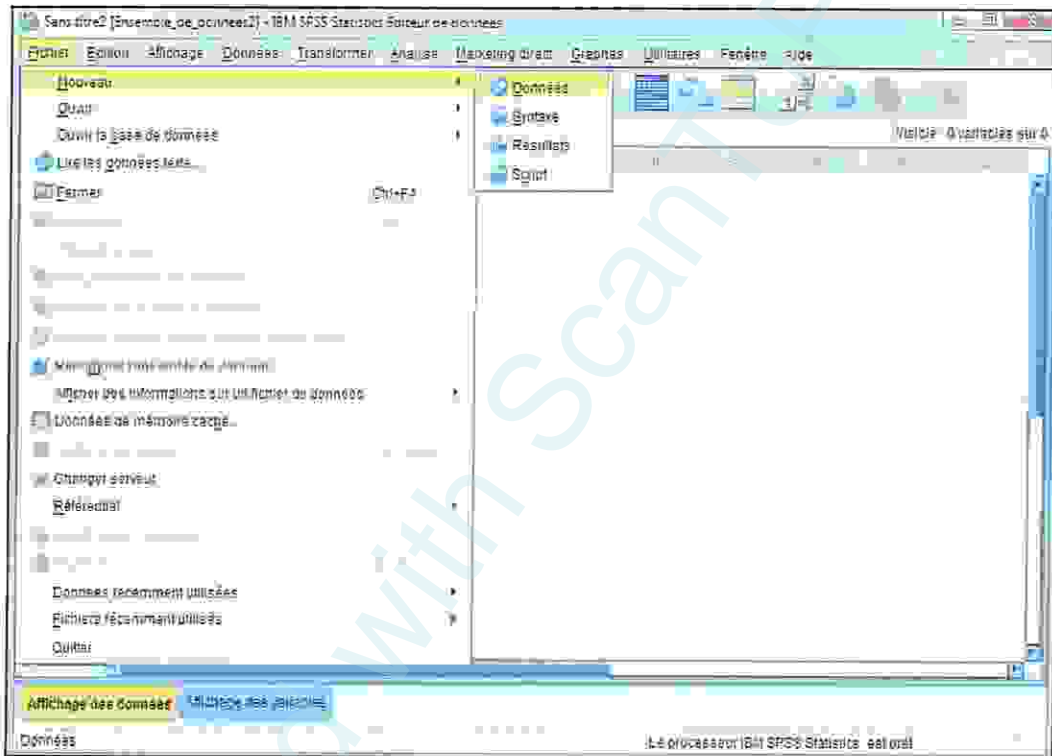


Figure 10 : La fenêtre de résultats d'un SPSS (OUTPUT).

### 9.1.5. Ouvrir un fichier de données avec SPSS :

Pour ouvrir un fichier de données SPSS (extension **.sav**), nous allons utiliser les menus déroulant. Cliquer sur « **File** », puis sur « **Open** » et finalement sur « **Data...** ». Là, vous pouvez chercher votre fichier dans le répertoire où vous l'avez enregistré.



**Figure 11** - Fenêtre d'ouvrir un fichier de données avec SPSS.

Une fois nos données ouvertes, nous pouvons explorer les deux affichages mentionnés en dessus : l'affichage des données (« **Data View** ») et l'affichage des variables (« **Variable View** ») :

Affichage des données :

- Chaque ligne représente un cas, par exemple un sujet (*case*).
- Chaque colonne représente une variable (*variable*).
- Chaque cellule contient une valeur d'un cas sur une variable.



	Profondeur	Hauteur Vgt	Diamètre ext ème	Diamètre int. me	Hauteur nid	Eau nid	Densité	Longueur	Largeur	Volume
1	55,00	240,00	18,00	7,00	4,00	30,00	40,00	38,60	26,44	13,02
2	77,00	200,00	22,00	9,00	6,50	60,00	40,00	38,04	26,88	12,78
3	66,00	170,00	20,00	7,50	4,50	35,00	30,00	37,16	26,86	13,46
4	66,00	200,00	23,00	3,00	6,00	29,00	20,00	36,62	26,26	12,66
5	75,00	180,00	20,00	7,00	5,00	36,00	40,00	36,20	26,32	11,81
6	70,00	155,00	28,00	7,00	6,00	00	30,00	38,70	26,20	13,34
7	70,00	180,00	28,00	9,00	6,00	25,00	30,00	38,12	27,74	15,17
8	70,00	210,00	25,00	7,50	6,50	20,00	30,00	37,50	28,50	17,80
9	90,00	140,00	26,00	8,00	7,00	30,00	40,00	37,00	28,42	13,15
10	100,00	100,00	28,00	8,00	8,00	00	20,00	37,10	28,54	13,30
11	120,00	150,00	24,00	7,50	6,00	00	30,00	38,22	28,60	12,95
12	65,00	140,00	25,00	7,00	4,50	38,00	30,00	37,00	26,10	12,89
13	65,00	160,00	29,00	9,00	7,50	20,00	48,00	36,94	27,00	13,71
14	87,00	170,00	28,00	8,00	5,50	30,00	25,00	37,60	26,00	12,94
15	65,00	150,00	26,00	8,00	6,00	00	10,00	38,10	26,18	12,59
16	100,00	160,00	24,00	8,00	5,50	00	10,00	38,78	26,12	13,47
17	120,00	180,00	23,00	3,50	6,00	30,00	40,00	39,32	26,40	13,95

Figure 12 : Fenêtre d'affichage de données avec SPSS.

Vue des variables :

- Chaque ligne représente une variable.
- Les colonnes décrivent les caractéristiques des variables.

Nom	Type	Largeur	
1	Profondeur	Nomérique	8
2	Hauteur Vgt	Nomérique	6
3	Diamètre ext	Nomérique	8
4	Diamètre int.	Nomérique	8
6	Hauteur nid	Nomérique	0
6	Eau nid	Nomérique	0
7	Densité	Nomérique	8
8	Longueur	Nomérique	9
9	Largeur	Nomérique	8
10	Volume	Nomérique	8
11	GP 12	Nomérique	8
12	succès éco	Nomérique	8
13	GP 11	Nomérique	3

Figure 13 : fenêtre d'affichage des variables avec SPSS.

## 9.2. Logiciel R :

### 9.2.1. Définition :

R est un logiciel statistique qui permet la lecture, la manipulation et le stockage de données. La grande majorité des méthodes statistiques actuelles y sont présentes par défaut ou au sein de « packages » dont la liste est en constante évolution. Les principaux facteurs qui expliquent son importance actuelle sont sa gratuité (licence GNU-GPL), sa fiabilité et sa disponibilité sous la plupart des systèmes d'exploitation (Windows, Mac OS, Linux, Unix). Initialement conçu pour illustrer l'enseignement de la statistique, R a connu une croissance exponentielle pendant les quinze dernières années dans le monde académique. Son développement actuel lui permet de rivaliser avec la plupart des logiciels payants utilisés dans les entreprises (Grégoire, 2012).

### 9.2.2 Pourquoi utiliser R :

Tout d'abord R est un logiciel gratuit et a code source ouvert (open-source). Il fonctionne sous UNIX (et Linux), Windows et Macintosh. C'est donc un logiciel multi-plates-formes. Il est développé dans la mouvance des logiciels libres par une communauté sans cesse plus vaste de bénévoles motivés [4].

Tout le monde peut d'ailleurs contribuer à son amélioration en y intégrant de nouvelles fonctionnalités ou méthodes d'analyse non encore implémentés. Cela en fait donc un logiciel en rapide et constante évolution. C'est aussi un outil très puissant et très complet, particulièrement bien Adapté pour la mise en œuvre informatique de Méthodes statistiques. Il est plus difficile d'accès que certains autres logiciels du marché (comme SPSS ou Minitab par exemple), car il n'est pas conçu pour être utilisé à l'aide de < clics > de souris dans des menus. L'avantage en est toute fois double :

\* L'approche est pédagogique puisqu'il faut maîtriser les méthodes statistiques pour parvenir à les mettre en œuvre.

\* L'outil est très efficace lorsque l'on domine le langage R puisque l'on devient alors capable de créer ses propres outils, ce qui permet ainsi d'opérer des analyses très sophistiquées sur les données [4].

### 9.2.3. R et les statistiques:

R est un logiciel dans lequel de nombreuses techniques statistiques modernes et classiques ont été implémenté. Les méthodes les plus courantes permettant de réaliser une analyse statistique telles que :

- statistique descriptive.
- tests d'hypothèses.
- analyse de la variance.
- méthodes de régressions linéaires (simple et multiple) ; sont enchâssé directement dans le cœur du système. Notez également que la plupart des méthodes avancées de statistique sont aussi disponibles au travers de modules externes appelés les packages.

Ceux-ci sont faciles à installer directement a partir d'un menu du logiciel. Ils sont tous regroupés sur le site internet du Comprehensive R Archive Network (CRAN) (<http://cran.r-project.org> ) sur lequel vous pouvez les consulter. Ce site fournit aussi, pour certains grands domaines d'études, une liste commentée des packages associés a ces thèmes (appelée *Task View*), ce qui facilite ainsi la recherche d'une méthode statistique particulière. Par ailleurs, une documentation détaillée en anglais de chaque package est disponible sur le CRAN. Il est par ailleurs utile de noter que les méthodes statistiques les plus récentes y sont régulièrement ajoutées par la communauté statistique elle-même [4].

### 9.2.4 .R et les graphiques :

Une des grandes forces de R réside dans ses capacités, bien supérieures a celles des autres logiciels courants du marché, a combiner un langage de programmation avec la possibilité de réaliser des graphiques de qualité. Les graphiques usuels s'obtiennent aisément au moyen de fonctions prédéfinies. Ces dernières possèdent de tres nombreux paramètres permettant par exemple d'ajouter des titres, des légendes, des couleurs, etc. Mais il est également possible d'effectuer des graphiques plus sophistiqués permettant de représenter des données complexes telles que des courbes de surface ou de niveau, des volumes axes avec un e et 3D, des courbes de densité, et bien d'autres choses encore. Il vous est également possible d'y ajouter des formules mathématiques. Vous pouvez aussi agencer ou superposer plusieurs graphiques sur une même fenêtre, et utiliser de nombreuses palettes de couleur [4] (Figure 15).

### 9.3.5 L'interface R:

L'interface R (Figure 14) est composée de quatre fenêtres :

- Fenêtre d'édition (en haut à gauche) : dans cette fenêtre apparaissent les fichiers contenant les scripts R que l'utilisateur est en train de développer. Enregistrer le fichier avec une extension permet une coloration syntaxique adaptée au langage R. En entête de cette fenêtre, des icônes permettent de sauvegarder le fichier, d'exécuter un morceau de code sélectionné (icône run) ou l'intégralité du code contenu dans le fichier (icône source).
- Fenêtre de commande (en bas à gauche) : cette fenêtre contient une console dans laquelle les codes R sont saisis pour être exécutés.
- Fenêtre espace de travail / historique (en haut à droite) : contient les objets en mémoire, que l'on peut consulter en cliquant sur leur noms, ainsi que l'historique des commandes exécutées.
- Fenêtre explorateur / graphique / package / aide (en bas à droite) : l'explorateur permet de se déplacer dans l'arborescence des répertoires, la fenêtre graphique contient les graphiques tracés via R (il est possible de les exporter), la fenêtre package montre les packages installés et actuellement chargés et la fenêtre d'aide contient la documentation sur les fonctions et packages [7].

## 10. Quelques statisticiens :

- **PEARSON Karl**

Statisticien anglais (1857 - 1936). "Il s'est donné pour mission de faire fructifier au niveau théorique les problèmes posés par l'application de la statistique à la biologie : il se consacre donc à l'étude des probabilités, mettant au point la fameuse formule du chi carré". F. Bédarida, 1977. Statistique et société en Angleterre au XIX e siècle. In *Pour une histoire de la statistique*, INSEE. Il soutint Galton et sa théorie de "l'eugénisme" (Kuhn, 1979).

- **FISHER Ronald**

Statisticien anglais (1890 - 1962). À partir de ses expérimentations agronomiques, il tenta de montrer que, même si les postulats de normalité relevaient souvent de l'abus de confiance, cela ne détériorait pas trop la validité des conclusions. Dans ce cadre, il fut l'un des premiers à développer les tests de permutations des rangs ou de randomisation avec Pitman et Welch. *The design of experiments*. Oliver & Boyd, Edimbourg, 1935 (Kuhn, 1979).

- **STUDENT (GOSSET) William Sealy**

Statisticien anglais (1876 - 1937). W.S. Gosset a publié sous le nom de *Student*. Il travaillait pour l'industrie de la bière (maison Guinness). Il fit progresser la statistique dans le domaine des probabilités (Kuhn, 1979).

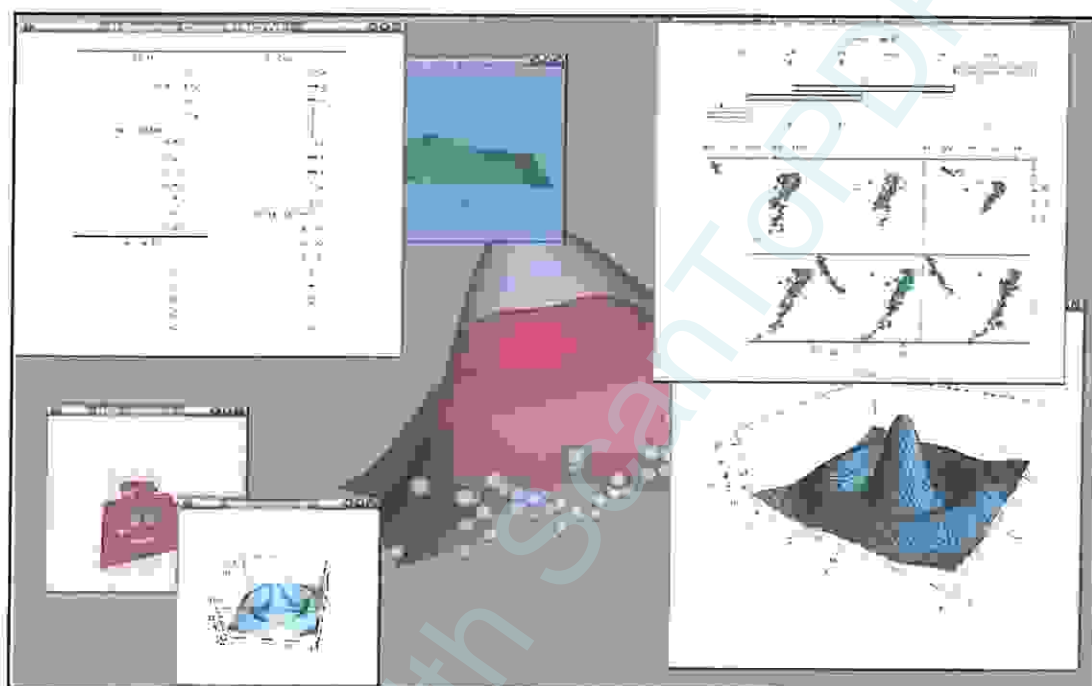


Figure 14 : Quelques possibilités graphiques offertes par R [4].

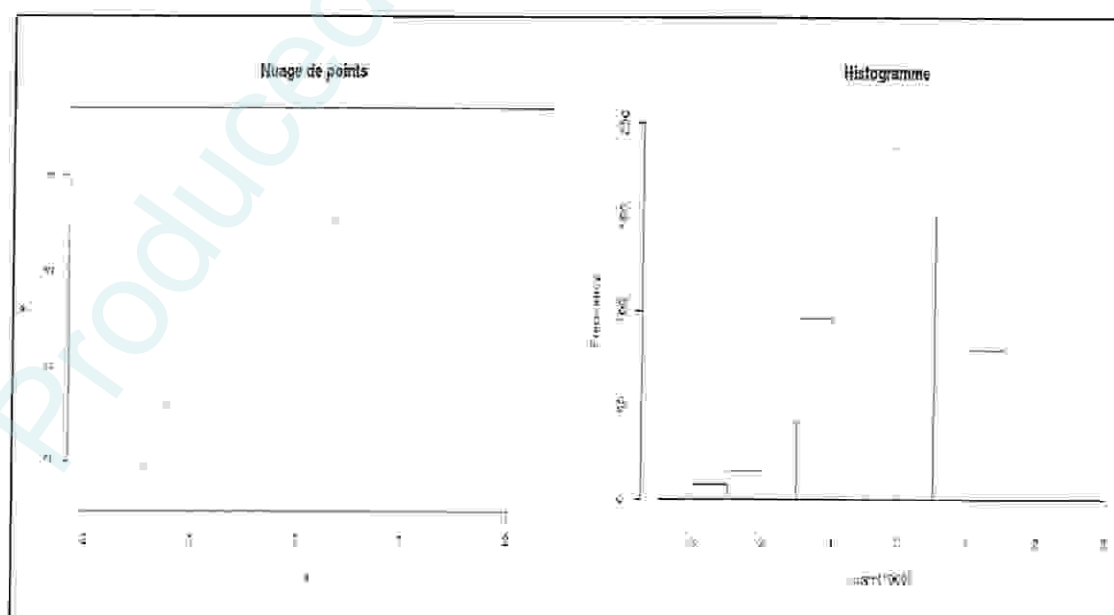


Figure 15 : Interface graphique R [5].

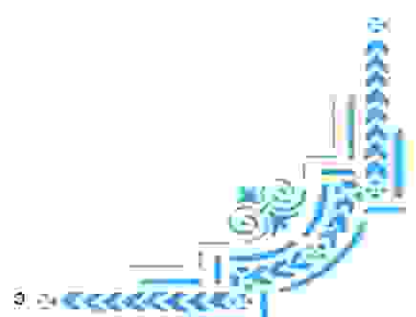
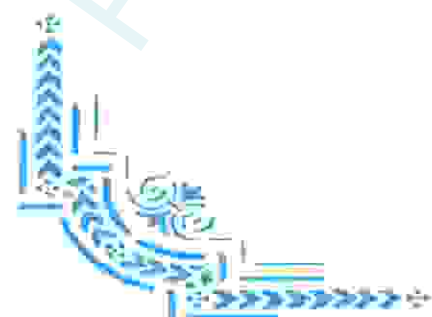


e

e

## *Chapitre: II*

### *Le système d'information géographique appliquée en biologie*



Produced by Scantopdf

Un système d'Information Géographique (S.I.G) est un système informatique permettant, à partir de diverses sources, de rassembler et d'organiser, de gérer, d'analyser et de combiner, d'élaborer et de présenter des information localisées géographiquement, contribuant notamment à la gestion de l'espace (figure 16) (Rahal, 2005).

### 1. Objectif :

- ✓ Les S.I.G sont des outils d'aide à la décision et de prospective.
- ✓ très utiles dans une stratégie d'aménagement de l'espace.
- ✓ Une des finalités des S.I.G est de produire des cartes et des analyses, afin d'élaborer un diagnostic, aider à la décision et servir d'appui aux débats. (Rahal, 2005)

### 2. Bref historique : Les premiers SIG

Les premiers SIG opérationnels sont apparus dans les années 1960, au Canada et aux Etats Unis. Le pionnier est indiscutablement la canadienne géographie information system (1964), qui rassemble des informations relatives à l'usage du sol, et des données concernant l'environnement, sur une grande partie de territoire canadien. Le logiciel a été développé pour ces besoins spécifiques. Deux autres réalisations précoces méritent d'être mentionnées : le New York Land Use Information System (1967) et le Minnesota Land Management (1969). Depuis cette époque les coûts et les difficultés techniques ont considérablement diminué, et de nombreux logiciels commerciaux sont aujourd'hui disponibles, offrant de bonnes performances à un prix raisonnable. (Hammoun & Bouzida, 2010)

### 3. Définition d'un Système d'Information Géographique (SIG) :

Un système d'information géographique est un ensemble de logiciels intégrés, spécifiquement conçu pour être utilisé avec des données géographiques, qui comprennent à la fois les données spatiales et les données attributaires décrivant les caractéristiques géographiques, il effectue une gamme complète de fonction de base comprennent La saisie (inputs), le stockage, l'extraction, la transformation et l'affichage (Output). En plus d'une grande variété de procédés descriptifs et analytiques, de cela on peut dire que Le concept de base des SIG décrit à la fois l'emplacement, la distribution spatiale et la relation entre les deux (Qi hao, 2010) (Figure.16).

Les S.I.G se placent au croisement de quatre disciplines auxquelles ils empruntent une partie de leurs histoires, de leurs principes et de leurs technologies. (Figure 17)(Rahal, 2005).

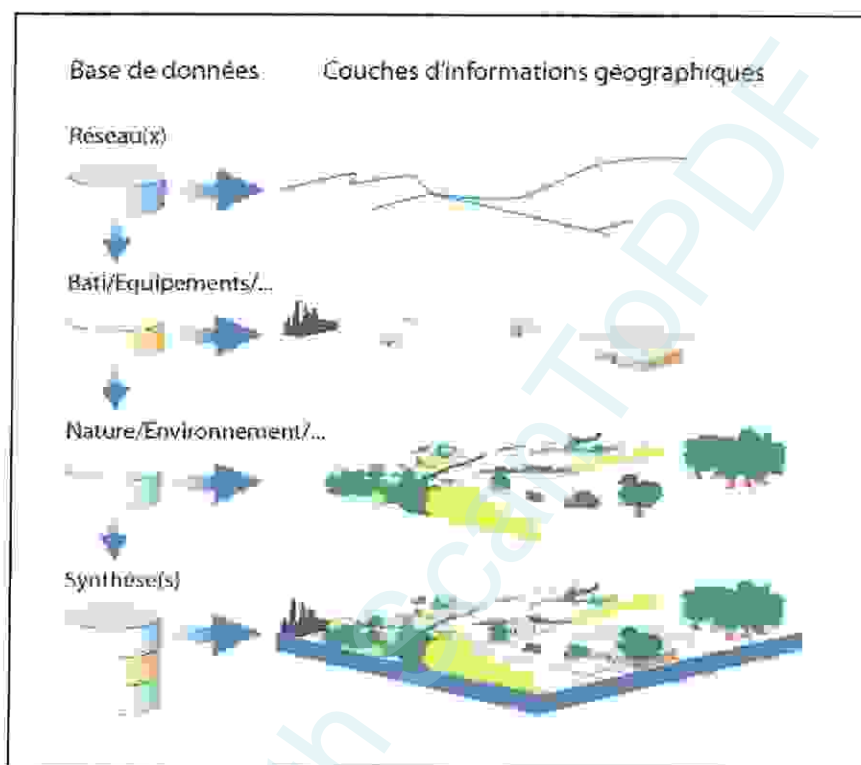


Figure 16 : Modes de représentation de l'information géographique dans un SIG[6].

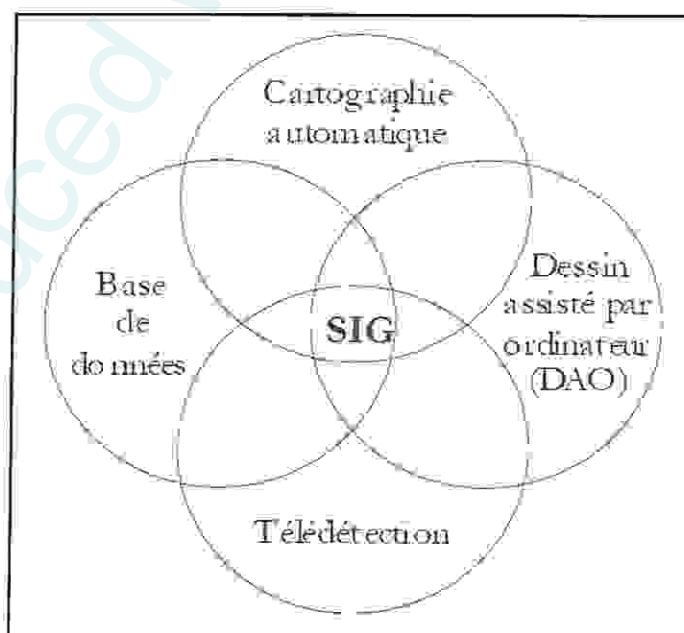


Figure17 : Les disciplines contribuant au S.I.G(Rahal, 2005).



#### 4. Eléments et fonctions d'un SIG :

Un SIG est un outil d'aide à la décision composé d'éléments (des personnes, du matériel, des logiciels, des données) et des fonctions (saisie, stockage, gestion, traitement, sélection, analyse).

Les éléments indispensables au fonctionnement d'un SIG sont les suivants :

- Les personnes : c'est l'élément le plus important dans un SIG. Les personnes doivent définir les traitements et développer les procédures d'exploitation. Disponibilité et exactitude des données : elles conditionnent les possibilités d'interrogation et d'analyse.
- Le matériel : Les capacités du matériel conditionnent la capacité d'exploitation, la facilité d'utilisation et le type de sortie possible.
- Les logiciels : On entend par là les SIG, mais aussi les logiciels de bases de données, de dessin, de statistiques d'imagerie ou d'applications spécifiques.

Afin d'obtenir les réponses aux questions posées, une série d'étapes doit être suivie pour le traitement de l'information. La combinaison de ces éléments donne accès aux fonctions suivantes :

- La saisie, le stockage et la gestion de données.
- Le traitement, l'analyse, la sélection et l'affichage de données.
- La production de cartes, rapports, tableaux et autres synthèses d'information.

(Hammoun & Bouzida, 2010).

#### 5. Structure d'un SIG :

La figure (19) met en évidence quatre groupes de fonctionnalités au-dessous d'une couche d'applications: l'acquisition des données géographiques d'origines diverses, la gestion pour le stockage et la recherche des données, l'analyse spatiale pour le traitement et l'exploitation et enfin la présentation des résultats sous forme cartographique. (Abdelbaki C., 2012)

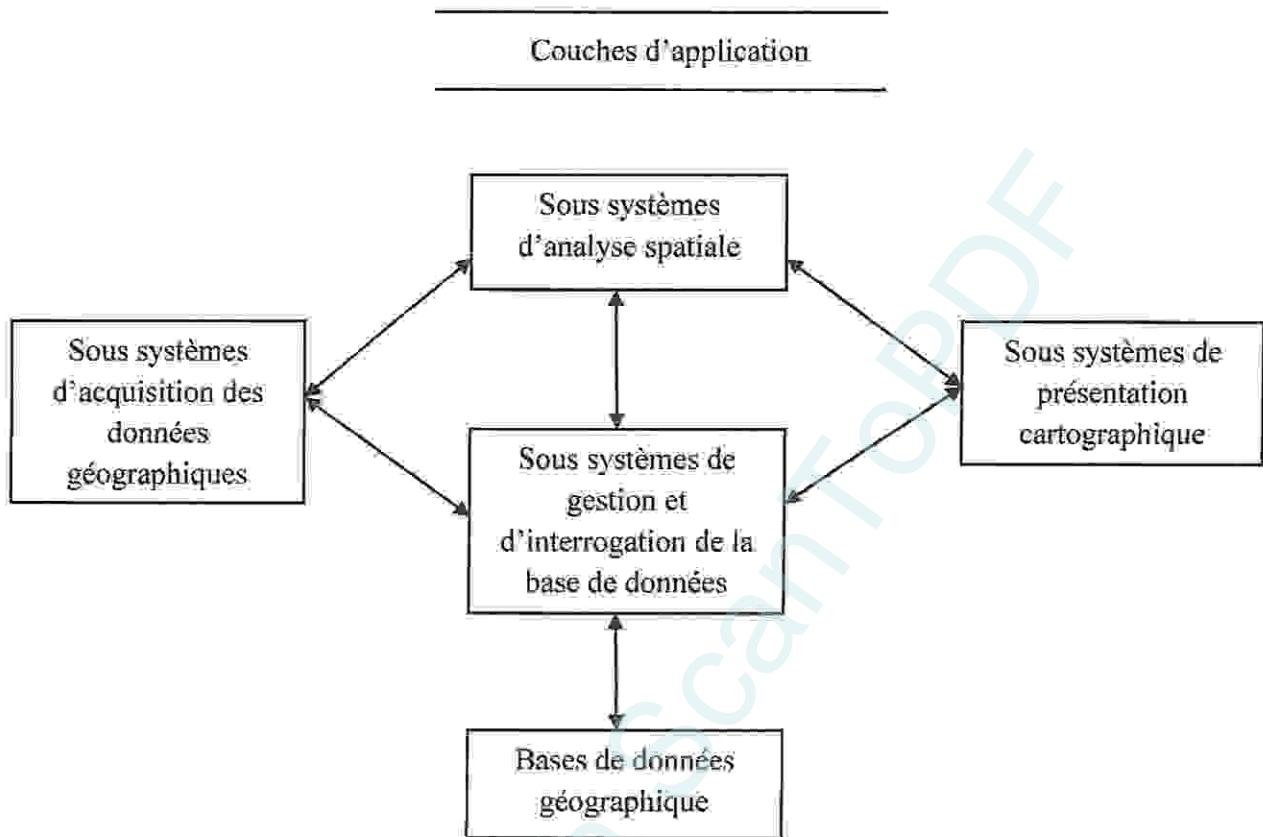


Figure 18: Structure d'un SIG (Abdelbaki, 2012).

## 6. Fonctionnalités de base des SIG:

La gamme de fonctionnalités couramment associés à des SIG en tant que technologie est:

**6.1. La saisie de données :** Les fonctions de saisie de données sont étroitement liées aux disciplines de, la photogrammétrie, la télédétection, en particulier, est le domaine qui offre des photographies et des images à partir duquel on obtient des séries de données (Rolf, 2001).

### 6.2. La Transformation:

Cela comprend un certain nombre de processus clés parmi lesquels : la transformation vecteurs-raster et vice versa, la transformation du système de coordonnées et de la projection des cartes, et le reclassement des attributs. (Verbyla, 2002)

### 6.3. Stockage de données :

Dans la plupart des systèmes de traitement disponibles, les données sont organisées en couches selon les différentes catégories thématiques, par exemple : formation végétale,

hydrologie et toponymie, la superposition de ces couches doit refléter les phénomènes et leurs relations aussi proches que possible de ce qui existe dans la réalité (Rolf, 2001).

#### 6.4. L'analyse spatiale :

Une des principales fonctions d'un système d'information géographique (SIG) est l'analyse spatiale, c'est-à-dire la description statistique ou bien l'explication soit de localisation ou des informations attributs. (Fotheringham *et al.*, 1994)

**6.4.1. Fonctions de mesure :** comprennent le calcul des distances entre les caractéristiques ou le long de leur périmètre, et le calcul de dimension d'une zone 2D ou bien le volume d'une zone 3D.

**6.4.2. La requête spatiale :** elle permet d'identifier les caractéristiques définies par l'utilisateur de manière sélective (conditions logiques). Ceci inclut la recherche par zone pour en extraire leurs attributs, recherche par attributs pour extraire les zones correspondantes et la sélection des caractéristiques d'une couche sur la base des caractéristiques d'une autre. (Verbyla, 2002).

- **La sélection spatiale Interactive :**

Dans la sélection spatiale interactive, on définit la condition de sélection par le pointage ou le dessin des objets spatiaux sur l'écran après avoir indiqué la couche(s) de données spatiales à partir de laquelle on sélectionne des fonctions. Les objets interactivement définis sont appelés les objets de sélection, ils peuvent être des points, lignes ou polygones.

La Sélection des caractéristiques conduit à la Sélection des enregistrements. Et Viceversa, on distingue :

- \* Sélection spatiale par les conditions d'attributs
- \* Sélection de caractéristiques qui sont à l'intérieur des objets sélectionnés :
- \* Sélection de caractéristiques contiguës à un objet sélectionné (MacEachren *et al.*, 2005)
- \* Sélection de caractéristiques en fonction de leur distance (Navrafil, 2009)

**6.4.3. Fonctions de superposition :** c'est quand deux ou plusieurs couches sont fusionnées pour créer une nouvelle couche. (Verbyla, 2002), Les personnes qui ont introduit la superposition au SIG pensées de ce type d'analyse de l'information géographique en terme de voir les interactions entre deux ou plusieurs cartes transparentes.

**6.4.4. Fonctions de proximité:** la fonction de voisinage évalue les caractéristiques d'une zone autour d'un emplacement. Cela permet de regarder des « zones tampons » autour de caractéristiques, et la propagation des effets si les caractéristiques sont une source de

quelque chose qui se propage-par exemple, les sources d'eau, les éruptions volcaniques, les sources de pollution(Wainwright *et al.*, 2005).

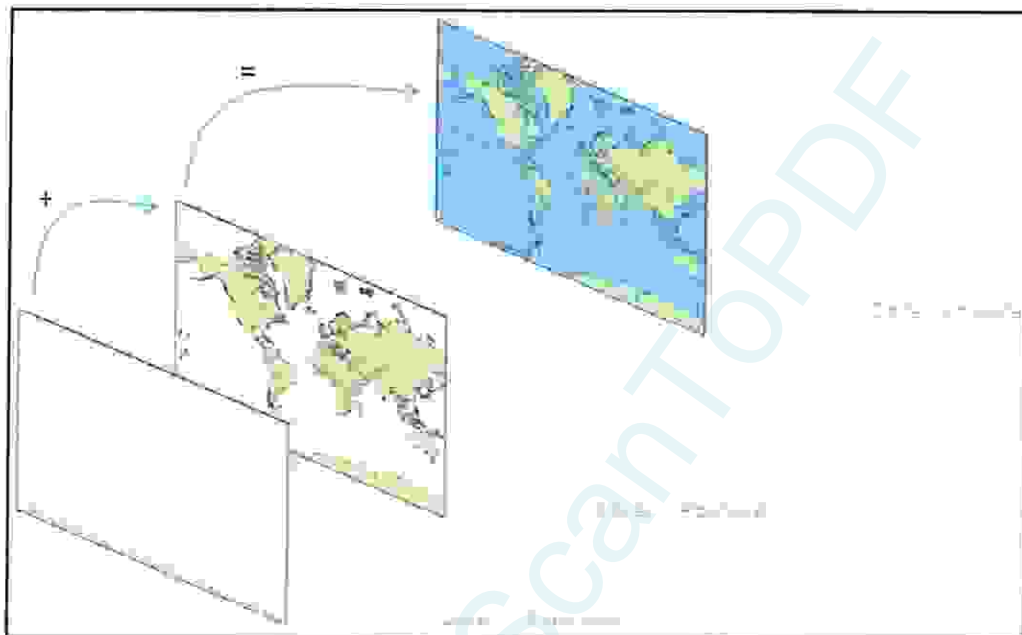


Figure 19: Procédé de superposition de couches[10].

**6.4.5. Fonctions de réseau :** permet d'évaluer comment les caractéristiques sont connectés. Ceci est utile dans les applications qui étudient les réseaux de caractéristiques connectées comme les réseaux routiers, les cours d'eau dans les zones côtières, et les lignes de communication dans la téléphonie mobile.

**6.4.6. Classification:** La classification est une technique de retirer intentionnellement les détails à partir d'un ensemble de données saisies, dans l'espoir de révéler des tendances importantes de la distribution spatiale.

### 6.5. Visualisation :

Le SIG est un support technique permettant une bonne visualisation des données spatiales (Bengueraï *et al.*, 2009), Cette fonction du SIG est étroitement liée aux disciplines de la cartographie, L'imprimerie et L'édition (Rolf, 2001).

**6.5.1. Définition de la cartographie :** La cartographie est une forme de communication et peut être vu comme une forme de langage spatial pour décrire les emplacements, discuter les lieux et interpréter les arrangements bidimensionnels des fonctionnalités, elle a été également définie comme « la production y compris la conception, la compilation, la construction, la

- **Carroyage** : un réseau des parallèles de latitude et des méridiens de longitude tracés sur carte à fin de faciliter le repérage sur terrain.
- **Cartouche** : contient le nom de l'institution, le groupe ou l'individu responsable de la réalisation de la carte. (C.D.S.P.I., 2009 in Zerroug, 2012)

- **Notion de la carte dans les SIG :**

Ici nous utilisons le terme cartographie pour décrire le produit numérique ou analogique d'un SIG qui affiche des informations géographiques à l'aide d'un support cartographique. La carte dans les SIG est le résultat final d'une série d'étapes de traitement de données SIG en commençant par la collecte des données, l'enregistrement, la maintenance, l'analyse et terminant avec une carte. Chacune de ces activités successives transforme une base de données d'informations géographiques jusqu'à ce qu'elle soit dans la forme appropriée à l'affichage sur une technologie donnée (Longley *et al.*, 2005).

#### 7. Les domaines d'application des S.I.G. :

Les domaines d'application des SIG sont aussi nombreux que variés :

- **Tourisme** : gestion des infrastructures, itinéraires touristiques.
- **Marketing** : localisation des clients, analyse d'un site sur le plan commercial.
- **Planification urbaine** : cadastre, POS, voirie, réseaux assainissement.
- **Protection civile** : gestion et prévention des catastrophes.
- **Transport** : planification des transports urbains, optimisation d'itinéraires.
- **Hydrologie** : état des eaux superficielles, réseau d'observations sur les eaux de charriage.
- **Forêt** : cartographie pour aménagement, gestion des coupes et sylviculture.
- **Géologie** : prospection minière.
- **Biologie** : études du déplacement des populations animales.
- **Télécoms** : implantation d'antennes pour les téléphones mobiles (Rahal, 2005).

#### 8. L'intérêt des S.I.G :

L'un des principaux intérêts des S.I.G est de pouvoir croiser les informations entre elles et d'en tirer une plus-value. D'après Robin (2002) les S.I.G permettent de :

- \* La création et la récupération des données sous divers formats, en mode vecteur comme, si possible, en mode raster ;
- \* La gestion de ces données (système de gestion de base de données) ;

projection, la reproduction, l'utilisation et la distribution des cartes », la cartographie de nos jours principalement menée en Systèmes d'Information Géographique (Anne Ruas, 2011 in Zerroug, 2012)

- **Types de cartes : il y a deux types de cartes :**

- **Carte topographique:** Ces cartes sont un outil de référence, montrant les contours d'une sélection de caractéristiques naturelles et artificielles de la Terre, agit souvent comme un cadre pour d'autres informations.

- **Carte thématique:** Ces cartes sont un outil pour communiquer des concepts géographiques tels que la distribution des espèces forestières (Ansari et al., 2008 in Zerroug, 2012)

), les cartes thématique sont importantes en matière du SIG, ils nous permettent choisir les combinaisons de couches, les intervalles de classe, les couleurs, les motifs et les symboles. (Brimicombe, 2001 in zerroug, 2012)

- **Composition de la carte :** La carte comporte une surface utile et un habillage :

- **La surface utile :** c'est où est représenté l'image carte

- **L'habillage :** comporte :

- **Titre:** fournit une description simple de la carte, il est très important car il est généralement la première chose à regarder sur une carte. Il doit être court, mais donne au lecteur une première idée de ce que la carte représente.

- **légende :** c'est un élément standard dans la plupart des configurations. Il donne la signification des symboles et des couleurs utilisées dans la carte.

- **Échelle :** montre la relation entre les mesures de distance sur la carte et la distance au sol réelle. Les Barres d'échelle sont les plus courants et pratique. Les valeurs sont indiquées en unités de carte (mètres, pieds ou en degrés).

- **Toponymie :** c'est le nom des lieux, ou le nom des détails impératifs pour une bonne lecture de la carte

- **Orientation (Flèche du Nord) :** c'est une figure montrant les principales directions (Nord, Sud, Est et Ouest) Sur une carte, elle est utilisée pour aider les gens à s'orienter eux-mêmes quand ils utilisent la carte.

- **Cadre:** Le cadre de la carte est une ligne qui définit exactement les bords de la zone indiquée sur la carte.

- **Date :** indique quand la carte a été produite et la date de recueillement des données.

- la sortie est la conversion des données d'un produit cartographique sur papier ou sur écran (affichage des données sous différentes formes, cartes, tableaux ou graphes) (Beldjazia, 2009).



Figure 21 ; Présentation du logo du logiciel MapInfo [12].

## 11.2. Environnement de MapInfo :

MapInfo est classiquement composé d'une barre des menus et de barres d'outils variées selon les besoins. C'est à partir de la barre des menus que l'on accède à la plupart des fonctions de MapInfo (Rahal, 2005).

### 11.2.1. La barre d'outils Général :

Cette barre est utile pour naviguer sur une fenêtre carte et pour y afficher les informations (figure 22) (Rahal, 2005).

### 11.2.2 La structure des données au format MapInfo :

MapInfo est un logiciel qui structure les informations en tables. Une table est un ensemble de fichiers qui sont manipulés ensemble par le logiciel (figure 24) (Rahal, 2005).

## 11. logiciel Mapinfo :

### 11. 1. Définition :

Mapinfo est un logiciel de traitement de l'information géographique. Il permet la représentation de données statistiques sur des cartes mais comporte aussi un volet gestion de bases de données géolocalisées et programmation. Le fait qu'il ait été pionnier dans le domaine explique en partie sa position privilégiée dans un créneau de logiciels qui connaît actuellement une forte croissance. (Gérardet *al*, 2012)

Ce logiciel permet de créer, d'afficher, de modifier toutes formes d'informations géographiquement référencées. Il est utilisé pour :

- Créer des cartes détaillées afin d'améliorer les présentations et faciliter la prise de décision.
- Mettre en évidence des tendances de vos données qui autrement ne sont pas perceptibles.
- Réaliser des analyses spatiales dans un environnement bureautique.
- Opérations de superposition : superposition de couches thématiques.
- Traitements statistiques.
- Echange de données avec les autres SIG (ArcGis, ArcInfo, etc.) et les logiciels de (CAO/DAO), tel que « AUTO CAD ».
- Type de sortie : représentation cartographiques, valeurs numériques ou textuelles, histogrammes, graphiques etc.
- Disponibilité d'une bibliothèque de symboles cartographiques, de trames et légendes modifiables de façon interactive.
- Géocodage par adresse.
- Accès aux bases de données externes (Access, Excel, Oracle, SQL server, ..).
- Son langage de requête qui permet de mixer les entrées graphiques et non graphiques.
- Personnalisation avec *MapBasi* (Vendé-Leclerc, 2008).

Il existe quatre fonctions de base de ce logiciel :

- l'entrée des données : cela se fait de multiples façons, manuellement ou non, interactivement ou par lecture de fichiers ou de table d'information,
- la gestion des données : (alphanumériques et spatiales – Système de Gestion de Base de données relationnel (INFO),
- l'analyse des données (alphanumériques et spatiales),
-



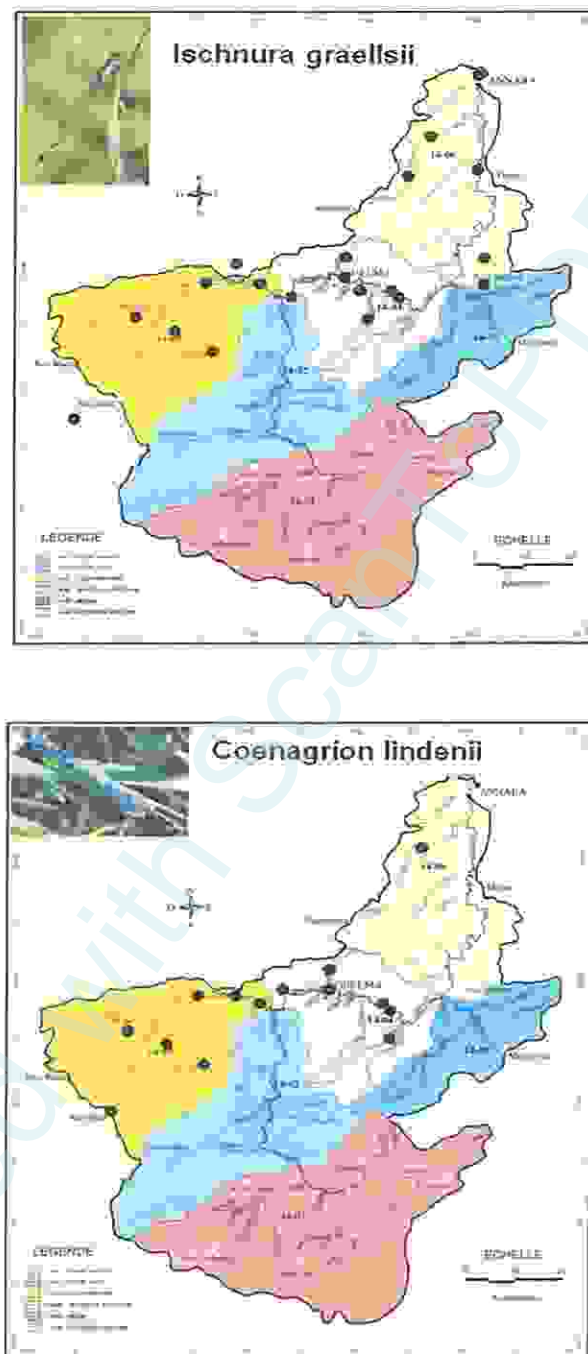


Figure 20 : Répartition spatiale de deux espèces d'odonate dans le bassin versant de Guelma (Satha, 2008).

radiotracking, la modélisation de la distribution des espèces, le suivi et l'évaluation de la biodiversité, du couvert végétal, de la nature des sols, des paysages... (eg. Moilanen & Hanski, 1998 ; Guinet et al.,2001 ; Tardif et al.,2005 ; Husté et al.,2006).

Si les données de base sont souvent disponibles sur la majeure partie du territoire, il reste des difficultés importantes dans l'obtention des données nécessaires à une gestion élaborée du paysage, tant pour ce qui concerne les objets paysagers (par exemple: assolement à l'échelle du parcellaire) que pour les facteurs d'évolution des paysages[9].

Les variables permettant de décrire la structure spatiale du paysage sont bien connues. Il s'agit essentiellement, des formes du relief, des textures de la végétation, des structures d'organisation (voirie, limites parcellaires, cours d'eau...), des objets caractéristiques spécifiques d'un paysage (arbres isolés, murets de terrasse, clôtures, édifices...). Les structures d'organisation peuvent être décrites sous forme linéaire: haies, cours d'eau. Les calculs de distance permettent de rendre compte d'un effet de proximité à une structure. Pour les routes, les calculs de distance dans un réseau (mode vecteur) ou selon une distance coût (mode raster) permettent de rendre compte de l'accessibilité au paysage. Enfin, les données ponctuelles (mares, édifices, arbres isolés...) peuvent faire l'objet de calculs de densité dans un voisinage défini [9].

### 10.1. Les outils utilisés pour cartographier les stations forestières:

Pour cartographier les stations forestières, certains outils sont utiles, voire indispensables, mieux connaître ces outils permet de définir leurs intérêts mais aussi de connaître leurs limites. Les outils ont été classés en trois grandes catégories :

- Ceux qui sont utilisés lors d'un travail préliminaire (avant d'aller sur le terrain). Ils permettent de rassembler des informations, de choisir la méthode de cartographie la plus adaptée et de préparer la phase de terrain.
- Ceux qui servent sur le terrain ; il s'agit d'outils au sens propre (tarière pédologique, couteau, fiole d'acide chlorhydrique, . .) ou au sens figuré (flores, guides botaniques...).
- Ceux qui facilitent la restitution de l'image et aident à la rédaction de la carte (fonds topographique de carte existants, logiciels spécifiques d'aide à la cartographie...).

Il existe, aussi, des outils spécifiques d'une ou plusieurs régions naturelles qui peuvent être utilisés à toutes les phases du projet; ce sont les catalogues des stations ou les guides simplifiés pour le choix des essences (Beldjazia, 2009).

- \* La représentation graphique de ces données (système de cartographie assistée par ordinateur; dessin assisté par ordinateur) ;
- \* L'analyse de ces données par des requêtes attributaires dans la base de données et leur visualisation graphique ;
- \* L'analyse de ces données par des requêtes spatiales ;
- \* L'édition des résultats de l'analyse sous forme cartographique.

#### 9. Liste de principaux logiciels de (SIG) :

- **MapGuide** : serveur cartographique, récemment offert à la communauté OpenSource par Autodesk.
- **Openmap** : permet de développer des applets à intégrer dans des sites web basés sur les Javabeans.
- **OrbisGIS** : OrbisGIS est capable d'afficher, de manipuler et de créer des données spatiales vecteur et raster. Il est complètement fait en Java et, pour ça, est multiplateforme.
- **QGIS** : logiciel de cartographie basé sur la bibliothèque Qt. Il est disponible sous Linux (KDE), Mac OS X, ou Windows. Il permet notamment la visualisation "à la volée" des couches de données comme des shapefiles ainsi que leur modification. Il permet aussi de produire des fichiers à publier sur MapServer. Son ergonomie le rend simple à utiliser.
- **SAGA GIS** : logiciel SIG sous GPL pour Windows et Linux.
- **uDig** : logiciel SIG développé en Java, utilisé pour le développement d'applications tierces.
- **Virtual Terrain Project** : logiciel SIG 3D.
- **MapInfo Professional, Spectrum Spatial, MapXtreme, Engage-3D, Encom Discover** de Pitney Bowes Software (PBS)[11].

#### 10. Usages du SIG en écologie:

L'utilisation des Systèmes d'Information Géographiques (SIG) dans le domaine de l'écologie s'est considérablement développée durant les dix dernières années. Les applications SIG destinées à la gestion de paramètres environnementaux sont multiples et illimitées. Ainsi, chaque spécialité peut intégrer avec succès la dimension géographique, comme l'inventaire des espèces, le suivi des espèces par satellite, calculer des déplacements issus d'études de

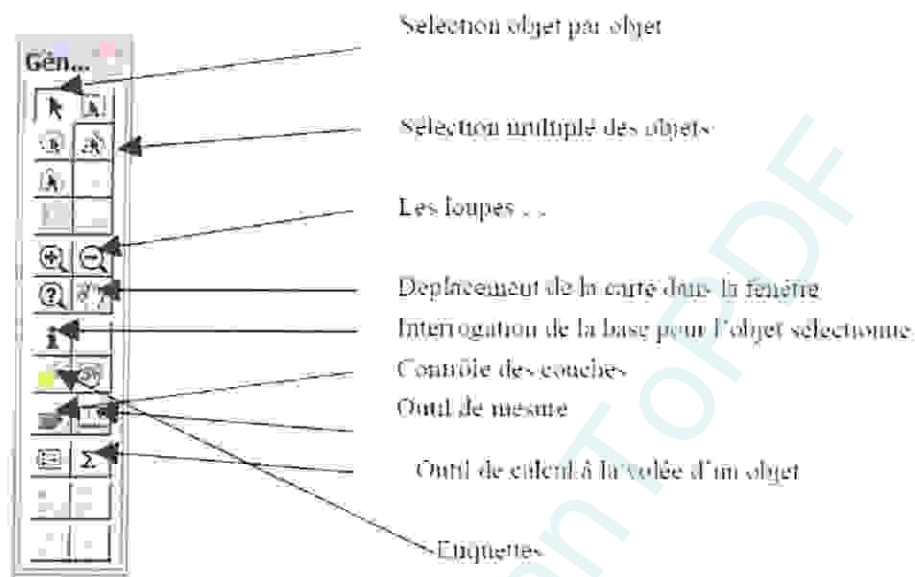


Figure 22: La barre d'outils Général de mapinfo(Rahal, 2005)

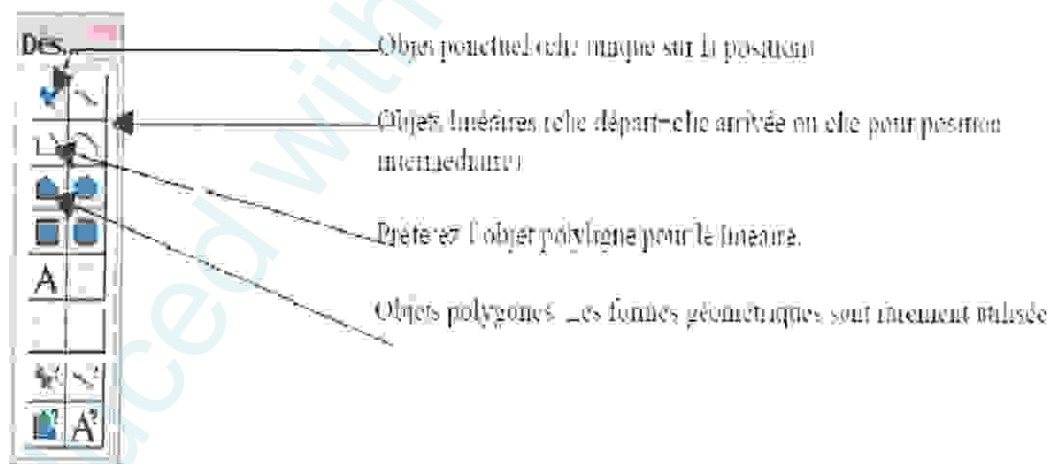


Figure23 : La barre d'outils Dessin de mapinfo(Rahal, 2005)

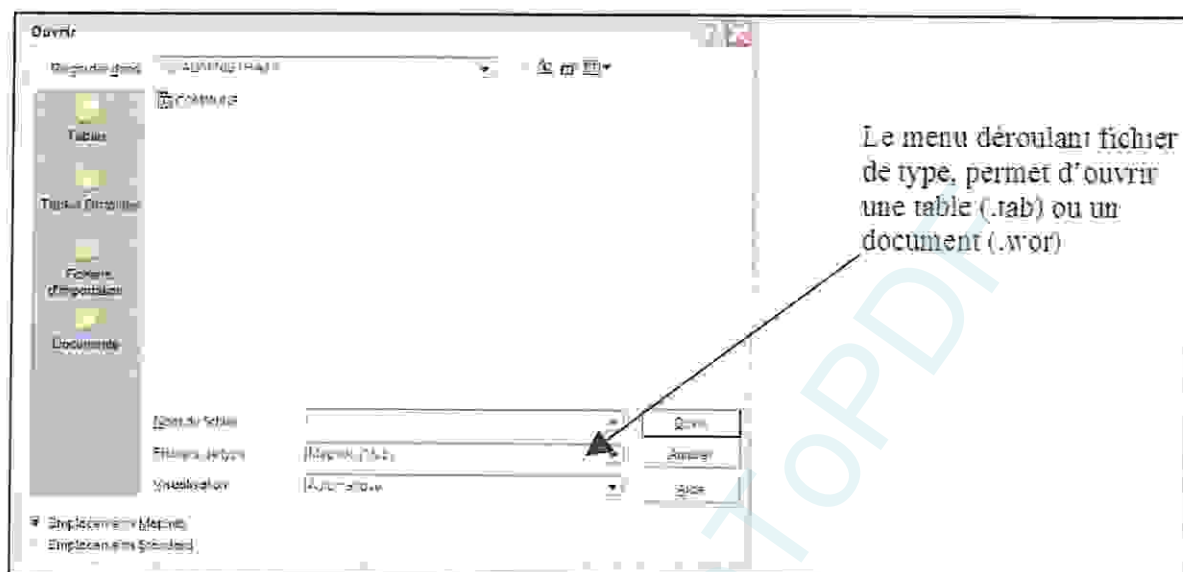


Figure 24: Menu Fichier ouvrir.

Une table est constituée de plusieurs fichiers liés entre eux :

\*.TAB: décrit la structure de la table MapInfo. C'est lui qui est appelé par le menu «ouverture table » et qui se charge de lister l'ensemble des fichiers nécessaires à son ouverture. C'est un répertoire d'adresses.

\*.DAT: contient les données tabulaires/attributionnaires. Il peut être remplacé par importation : par des fichiers Excel (\*.xls), Dbase (\*.dbf), Access (\*.mdb), ...

\*.MAP: C'est le fichier qui décrit les objets graphiques (nature, localisation ...).

\*.ID: C'est le fichier de références croisées entre les données tabulaires/attributionnaires et les objets cartographiques.

\*.IND: C'est un fichier d'index qui permet d'accélérer les requêtes sur la table.

\*.BMP, \*.TIF, etc. : Les données tabulaires peuvent être remplacées par des images raster.

\*.WOR: C'est le fichier qui enregistre l'espace de travail c'est à dire toutes les tables et leurs fichiers ouverts ainsi que l'ensemble des couches temporaires (requêtes, dessins, étiquettes). (Rahal, 2005)

### 11.2.3. Le contrôle des tables (ou des couches) :

Bien comprendre comment s'organisent les couches les unes par rapport aux autres et comment s'organise l'articulation entre la donnée géographique et la table attributaire qui lui est associée, est la base de la bonne compréhension des S.I.G. Une fois la ou les tables



11.2.5.Calculs simples à partir de la fenêtre carte :

Avec l'outil flèche noire, on effectue une sélection d'un ou plusieurs objets. On clique ensuite sur l'outil Somme. Un tableau s'affiche et propose les moyennes et sommes pour chaque champ (Figure 27)(Rahal, 2005).

Autres calculs :

On peut réaliser d'autres calculs pour interroger la base à partir du menu Sélection>statistiques.

Choix dans les menus déroulants de :

1/ la table ou la sélection

2/ Le champ sur lequel on souhaite les statistiques.

On obtient ensuite les statistiques sur le champ population dans l'exemple ci-contre (Figure28) (Rahal, 2005).

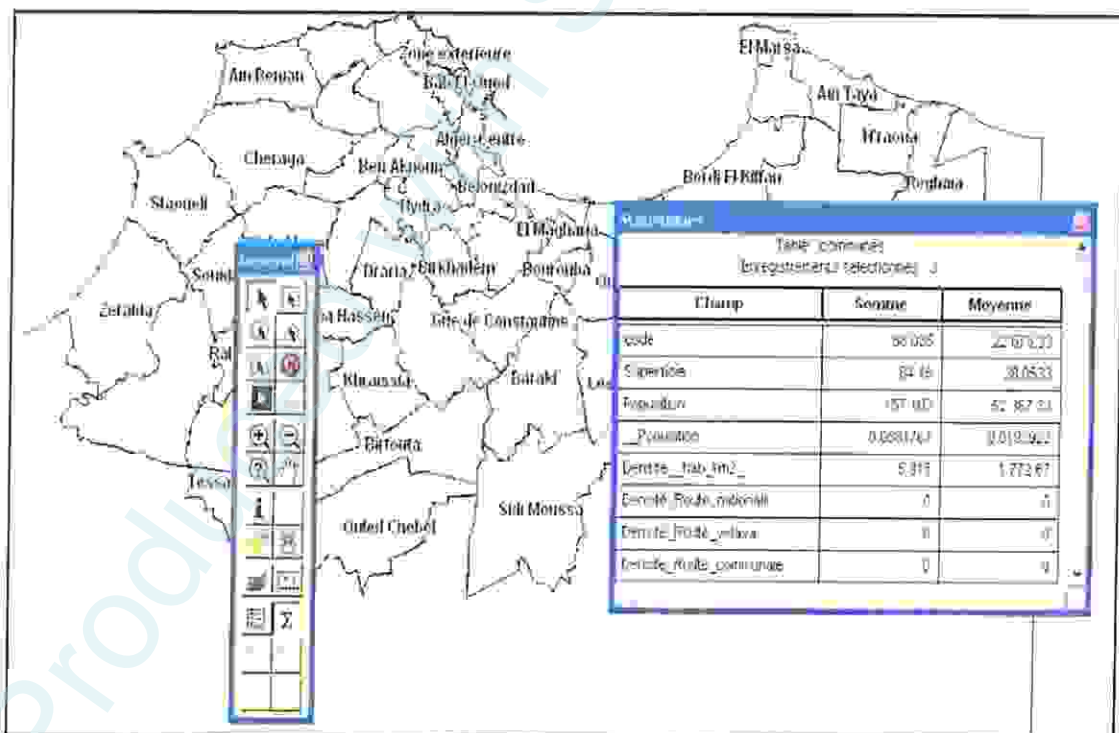


Figure27: L'outil somme de MapInfo.

zone d'étude (figure 30) (Beldjazia, 2009).

La deuxième étape consiste à choisir une nouvelle table pour dessiner la limite de la

et on obtient donc une carte géo référencée (Figure 29) (Beldjazia, 2009).

Une valeur d'erreur égale à zéro indique que la position des points de calage semble correcte ; faut rentrer les coordonnées des points de calage que nous avons déterminés préalablement. topographique ; pour cette étape il faut préciser la projection puis l'unité appropriée. Après il

Pour limiter la zone d'étude, on doit premièrement caler l'image scannée de la carte

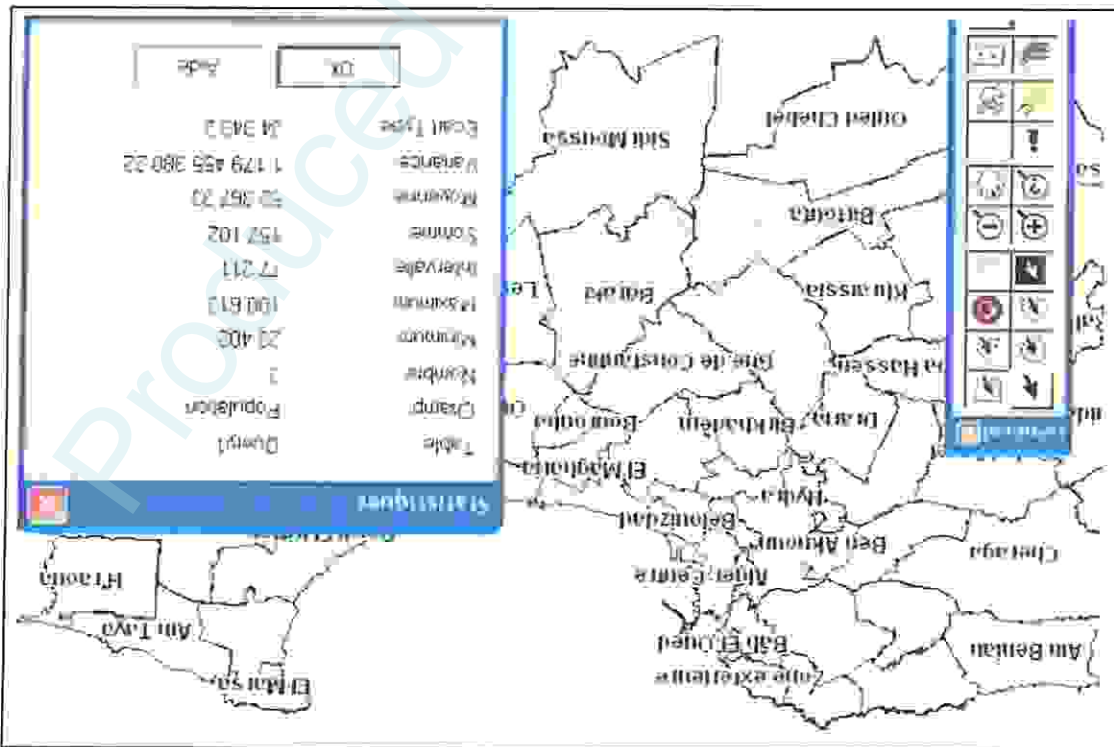
logiciels de SIG.

Pour cette raison, on a utilisé le logiciel MapInfo Professional 8.0 qu'est l'un des

✓ La limite de la zone d'étude :

#### 11.2.6. Application en biologie (Model):

Figure 28: L'outil statistique de MapInfo.





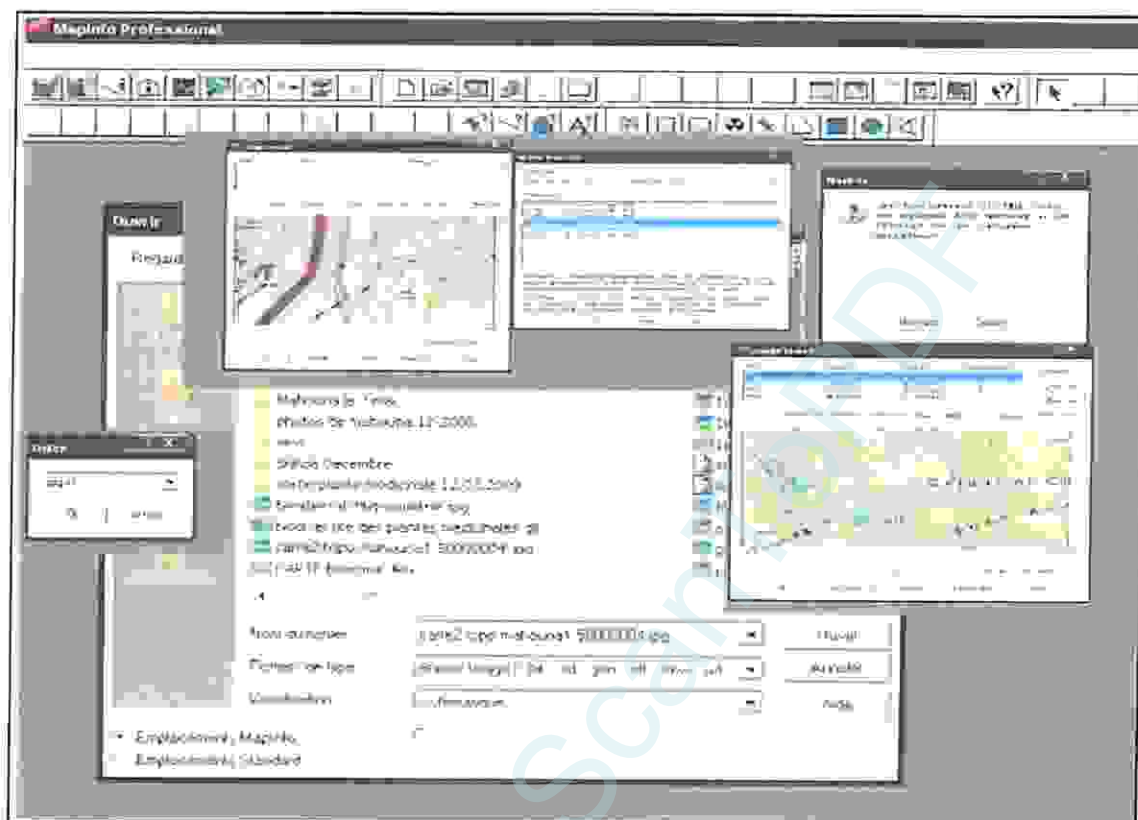


Figure 29 : Calage de la carte topographique par le MapInfo 8.0

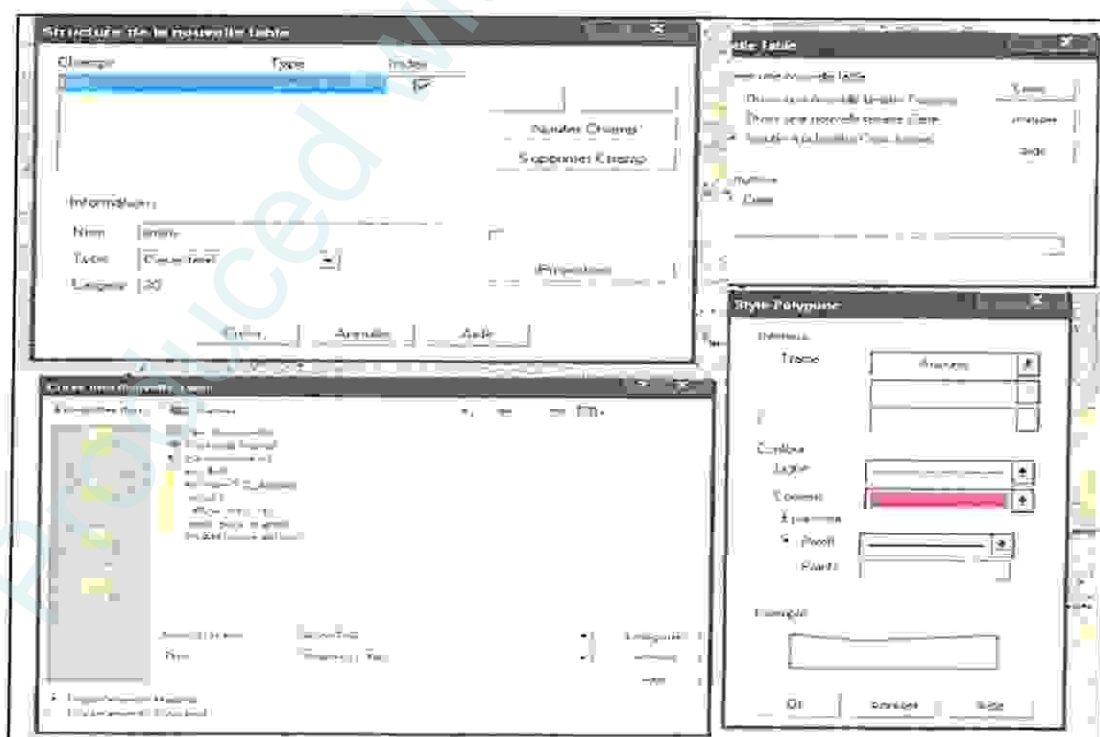


Figure 30 : Méthode de dessin de la limite de la zone d'étude de la Mahoung.

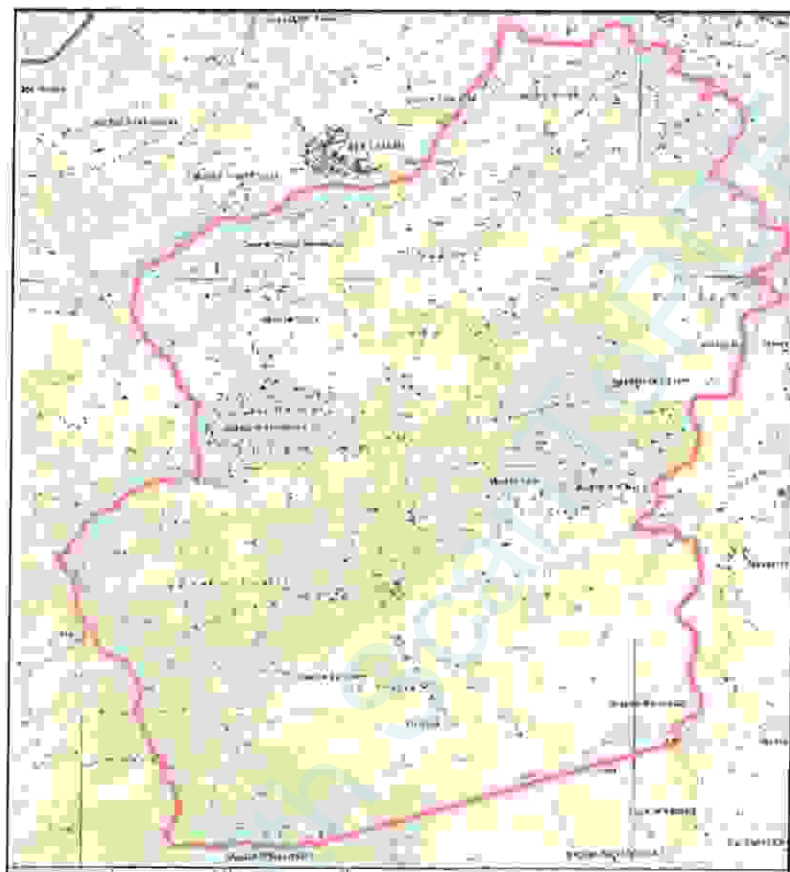
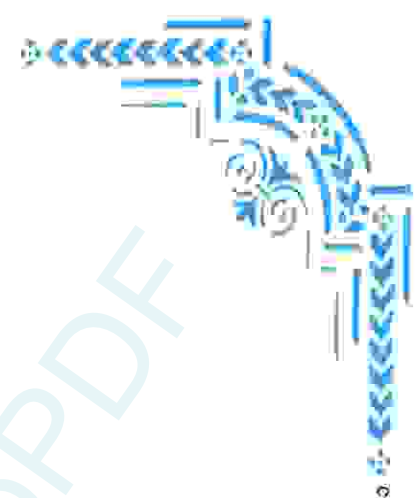
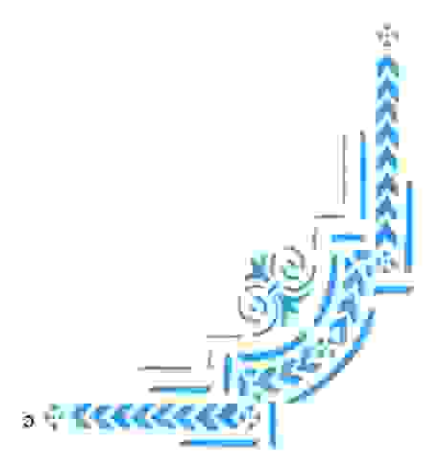


Figure 31 : Localisation de la zone d'étude dans la carte topographique de Guelma (feuille nj-32-II-5 Est) (Echelle 1/50 000 ème)(Beldjazia, 2009)



# *Chapitre III*

## *La bioinformatique*



Produced with ScanTOPDF

### 1. Définition :

C'est l'approche *in silico*, qui vient compléter les approches classiques *in situ* (dans le milieu naturel), *in vivo* (dans l'organisme vivant) et *in vitro* (en éprouvette) de la biologie traditionnelle [13].

La bioinformatique correspondait à l'utilisation de l'informatique pour stocker et analyser les données de la biologie moléculaire. Cette définition originale a maintenant été étendue et le terme bioinformatique est souvent associé à l'utilisation de l'informatique pour résoudre les problèmes scientifique poses par la biologie dans son ensemble. Il s'agit dans tous les cas d'un champ de recherche multidisciplinaire qui associe informaticiens, mathématiciens, physiciens et biologistes (Beroud, 2011).

Les deux thèmes majeurs de la bioinformatique sont :

- Identifier les gènes et prédire leur fonction .
- Prédire le repliement des séquences protéiques avec identification de l'activité biologique. (Ali Chaibou et al., 2009)

Pour cela la bioinformatique utilise plusieurs méthodes, dont la plus compréhensible est la comparaison de séquences . L'exemple du scénario présenté illustre la diversité des étapes nécessaires pour atteindre le but fixé. (Ali Chaibou et al., 2009)

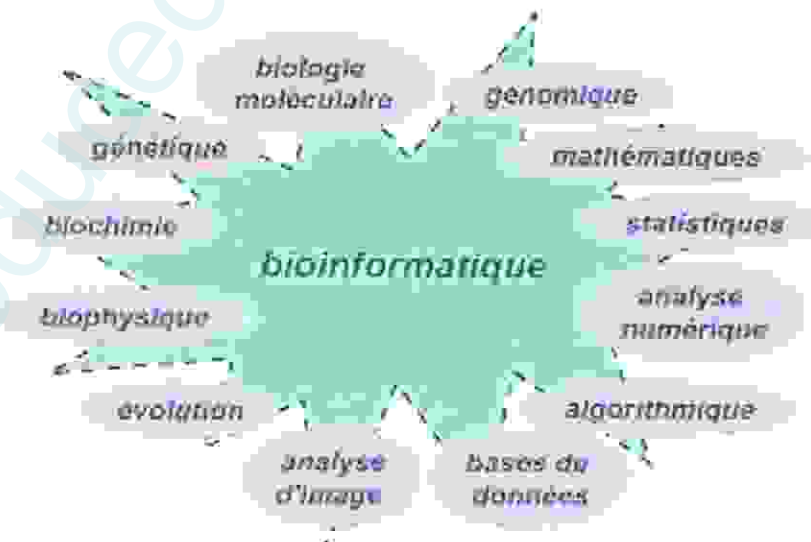


Figure 32 : La bioinformatique multidisciplinaire [13].

## 2. Les banque de données :

Elles permettent de stocker des données afin de pouvoir y accéder aisément dans le futur est un des buts premiers de la bioinformatique. Ces données sont essentiellement le résultat du séquençage des génomes, duquel découle une énorme quantité de données qu'on ne peut conserver sur papier d'une part à cause de la place que cela prendrait d'autre part, du fait qu'il serait difficile d'y chercher et de traiter les enregistrements. Les bases de données sont donc des outils indispensables. (Ali Chaibo *et al*, 2009).

Souvent les termes de banque ou base sont utilisés sans distinction particulière. Toutefois il existe une différence non seulement pour l'utilisateur mais aussi pour l'implantation informatique de ces dernières :

- **Base de données** : ensemble de données organisé en vue de son utilisation par des programmes correspondant à des applications distinctes et de manière à faciliter l'évolution indépendante des données et des programmes.
- **Banque de données** : ensemble de données relatif à un domaine défini des connaissances et organisé pour être offert aux consultations d'utilisateurs. (Karlin *et al*, 1990)

### 2.1. Les différents types de base de données :

Il existe un grand nombre de bases de données d'intérêt biologique. D'une façon générale, on distingue les banques ou base de données généralistes et les banques de données spécialisées. (Ali Chaibo *et al*, 2009)

#### 2.1.1. Les banques généralistes :

Cette section présente les principales collections généralistes de séquences nucléotidiques et protéiques ainsi que les centres de saisie qui leur sont associés.

##### 2.1.1.1 Séquences nucléiques :

Trois banques de séquence nucléiques :

- **EMBL** : banque européenne créée en 1980 et financée par l'EMBO (European Molecular Biology Organization), elle est aujourd'hui diffusée par l'EBI (European Bioinformatics Institute, Cambridge, UK).
- **GenBank** : créée en 1982 par la société IntelliGenetics et diffusée maintenant par le NCBI (National Center for Biotechnology Information, Los Alamos, US)
- **DDBJ** : créée en 1986 et diffusée par le NIG (National Institute of Genetics, Japon). Ces trois banques s'échangent systématiquement leur contenu depuis 1987 et ont adopté un système de conventions communes. (Karlin *et al.*, 1990)

#### 2.1.1.2. Séquences protéiques :

Deux banques protéiques :

- **PIR-NBRF** : créée en 1984 par la NBRF (National Biomedical Research Foundation). Elle est maintenant un ensemble de données issues du MIPS (Martinsried Institute for Protein Sequences, Munich, Allemagne) et de la banque japonaise JIPID (Japan International Protein Information Database).
- **SwissProt** : créée en 1986 à l'Université de Genève et maintenue depuis 1987 dans le cadre d'une collaboration, entre cette université (via ExPASy, Expert Protein Analysis System) et l'EBI. Celle-ci regroupe aussi des séquences annotées de la banque PIRNBRF ainsi que des séquences codantes, traduites de l'EMBL. (Karlin *et al.*, 1990)

#### 2.1.1.3. Les banques spécialisées :

Elles ont pour but de recenser des familles de séquences autour de caractéristiques biologiques comme les gènes identiques issus d'espèces différentes. Elles peuvent aussi regrouper des classes spécifiques de séquences comme les vecteurs de clonage ou toutes les séquences d'un même génome. . (Ali Chaïbo *et al.*, 2009).

Il existe plus de 250 bases de données d'intérêt biologique spécialisées :

-**Bases de motifs nucléiques** : Base de facteurs de transcription TFD (Transcription Factor Database), IMD (Information Matrix Database ),

-**Bases de motifs protéiques** : Prosite : elle peut être considérée comme un dictionnaire qui recense des motifs protéiques ayant une signification biologique.

-**Bases de structure** : PDB (Protein Data Bank)

-**Base d'expression**: YPD (Yeast Proteome Database) , MGED ( Microarray Gene Expression Data )

-**Bases de voies métaboliques**: KEGG (Kyoto Encyclopedia of Genes and Genomes)

-**Bases de cartographie**: GDB (Genome Data Base) MGD ( Mouse Genome Database )

-**Base de phénotypes** : MIM (Mendelian Inheritance in Man), MIA (Molecular Information Agent) (Ali Chaibo *et al.*, 2009).

## 2.2. Diffusion et utilisation des banques de données :

### 2.2.1. Diffusion :

Les bases de données sont mises à jour plusieurs fois par an. Pendant longtemps, le mode de distribution était l'envoi postal. Depuis 1990 et le développement des réseaux informatiques à haut débit, un grand nombre de bases sont stockées sur des serveurs publics. Ainsi beaucoup de serveurs mettent à disposition de nombreuses bases avec une mise à jour quotidienne des données (comme la banque EMBL) (Adleman, 1994)

### 2.2.2. Interrogation des banques de données :

Il existe deux types de logiciels pour que les utilisateurs puissent extraire les informations qui les intéressent. Les premiers sont des logiciels déjà programmés (comme le logiciel ACNUC ou SRS). Par contre, les deuxièmes sont des programmes établis à l'aide de systèmes de gestion de bases de données (SGBD) qui utilisent un langage de requête (Adleman, 1994)

### 3. Les méthodes utilisées par la bioinformatique :

La bioinformatique recourt à trois méthodes complémentaires :

#### 3.1 Méthode comparative :

Des logiciels de comparaison explorent les bases de données (séquences, structures) de gènes et de protéines déjà identifiés et annotés par d'autres équipes scientifiques pour établir des rapprochements (similarités, homologues ou identités) avec des séquences ou structures inconnues que l'on étudie et que l'on cherche à identifier [14].

#### 3.2. Méthode statistique :

Des logiciels appliquent des analyses statistiques aux données (sur la syntaxe des séquences) pour tenter de dégager et de repérer des règles et des contraintes présentant un caractère systématique, régulier ou général [14].

#### 3.3. Approche par modélisation :

Approche probabiliste. Elle consiste à étudier les objets (ex. : séquences, structures, motifs, etc., ...) à travers la construction d'un modèle qui tente d'en extraire les propriétés communes. La relation entre les objets d'étude (et/ou leur reconnaissance) est alors exprimée en référence à ce modèle optimal commun [14].

### 4. Les domaines de la bioinformatique :

Les domaines d'études en « ome » ou « omiques » le plein essor de la bioinformatique (telle qu'on la conçoit maintenant) a commencé avec la génomique (début des années 80) marquée par des événements phares comme le séquençage du génome humain.

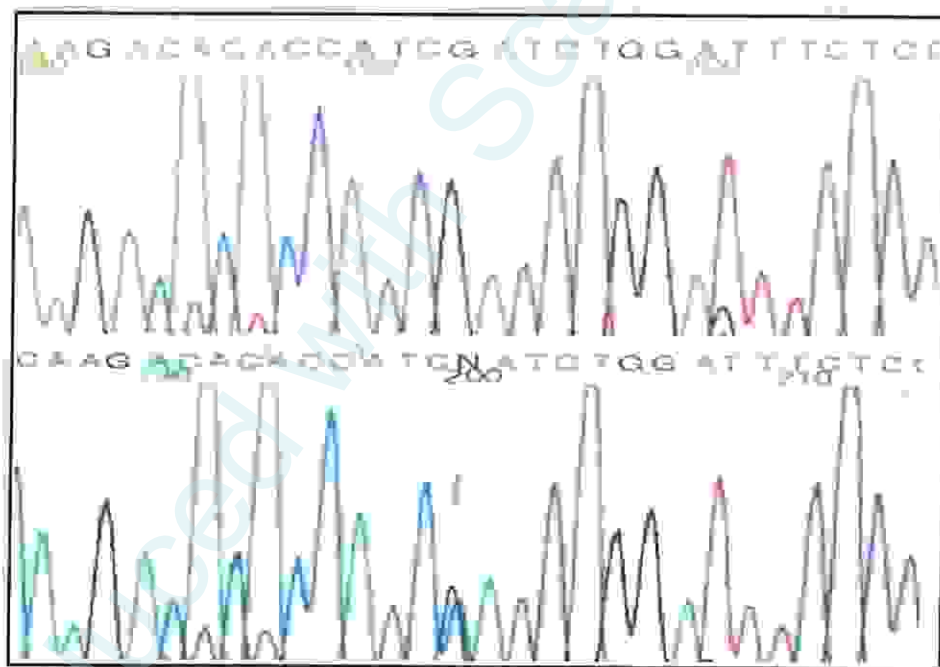
Puis la quantité et la multiplicité d'informations biologiques obtenues de même que les avancées technologiques ont fait apparaître au fur et à mesure d'autres domaines d'étude en « ome » ou « omique » [14].

#### 4.1. La génomique :

La génomique est une nouvelle discipline de la biologie qui vise à l'analyse moléculaire et physiologique complète du matériel héréditaire des organismes vivants. Il s'agit



déduire les fonctions des gènes et leurs interactions à partir de leurs séquences, ce qui facilite l'intégration de la génomique dans la physiologie. La génomique repose sur des techniques qui évoluent à une vitesse prodigieuse et l'ensemble des bases de données obtenues, qui représente des dizaines de milliards de caractères répartis en dizaines de milliers de fichiers, pose déjà aux informaticiens des problèmes considérables pour classer et interpréter cette énorme masse de données. la génomique est l'étude exhaustive des génomes et en particulier de l'ensemble des gènes, de leur disposition sur les chromosomes, de leur séquence, de leur fonction et de leur rôle. La cartographie physique est le positionnement de repère sur le génome. Le séquençage : pour connaître les « instructions » que renferme un fragment d'ADN, on lit la succession des bases puriques et pyrimidiques (A,T,C,G) de l'enchaînement. cette lecture est appelée séquençage (Figure 33) [15].



**Figure 33** : Illustration du résultat de séquençage [15].

#### 4.2. La protéomique :

La protéomique étudie le protéome : ensemble des protéines identifiées à partir d'un génome. Tout les cellules de l'organisme possèdent le même génome, mais ont un protéome différent selon l'organe et le moment de développement de l'individu. La protéomique s'attache

à déterminer la localisation, la structure et la fonction de ces protéines. Elle analyse leurs interactions et leurs modifications au cours du temps (Figure 34).

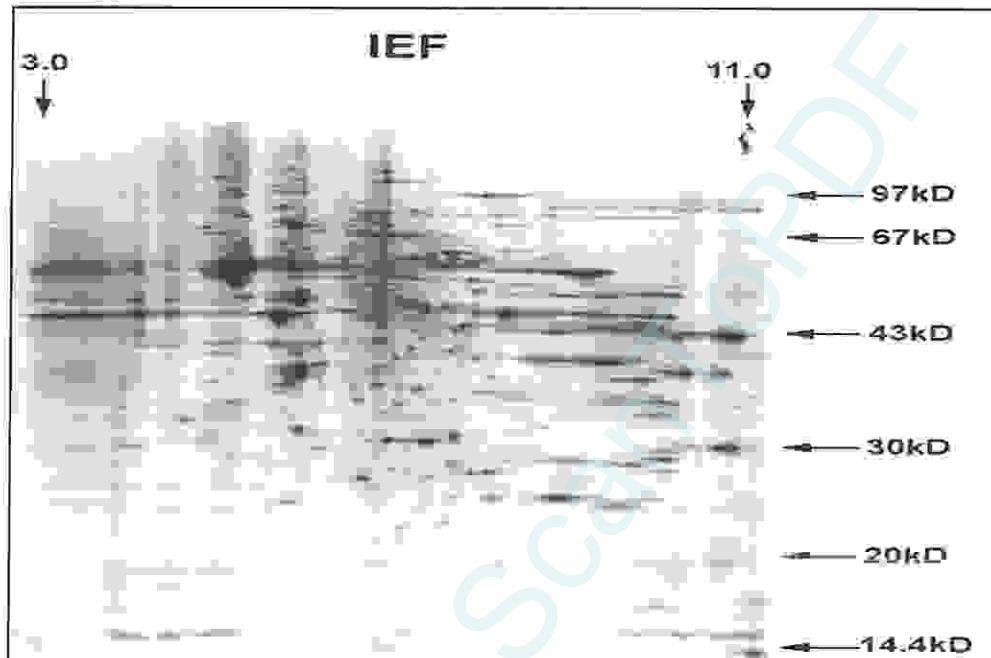


Figure 34 : Electrophorèse en 2D pour l'analyse protéique [16].

#### 4.3. Les autres « omes » et « omique » :

Métabolome – métabolomique : description de la population des différents métabolites d'une cellule dans un stade physiologique donné.

Interactome – interactomique : description des interactions entre toutes les macromolécules d'une cellule. on trouve aussi dans la littérature : lipidome ,protéasome , régulome , spmiceosome , prédictome ,réactome , glycomique [16].

### 5. Les applications de la bioinformatique :

#### 5.1. La santé :

Avec le traitement des maladies génétiques et des maladies multifactorielles, comme le diabète ou l'obésité : la sélection de gènes candidats, assistée par l'exploitation intelligente des informations présentes dans les banques de données publiques ou propriétaires, peut

accélérer l'identification des gènes responsables de pathologies complexes et par suite le développement de thérapies innovantes et adaptées aux individus. (Ali Chaibo *et al.*, 2009)

### 5.2. L'industrie pharmaceutique :

Avec la découverte de nouvelles cibles thérapeutiques et la conception rationnelle de médicaments et connaissances la structure exacte des protéines, les concepteurs de médicaments élaborent des molécules qui s'adaptent aux sites protéiques et les activeront ou les empêcheront d'agir (Ali Chaibo *et al.*, 2009)

### 5.3. L'industrie agro-alimentaire :

Les connaissances sur la structure tri-dimensionnelle de certaines protéines permettent de concevoir des ligands et autres molécules capables interagir avec ces macro-molécules et par exemples d'inhiber une fonction nocive ou réduire sa fréquence d'apparition ou améliorer le rendement ou la qualité (Ali Chaibo *et al.*, 2009)

### 5.4. L'écologie :

Pour la lutte contre la pollution des eaux et la pollution chimique ou biologique, les risques d'utilisations des cosmétiques, la préservation de la biodiversité, et la gestion du milieu marin l'étude la structure et du fonctionnement des écosystèmes Marins l'apport des modèles marins à la compréhension de l'évolution des fonctions biologiques et les nouvelles sources d'énergie (Ali Chaibo *et al.*, 2009)

### 5.5. Les biopuces :

Le concept de biopuce ou puce à ADN remonte au début des années 1990. Il repose sur une technologie pluridisciplinaire intégrant la micro-électronique, la chimie des acides nucléiques, l'analyse d'images et comme brique de base la bioinformatique. le principes de fonctionnement de ces puces repose sur le phénomène de l'appariement par complémentarité des bases se deux séquences d'ADN. Ceci permet d'identifier une séquence de nucléotides, c'est-à-dire l'enchaînement des bases d'un fragment d'ADN en mettant ce dernier en présence d'autres brins d'ADN, dont la séquence est connue. Par exemple, face à des brins d'ADN synthétiques représentatifs d'une maladie, les brins extraits de l'ADN du patient vont s'apparenter si le malade est porteur de l'affection recherchée. Cette technologie permet de

mesurer l'expression de plus de 40 000 gènes dans une cellule ou un tissu et de comparer leur expression entre différentes conditions (normal / pathologiques, traité/non traité, cinétiques temporelles, ...) (Ali Chaibo *et al.*, 2009)

#### 5.6. La guerre bactériologique ou chimique :

En déterminant par avance les modifications du fonctionnement génétique des cellules immunitaires occasionnées par des agents toxiques, on peut identifier très rapidement et à un coût bas les produits chimiques (mercure, dioxine ...) ou bactériologiques (bacille du charbon ou de la diphtérie ...) disséminés par un éventuel agresseur. (Ali Chaibo *et al.*, 2009)

#### 5.7. Les molécules darwiniennes :

Faisant appel aux principes de la sélection naturelle, un logiciel engendre des molécules-molécules inédites, futures candidates au statut de médicament. Plusieurs modèles chimiques permettent déjà de trouver informatiquement de nouvelles molécules actives. Le dernier logiciel en date propose d'appliquer la sélection naturelle darwinienne sur des modèles de molécules. Il part d'une molécule fournie par un laboratoire, connue pour son action sur certaines maladies. Il en construit ensuite jusqu'à trente « parents », des molécules élaborées automatiquement selon des critères géométriques, physico-chimiques ou autres. (Ali Chaibo *et al.*, 2009).

A chaque molécule est associée un « chromosome » fictif représentant ses caractéristiques. L'algorithme déclenche alors un processus de reproduction sur cette population de chromosome, en autorisant les mutations (changement actuel et aléatoire) et les échanges d'une séquence entre deux chromosomes. A chaque génération, le logiciel choisit les meilleurs reproducteurs, par exemple les molécules ayant la configuration la plus basse d'énergie.

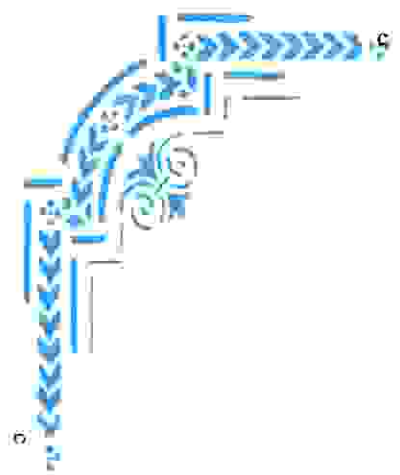
En jouant sur les paramètres de sélection, on peut aussi modéliser la fixation des ligands sur une protéine, ou encore prédire la forme exacte d'une molécule d'après sa formule chimique ou un temps record. (Ali Chaibo *et al.*, 2009)

### 5.8. Le domaine anténatal :

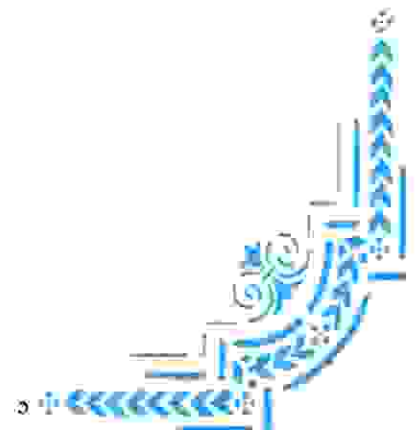
Le dépistage anténatal se fait par analyse des chromosomes placentaire ou du liquide dans lequel baigne l'embryon et donc indirectement de l'embryon.

Cette analyse va permettre de repérer très tôt toute anomalie grave pouvant toucher le futur bébé. En effet, il est aujourd'hui possible d'étudier les structures des chromosomes en un temps réduit et avec un cout faible et de prédire les éventuels problèmes et malformations pouvant en découler.

En cas d'anomalie, il importe aux parents, avec l'aide du médecin, de prendre la décision de poursuite ou d'interrompre la grossesse. (Ali Chaibo *et al.*, 2009)



# Conclusion



**Conclusion :**

Ce travail a consisté en une étude des logiciels statistique, cartographique et bioinformatique, l'utilisation de ces logiciels très importante dans des vastes domaines comme biologie, mathématique, physique ...

Notre travail se divise en trois chapitres :

Dans le premier chapitre concerne à étudier un grand champ de statistique qui définie par la science dont l'objet est de recueillir, de traiter et d'analyser des données issues de l'observation de phénomènes aléatoires, c'est-à-dire dans lesquels le hasard intervient.

L'analyse des données est utilisée pour d'écrire les phénomènes étudiés, faire des prévisions et prendre des décisions à leur sujet. En cela, la statistique est un outil essentiel pour la compréhension et la gestion des phénomènes complexes.

Les données étudiées peuvent être de toute nature, ce qui rend la statistique utile dans tous les champs disciplinaires et explique pourquoi elle est enseignée dans toutes les filières universitaires, de l'économie à la biologie en passant par la psychologie, et bien sur les sciences de l'ingénieur.

Dans la deuxième chapitre on essaye de définir le système d'information géographique, on démontrant les fonctions et les étapes nécessaires à la mise en place, ainsi que les moyens fournis par cet outil en matière d'affichage et de production de cartes, nous abordons également le rôle et l'importance du SIG dans le domaine environnemental et écologique.

Dans le dernier chapitre pour objectif est donc de faire le point sur les apports de la bioinformatique notamment par les différentes bases de données et outils bioinformatiques qu'elle a permis de créer ces dernières années.

*Références  
bibliographiques*



## Références Bibliographiques

- Abdelbaki, C., Benhamouda, F., & Chikih, M. (2012). SIG: outils de gestion des réseaux d'alimentation en eau potable: cas du réseau de la ville de Birtouta-Alger, Algérie. Éditions universitaires européennes.
- Adleman, L. M. (1994). Molecular computation of solutions to combinatorial problems. *Nature*, 369, 40.
- Chaibou M, Bougueffa I, Douad H, Séridi A. (2009). La bioinformatique outil de la biologie moléculaire. université 08 mai 1945 guelma. Mémoire d'Ingénieur d'état. 47 p
- Beldjazia A, (2009). Etude écologique et cartographique de la végétation du massif de la mahouna (guelma). université Badji mokhtar annaba. Mémoire de magistère. 124 p.
- Denguerai, A., & Douabdeli, R. (2009). Contribution à l'élaboration d'un SIG pour une région steppique: wilaya de Naâma (Algérie).
- Beroud, C. (2011). Bases de données et outils bioinformatiques utiles en génétique. Université Médicale Virtuelle Francophone. 19 p
- Dagnelie, P. (1998). Statistique théorique et appliquée (Vol. 2). De Boeck Université. 736 p.
- De By A, De By R., Knippers R, Sun Y, et Ellis M, et Kraak., M, Weir, M, Georgiadou, Y, Radwan M, et Westen C, et Kainz W, et Sides E. (2000) Principles Of Geographic Information Systems, The International Institute for Aerospace Survey and Earth
- Fotheringham S., Rogerson P. (1994). Spatial analysis and GIS, Edited by CRC Press. (ISBN 10: 0748401040), p 80.
- Gilbert D, Manolo G. (2015) Bioinformatique cours et application. Dunod. Paris .203 p.
- Golmard, J. L., Mallet, A., & Morice, V. (2007). Biostatistique.
- Grégoire, G., Jollois, F. X., Petiot, J. F., Qannari, A., Sabourin, S., Swertwaegher, P., & Viguier-Pla, S. (2012). Les logiciels et l'enseignement de la statistique dans les départements" Statistique et Informatique Décisionnelle"(STID) des IUT. *Statistique et Enseignement*, 2(2), 5-24.

- Grégoire, G., Jollois, F. X., Petiot, J. F., Qannari, A., Sabourin, S., Swertwaegher, P., & Viguier-Pla, S. (2012). Les logiciels et l'enseignement de la statistique dans les départements "Statistique et Informatique Décisionnelle"(STID) des IUT. *Statistique et Enseignement*, 2(2), 5-24.
- Guinet, C., Dubroca, L., Lea, M. A., Goldsworthy, S., Cherel, Y., Duhamel, G., ... & Donnay, J. P. (2001). Spatial distribution of foraging in female Antarctic fur seals *Arctocephalus gazella* in relation to oceanographic variables: a scale-dependent approach using geographic information systems. *Marine Ecology Progress Series*, 219, 251-264.
- Hammoun, H., Bouzida R. (2010). *Pratique des Systèmes d'information géographiques (SIG)*. Pages bleues : Alger.196 p.
- HERVÉ, M. (2014). Aide-mémoire de statistique appliquée à la biologie. Constr Son Étude Anal Résultats à l'aide du logiciel R. Version, 5.
- Husté, A., Clobert, J., & Miaud, C. (2006). The movements and breeding site fidelity of the natterjack toad (*Bufo calamita*) in an urban park near Paris (France) with management recommendations. *Amphibia-Reptilia*, 27(4), 561-568.
- Karlin, S., Stephen, F. Altschul. (1990). Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl.Acad. Sci. USA* 87 :2264-68.
- Khallef, M., Liman, R., Konuk, M., Cigerci, I. H., Benouareth, D., Tabet, M., & Abda, A. (2015). Genotoxicity of drinking water disinfection by-products (bromofom and chlorofom) by using both *Allium* anaphase-telophase and comet tests. *Cytotechnology*, 67(2), 207-213.
- Kuhn, W. E. (1979). *Pour une histoire de la statistique*. Tome I. Insee, Imprimerie Nationale, 593 pp.
- Longley, P. (2005). *Geographic information systems and science*. John W & Sons.
- Maxam, M et Gilbert, W (1977). A New Method for Sequencing DNA. *Proceedings of the National Academy of Sciences of the USA*, vol. 74, p. 560-564.
- Millot, G. (2009). *Comprendre et réaliser les tests statistiques à l'aide de R*. De Boeck, Bruxelles.
- Moilanen, A., & Hanski, I. (1998). Metapopulation dynamics: effects of habitat quality and landscape structure. *Ecology*, 79(7), 2503-2515.

- Rahal, F. (2005), Modélisations et S.I.G pour l'aide à la gestion de l'environnement en milieu urbain. Etude du cas de la pollution atmosphérique au niveau de la ville d'Oran. Mémoire de Magister : Université d'Oran
- Robin, M. (2002). Télédétection, Des satellites au SIG. Une analyse complète du processus de création d'un type essentiel d'information géographique. Nathan Université, 318.
- Rolf, A., de Knippers, B. R. A., Sun, Y., Ellis, M. C., Kraak, M. J., Weir, M. J., ... & Sides, E. J. (2001). Principles of geographic information systems. The International Institute for Aerospace Survey and Earth Sciences (ITC), Hengelosestraat, 99.
- Satha Yalles A. (2008). Caractérisation du peuplement odonatologique des bassins versants de Bouhamdane et Seybouse. Univ.Guelma. Mémoire de Magister. 124p.
- Tardif, B., Lavoie, G., & Lachance, Y. (2005). Atlas de la biodiversité du Québec. Les espèces menacées ou vulnérables. Gouvernement du Québec, ministère du Développement durable, de l'Environnement et des Parcs, Direction du développement durable, du patrimoine écologique et des parcs, Québec.
- Tillé, Y. (2010). Résumé du Cours de Statistique Descriptive.
- Triola, M. Triola, M. F. (2012). Biostatistique pour les sciences de la vie et de la santé: édition revue et corrigée. Pearson Education France.
- Vendé-Leclerc, M. (2008). Prise en main du logiciel de SIG MapInfo, 37 p.
- Verbyla, (2002) .Practical GISAnalysis , First published 2002 by Taylor & Francis 11 New Fetter Lane, London EC4P 4EE (ISBN-10: 0415286093), 125 p.
- W. S. McCulloch et W. Pitts. (1943). A Logical Calculus of the Ideas Immanent i Nervous Activity. Bulletin of Mathematical Biophysics, vol. 5, p. 115-133.
- Wainwright, J., & Mulligan, M. (Eds.). (2005). Environmental modelling: finding simplicity in complexity. John Wiley & Sons.
- Weng, Q. (2010). Remote sensing and GIS integration: theories, methods, and applications (p. 416). New York: McGraw-Hill.
- Zerroug, K. (2012). Elaboration d'un système d'information géographique (flore) dans la Wilaya de Sétif. Université farhet abbas sétif. Mémoire de magister. 146 p.

• Références site web :

- [1] : <http://www.cours-gratuit.com/cours-statistique/support-de-formation-sur-les-parametre-d-une-serie-statistiques-a-une-variable>
- [2] : [https://fr.wikipedia.org/wiki/Liste\\_de\\_logiciels\\_de\\_statistiques](https://fr.wikipedia.org/wiki/Liste_de_logiciels_de_statistiques)
- [3] : <http://tecfaetu.unige.ch/etu-maltr/maya/tio/stic/ex5/decouverte.html>
- [4] : [http://www.biostatisticien.eu/springerR/livreR\\_presentation.pdf](http://www.biostatisticien.eu/springerR/livreR_presentation.pdf)
- [5] : <https://eric.univ-lyon2.fr/%l8jjacques/>
- [6] : [http://fsmarketing.free.fr/cours\\_SPSS/cours1.pdf](http://fsmarketing.free.fr/cours_SPSS/cours1.pdf)
- [7] : [http://www.icedd.be/17/index.php?option=com\\_k2&view=item&id=1217:les-installations-d%E2%80%99outils-sig-en-communes&Itemid=606&lang=fr](http://www.icedd.be/17/index.php?option=com_k2&view=item&id=1217:les-installations-d%E2%80%99outils-sig-en-communes&Itemid=606&lang=fr)
- [8] : <https://www.r-project.org/>
- [9] : <https://alexandreboissinot.jimdo.com/sig-et-%C3%A9cologie/>
- [10] : [http://www.memoireonline.com/07/08/1287/m\\_mise-en-place-d-une-plate-forme-de-cartographie-dynamique4.html](http://www.memoireonline.com/07/08/1287/m_mise-en-place-d-une-plate-forme-de-cartographie-dynamique4.html)
- [11] : [https://fr.wikipedia.org/wiki/Liste\\_des\\_logiciels\\_SIG](https://fr.wikipedia.org/wiki/Liste_des_logiciels_SIG)
- [12] : <http://www.obyga-engineering.com/uploaddir/upload/tools/mapinfo.jpg>
- [13] : [http://silico.biotoul.fr/enseignement/L3-Info/Introduction\\_bioinfo\\_L3info.pdf](http://silico.biotoul.fr/enseignement/L3-Info/Introduction_bioinfo_L3info.pdf)
- [14] : <http://fdanieau.free.fr/cours/master/Semestre1/RIni/cours1.pdf>
- [15] : <http://perso.ibcp.fr/gilbert.deleage/Cours/IBIS.pdf>
- [16] : [http://biochimej.univangers.fr/Page2/BIOINFORMATIQUE/7ModuleBioInfoJMGE/4IntroDefBioInfo/1Intr\\_DefBioInfo.htm](http://biochimej.univangers.fr/Page2/BIOINFORMATIQUE/7ModuleBioInfoJMGE/4IntroDefBioInfo/1Intr_DefBioInfo.htm)
- [17] : [http://www.pqs.ulg.ac.be/attachments/divers/Intro\\_info.pdf](http://www.pqs.ulg.ac.be/attachments/divers/Intro_info.pdf)