

République Algérienne démocratique et populaire
Ministère de l'enseignement supérieur
Et de la recherche scientifique
Université de Jijel

Faculté de science
Département d'Ecologie
et d'Environnement



جامعة جيجل
كلية العلوم
قسم البيئة والمحيط

Eco. 08/08

01
01

Mémoire

De fin d'Etudes en vue de l'Obtention du Diplôme d'Ingénieur d'Etat
en Ecologie Végétale et Environnement
Option : Pathologie des Ecosystèmes

Thème :

*Modélisation de l'évolution d'un
phénomène biologique, cas d'un
écosystème.*

Jury :

Président : M^r Bouljedri M.
Examineur : M^r Boudjlél F.
Encadreur : M^r Gherda M.

Présenté par :

Benmebarek Khadidja.
Boumedra Nadia.

Session : 2008-2009



Remerciement

• Avant tout, nous remercions Dieu le tout puissant de nous avoir aidé et éclairé notre chemin.

• Un grand remerciement à celui qui nous a consacré une partie de son temps et de sa compréhension pour ses conseils et ses critiques positives: notre encadreur:

M^{ieur} Gherda Melrouk.

• A tous les personnes du département d'écologie

• D'autre part, nous tenons à exprimer notre gratitude au président et aux membres du jury qui ont accepté de juger notre projet.

• A toutes les personnes qui ont, de près ou de loin contribué à la réalisation de ce mémoire.

LISTES DES FIGURES :

<u>Figures1</u> :Histogramme de H_1	55
<u>Figures 2</u> :Nuage de points ($\hat{\text{age}}_1, H_1$).....	55
<u>Figures 3</u> :Droite de régression linéaire de la hauteur sur l'âge.....	56
<u>Figures 4</u> : Histogramme de H_2	57
<u>Figures 5</u> : Nuage de points ($\hat{\text{age}}_2, H_2$).....	57
<u>Figures 6</u> : Droite de régression linéaire de la hauteur sur l'âge.....	58
<u>Figures 7</u> : Nuage de points de H_2 en fonction de $\log(\hat{\text{age}}_2)$	59
<u>Figures 8</u> : Droite de régression linéaire de la hauteur sur $\log(l'\hat{\text{age}}_2)$	60
<u>Figures 9</u> : Diagramme en boites de H_1 et H_2	61
<u>Figures 10</u> :Les régression linéaires.....	75

LISTES DES TABLEAU :

<u>Tableau 1:</u> L'évolution de H_1 en fonction de âge $_1$, pour un peuplement implanté en 1850...	54
<u>Tableau 2:</u> Statistique descriptive de la hauteur H_1	55
<u>Tableau 3:</u> Statistique descriptive de la hauteur H_2	57
<u>Tableau 4:</u> L'évolution de H_2 en fonction de âge $_2$, pour un peuplement implanté en 1950...	59
<u>Tableau 5:</u> Statistique descriptive de l'âge	61
<u>Tableau 6:</u> Tables des valeurs théoriques et observées pour P_2	62
<u>Tableau 7:</u> Tables des valeurs théoriques et observées pour P_1	63
<u>Tableau 8:</u> Tables des valeurs théoriques et observées pour P'_2	64
<u>Tableau 9:</u> Table des observations de la hauteur des deux peuplements pendant 55ans.....	65
<u>Tableau 10:</u> Données recueillies à 20 sites de l'étange de Thau le 25 oct. 1988.....	68

Liste des abréviations :

CEFE : Centre d'Ecologie Fonctionnelle Evolutive.

LASB : L'Analyse des Systèmes et Biométrie.

MCMC : Monte –Carlo par Chaîne Markov.

EDO : Equation Différentielle Ordinaire.

EDP : Equation Différentielle Potentielle.

CD : Cadmium.

Var(x) : la variance de x.

Me : Médiane de x.

R : coefficient de corrélation linéaire de Spearman.

S^2 : La variance.

Cov : Covariance.

ANOVA : Analyse de variance.

ANCOVA : Analyse de variance.

AFM : Analyse Factorielle Multiple.

ACP : Analyse en Composantes Principales.

ACM : Analyse Correspondances Multiple.

EMCO : Estimation des Moindres carrés Ordinaires.

X^2 : test de Chi-deux.

P_1 : Peuplement d'hêtre en 1850.

P_2 : Peuplement d'hêtre en 1950.

P'_2 : Peuplement d'hêtre en 1950 après transformation logarithmique.

H_1 : La hauteur du peuplement de 1850.

H_2 : La hauteur du peuplement de 1950.

Age_1 : l'Age du peuplement de 1850.

Age_2 : l'Age du peuplement de 1950.

Obs : Observées.

Théo : Théoriques.

U : Universités.

Fig. : Figures.

Bna : Bactéries aérobiques.

NH_4^+ : Ammoniac.

Var : variable.

Id : identités.

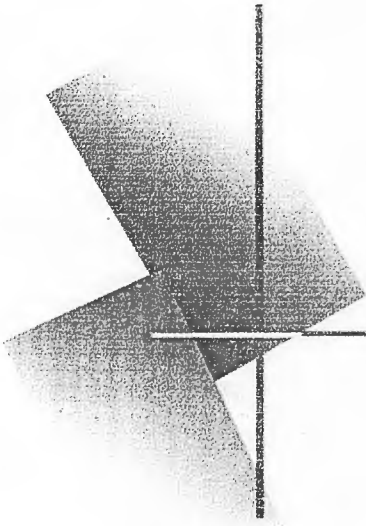
Prod- bac : production bactérienne.

Table des matières

0.1	INTRODUCTION :	4
1	Concepts biologiques	7
1.1	DYNAMIQUE DES POPULATIONS :	9
1.2	LES DIFFERENTS TYPES D'INTERACTIONS :	11
1.2.1	Symbiose :	11
1.2.2	Mutualisme ou coopération :	12
1.2.3	Commensalisme :	12
1.2.4	Neutralisme :	12
1.2.5	Parasitisme :	12
1.2.6	Prédation :	13
1.2.7	Concurrence :	13
1.3	PHENOMENES :	13
1.3.1	Phénomène déterministe :	13
1.3.2	Phénomène aléatoire :	15
1.4	MODELES ET MODELISATION :	17
1.4.1	Définitions :	17
1.4.2	Approches déterministes, stochastiques :	18
1.4.3	Comment modéliser ?	18
1.5	ETAPES COURANTES D'UNE MODELISATION	20

1.6	Test d'ajustement au modèle	22
1.6.1	Caractéristiques d'un modèle :	23
1.6.2	Construction d'un modèle statistique :	24
1.6.3	Interprétation du modèle :	25
1.6.4	Validation du modèle :	25
1.6.5	Conclusions :	26
2	Concepts mathématiques	27
2.1	ÉQUATIONS DIFFERENTIELLES :	27
2.2	Statistiques descriptives :	29
2.2.1	Objective :	29
2.2.2	Les séries simples :	30
2.2.3	Les séries multiples :	35
2.3	LES ANALYSES STATISTIQUES :	38
2.3.1	Régression linéaire :	38
2.3.2	ANOVA :	39
2.3.3	ANCOVA :	39
2.3.4	Régression logistique :	39
2.3.5	Analyse Factorielle Multiple (AFM) :	40
2.3.6	Analyse descriptive :	41
2.4	LES MATRICES :	42
2.4.1	Définitions	42
2.5	REGRESSION LINEAIRE MULTIPLE	44
2.5.1	Equation de régression et objectifs :	44
2.5.2	Notation matricielle :	45
2.5.3	Hypothèses :	45
2.5.4	La méthode des moindres carrés ordinaires :	46
2.6	LES TESTS STATISTIQUES	47

2.6.1	Introduction	47
2.6.2	Tests de corrélation :	48
2.6.3	Test de chi-deux (χ^2) :	49
2.6.4	Tests de comparaison :	49
2.7	AJUSTEMENT ET CORRELATION	51
2.7.1	Observation :	51
3	Applications numériques :	53
3.1	Application 1 :	53
3.2	Application 2 :	60
3.3	Application 3 :	68
3.4	Application 4 :	72
3.5	Conclusion	75



Introduction

0.1 INTRODUCTION :

La recherche en Biologie passe par une démarche expérimentale permettant le recueil des données de terrain par exemple en écologie ou dans le domaine des sciences de l'environnement. Les données de terrain sont ensuite analysées et traitées. Enfin, bien souvent, la modélisation mathématique est utilisée pour rechercher les mécanismes responsables de l'évolution et de la dynamique observée.

Les mathématiques interviennent à chacune des étapes précédentes. Les plans d'échantillonnage sont très utiles et permettent de déterminer les mesures qui vont être effectuées en fonction des questions biologiques posées. Cette étape est très importante pour éviter de réaliser des expériences coûteuses qui s'avèrent ensuite inexploitable. L'analyse des données permet d'extraire des informations des données de terrain et d'établir des relations entre les différentes variables.

L'étape ultime consiste à construire un modèle mathématique et/ou informatique qui a pour but de reproduire la dynamique observée et de l'expliquer. Les systèmes dynamiques constituent l'outil privilégié pour cette étape de modélisation mathématique et informatique.

On peut constater que dans ces dernières années, la part de la modélisation mathématique en Biologie n'a fait qu'augmenter. Aujourd'hui, la plupart des articles publiés par les grandes revues de Biologie présentent des modèles mathématiques. C'est particulièrement vrai dans le domaine de l'écologie et des sciences de l'environnement, où la modélisation mathématique apparaît comme une étape incontournable de la recherche en Biologie.

Deux risques majeurs, concernent la modélisation mathématique en Biologie :

- construire des modèles mathématiques sans biologie, c'est-à-dire ne s'appuyant sur aucune ou trop peu de données biologiques. Il s'agit de modèles qui ne sont jamais validés mais qui peuvent être utilisés pour faire des prédictions qui évidemment sont en général totalement inexactes.
- construire des modèles sans analyse mathématique. La floraison des logiciels de simu-

lisation numérique peut conduire certains chercheurs plutôt expérimentateurs à construire un modèle mathématique ayant pour but de représenter la dynamique du système qu'ils étudient. Il peut s'agir de modèles avec un grand nombre de variables et de paramètres.

-La communauté des mathématiciens peut sans doute et doit jouer un rôle important pour aider les biologistes dans la construction des modèles et pour l'étude analytique de ceux-ci lorsque celle-ci est possible.

-Ou pas du tout. Il y a des raisons assez fondamentales pour expliquer ce manque de contacts entre biologistes et mathématiciens :

- les mathématiciens considèrent trop souvent la modélisation mathématique en biologie comme une simple application sans grand intérêt mathématique, en tout cas ne relevant pas du tout des mathématiques pures mais plutôt du domaine de la Biologie elle-même. Certains voient même le risque d'être rejetés de leur communauté mathématique d'origine s'ils devaient pratiquer un peu trop une application plutôt que des mathématiques plus fondamentales.

- les biologistes voient dans la modélisation mathématique, en particulier lorsque l'on procède à des études analytiques des systèmes d'équations, des développements strictement mathématiques ne relevant absolument pas de leur discipline . Ils considèrent qu'il s'agit d'un exercice de mathématiques n'ayant aucune conséquence en Biologie(Auger.2000).

Description du mémoire :

Notre travail est composé en trois parties principales comme suit :

- La première est consacrée pour des rappels de quelques notions et définitions biologiques.

- La deuxième partie a essentiellement pour but d'exposer les différentes notions théoriques, dont la connaissance peut simplifier considérablement la présentation et faciliter la compréhension des méthodes statistiques utilisées.

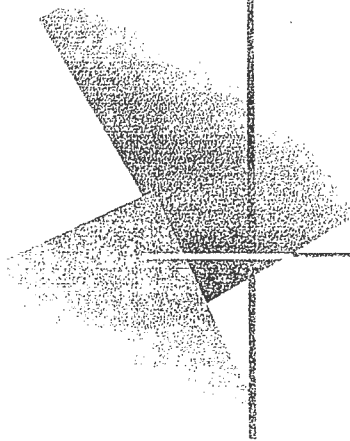
- Dans la troisième partie, nous avons essayé d'illustrer ce travail par des applications

numériques en utilisant quelques techniques mathématiques et des outils informatiques (langage R et workPlace).

Les données explorées sont inspirées de (des données écologiques recueillies à 20 sites de l'étang de thau (côte méditerranéenne de la france) dans le cadre du programme ECO-THAU (Amanieu *et all.*,). Les données, représentées au tableau 1 ci-dessous, sont tirées de Legendre et Legendre (tableau 13.2)-VertigO- La revue en sciences de l'environnement, Vol 6 no2, septembre 2005).

Chapitre I

Concepts Biologiques



On peut décomposer la modélisation en 5 étapes distinctes, décomposition que nous suivront dans l'ensemble des chapitres sur la modélisation :

Etape 1 :

- Définir le cadre d'étude et ses propriétés (analyse de la réalité).
- Dégager une ou plusieurs problématiques.
- Qui ? Les biologistes.

Etape 2 :

- Choisir et définir les hypothèses.

But : Trouver un équilibre entre réalisme des hypothèses et simplicité des modèles satisfaisant de telles hypothèses.

Démarche : partir d'hypothèses restrictives et d'un modèle simple. Elargir les hypothèses les plus contraignantes et complexifier le modèle.

On peut distinguer deux types d'hypothèses :

- Les hypothèses "biologiques" : hypothèses choisies en fonction du cadre dans lequel on se place ;

- Les hypothèses "mathématiques" : hypothèses techniques qui permettent d'obtenir des résultats théoriques intéressants. Lorsque l'on choisie ce type d'hypothèse, il faut étudier l'impact biologique de celles-ci. Elles peuvent, ou non, avoir des conséquences importantes sur la pertinence du modèle.

- Effectuer une description mathématique de ces hypothèses.
- Qui ? Travail en commun entre biologistes et mathématiciens.

Etape 3 :

- Proposer une modélisation :

Analyse mathématique des phénomènes et des hypothèses, choix de la modélisation.

Retour possible à l'étape 2 pour rajouter des hypothèses mathématiques.

Définir les paramètres du modèle, notamment les entrées (input en anglais).

- Qui ? Travail en commun entre biologistes et mathématiciens.

Etape 4 :

- Résolution mathématique du problème au sein du modèle choisi.
- Simulation numérique des résultats mathématiques obtenus à partir d'estimations des paramètres d'entrée du modèle (input du modèle).
- Qui ? Mathématiciens, informaticiens et biologistes (essentiellement pour l'estimation des paramètres du modèle).

Etape 5 :

- Comparaison entre le comportement du modèle et la réalité. Critique du modèle et des hypothèses avec trois possibilités :
 - Le modèle et la réalité ne concordent pas : rejet soit de certaines hypothèses, soit de la modélisation.
 - Le modèle et la réalité concordent mais restent éloignés : élargissement des hypothèses et amélioration de la modélisation.
 - Le modèle et la réalité se confondent : utopique.
- Dans les deux premiers cas : Interprétation des résultats (output du modèle) puis retour aux étapes 2 et 3.
- Dans le troisième cas : Résolution de la problématique.
- Qui ? Biologistes, Mathématiciens et informaticiens.

Dans le cadre de ce cours, nous insisterons sur les étapes 2 et 3 de la modélisation. L'étape 4, plus mathématiques est développée dans le cadre de cours de second cycle de mathématiques appliquées ou de biologie. Les étapes 1 et 5, basées sur des connaissances biologiques, sont développées dans les cours classiques de biologie (Cours U. Nice 2005-2006).

1.1 DYNAMIQUE DES POPULATIONS :

Se propose d'étudier changements dans les effectifs et la composition d'une ou plusieurs populations et les facteurs environnementaux, biologiques ou humains qui expliquent ces changements.

Si l'on s'occupe de populations animales-végétales sauvages, l'outil naturel est un recensement répété dans le temps, pas voie aérienne ou terrestre. L'autre idée consiste à définir un modèle démographique qui décrit l'évolution des effectifs de la population dans le temps et on estime les paramètres par la méthode dite de capture – marquage-recapture. Celle-ci consiste à en suivre dans le temps d'individus marqués pour lesquels on enregistre la présence ou l'absence et l'état s'ils sont détectés (un emplacement géographique ou bien leur état physiologique par exemple).

Une approche innovante dite par la modélisation intégrée à espace d'état a été développée par CEF et le LASB. Elle consiste à représenter sous la forme d'un unique modèle dynamique à espace d'état, le modèle de recensement et le modèle démographique propre à la population étudiée. Cette approche rend caduque l'approche traditionnelle par approximation de la vraisemblance conjointe aux deux modèles séparés (Pradel et Gimenez).

Dans cette nouvelle approche en effet , l'estimation à partir des deux types de données recueillies , des paramètres démographiques d'intérêt ainsi que celle des valeurs des différentes variables d'état du modèle intégré peuvent être effectuées efficacement par différentes méthodes statistiques : les méthodes dites de filtrage et les méthodes dites Monte-Carlo par chaînes de Markov (MCMC), ces deux classes de méthodes présentent des avantages mais aussi des inconvénients

Toutes les applications de la méthode de filtrage classique par filtre de Kalman impliquent une linéarisation du modèle d'état intégré et une normalisation des erreurs, deux approximations aux conséquences incertaines. Elle permet d'état intégré par contre un traitement rapide des enregistrements disponibles (Besbeas et al .2003). son remplacement avantageux (en rigueur et précision) par des méthodes de filtrages plus récentes de type particulière (Rossi et Vila 2005- 2006) qui respectent l'intégrité des modèles (non linéaires et lois des aléas)

1.2.2 Mutualisme ou coopération :

C'est une association bénéfique pour les deux membres de l'association qui permet souvent de se procurer nourriture et défense contre les prédateurs.

L'intérêts :

le mutualisme améliore les conditions de la vie des espèces s'y rapportant. De plus, le mutualisme ne génère aucune obligation d'association dans le cas de la proto-coopération.

Exemple : les bactéries de l'intestin humain de la nourriture et fournissent en échange des vitamines.

1.2.3 Commensalisme :

Une espèce profite de l'autre mais sans lui nuire. Les gîtes communes et les moyens de transport sont de bons.

Exemple :

1 - les balanes sont fixées sur le dos des baleines ou des limules et profitent ainsi de leur déplacement.

2 - les poissons hémoras sont attachés à l'abdomen des requins par leur nageoire dorsale qui fonctionne comme une ventouse et se nourrissent ainsi des restes alimentaires de requins.

1.2.4 Neutralisme :

Absence d'interaction concurrentielles ou mutualiste entre deux espèces, ni le bénéfice, ni le détriment d'une espèce sur l'autre ne sont mesurables (Dajoz 2006).

1.2.5 Parasitisme :

C'est une relation interspécifique dans laquelle un des individus de l'association, le parasite, vit aux dépens de celui qui l'héberge, dont il tire sa nourriture.

A titre d'exemple, nous développerons chacun des quatre types de parasitisme :

- la parasitisme des végétaux par des végétaux.
- la parasitisme des animaux par des végétaux.
- le parasitisme des animaux par des animaux.
- le parasitisme des végétaux par des animaux.

Mais il ne faut pas oublier, que les microbes, virus bactéries pathogène, protozoaires ou champignons peuvent parasiter tous les groupes.

1.2.6 Prédation :

Une autre relations d'exploitation, celle du mangé au mangeront se déférence du parasitisme, par le fait que l'un des partenaires consomme purement et simplement le second, c'est la prédation, au sens strict du terme.. Un animal, le prédateur doit pour vivre, capturer un autre animal vivant, qui constitue sa proie, auquel il donne obligatoirement la mort pour l'ingérer après l'avoir tué. On trouve des prédateurs dans tous les milieux, dans l'air, l'eau et le sol.

1.2.7 Concurrence :

Interaction indirect, de type compétition pour une ressource insuffisant pour deux espèces occupant une même niche écologique.

1.3 PHENOMENES :

1.3.1 Phénomène déterministe :

Nous introduisons le problème à travers un exemple illustratif. Nous reprenons la méthodologie proposée dans l'introduction "modélisation" :

Etape I : Cadre et Problématique :

– **Cadre** : Des biologistes ont tenté d'acclimater le renne dans les îles de Beiring. Dans l'une d'elles, 21 individus furent introduits en 1911. En 1938, avant l'effondrement de la population, 2000 rennes furent dénombrés.

– **Problématique** : Quelle est l'évolution de la population de rennes sur cette île ?

Etape 2 : Hypothèses :

– Caractéristique de l'environnement des rennes :

La population étudiée vit sur un site isolé, comprenant des ressources suffisantes pour le développement de la population initiale de rennes. Les ressources de l'île permettent à la population initiale de subsister sans contraintes.

– L'évolution des populations est de type déterministe.

Modélisation :

A partir de ces premières hypothèses, le problème de modélisation est alors de proposer une formalisation abstraite de la description suivante :

"La population de rennes sur une île de Beiring évolue de manière déterministe, avec un effectif en 1911 de 21 individus et un effectif en 1938 de 2000 individus."

On modélise l'effectif de la population par une fonction du temps $y(t)$. A chaque instant t ,

$y(t)$ représente le nombre de rennes.

La dynamique d'une population est notion de fonction dérivée (Cours U. Nice 2005-2006). Etudier la dynamique d'une population revient alors à étudier la fonction dérivée de l'effectif de la population. Dans une première partie, nous allons présenter (rappeler) certaines propriétés des dérivées ainsi que les champs de vecteurs. Nous reviendrons alors sur la modélisation de l'évolution de la population de rennes et l'étude de cette modélisation. Avant cela, nous revenons sur la distinction entre modèle déterministe et modèle stochastique Commentaires :

Comme nous l'avions vu dans le premier chapitre, la dynamique d'un phénomène biologique peut être vue sous deux angles différents : déterministe ou stochastique.

Nous avons étudié la modélisation stochastique avec les chaînes de Markov. Nous

présentons ici des modèles déterministes. La modélisation déterministe s'applique à des phénomènes que l'on maîtrise mieux, pour lesquels on peut prédire le futur de manière convenable à partir des informations présentes.

On peut noter deux cadres d'applications :

- L'étude de phénomènes peu aléatoire.
- L'étude de population d'effectif suffisamment grand.

Le premier est naturel. Il faut toutefois noté que tout (ou presque) phénomènes aléatoires a une partie stochastique et une partie aléatoire. Si la partie stochastique est faible, on préférera une modélisation déterministe. Si la partie stochastique est forte, on choisira une modélisation stochastique.

La combinaison des deux est bien souvent une solution pertinente. Le second cas est une conséquence de la loi des grands nombres en probabilité. On étudie l'évolution d'un individu et on la suppose aléatoire. Si l'on regarde maintenant l'évolution d'un grand nombre d'individus semblables, sous certaines conditions l'évolution de cette population est déterministe.

1.3.2 Phénomène aléatoire :

Nous introduisons le problème à travers un exemple illustratif. Comme il s'agit d'une modélisation d'un phénomène biologique, nous suivons la méthodologie proposée.

Etape 1 : Cadre et Problématique :

- **Cadre :** Une parcelle en friche dans un milieu tempéré. On dispose de l'ensemble des informations disponibles sur cette parcelle (ensoleillement, caractéristiques du sol,...). On dispose aussi d'informations sur l'évolution de la végétation sur d'autres parcelles en friche.

On effectue des observations de l'évolution de la végétation tous les trois ans.

- **Problématique :** Quel type de végétation risque-t-il de se développer sur une telle parcelle ?

En particulier, va-t-on avoir l'apparition d'une forêt ?



Etape 2 : Hypothèses :

– Le choix des états possibles : on suppose dans cet exemple que trois types de végétation peuvent se développer sur ce type de parcelle : des herbes ou autres espèces pionnières (notées h), des arbustes (notés a) et enfin des arbres plus gros qui constituent des éléments d'une forêt (notés f).

Cette hypothèse est fortement restrictive car on peut observer dans la réalité d'autres types de végétation possible. Par contre cette restriction aux états les plus fréquents permet de développer un modèle compréhensible.

– Le type d'évolution sur une parcelle en friche en milieu tempéré.

Deux points de vue sont possibles :

Soit, à partir des informations sur la parcelle, on est certain du type de végétation qui poussera. Dans ce cas la modélisation de l'évolution de la végétation sera déterministe (Cours U. Nicc 2005-2006).

Nous développerons ce point de vue dans le chapitre suivant.

Soit, toujours à partir de la même source d'informations disponibles, on n'est pas certain de l'évolution de la végétation sur cette parcelle: On fait l'hypothèse de plusieurs évolutions possibles à partir de la situation présente. On suppose alors que l'évolution de la végétation est aléatoire (ou stochastique). Dans l'ensemble chapitre, on choisit cette seconde hypothèse.

Pour le moment, on ne peut proposer que des hypothèses "biologiques". Il faudrait connaître les conditions d'existence des outils mathématiques utiles à ce type de problème pour proposer les hypothèses mathématiques.

A partir de ces premières hypothèses, le problème de modélisation est alors de proposer une formalisation abstraite de la description suivante :

"La végétation sur une parcelle en friche évolue de manière aléatoire dans trois états distincts (h,a,f)":

L'objet mathématique le plus simple qui répond aux critères ci-dessus est la chaîne de Markov homogène, que nous introduisons dans la section suivante. La modélisation

de l'évolution par une chaîne de Markov homogène imposera de nouvelles hypothèses, les hypothèses mathématiques, qu'il faudra étudier d'un point de vue biologique.

1.4 MODELES ET MODELISATION :

1.4.1 Définitions :

Définition 1. Un modèle est une représentation symbolique de certains aspects d'un objet ou d'un phénomène du monde réel.

Le modèle donne donc une vision simplifiée ou idéalisée de la réalité.

Définition 2.

La modélisation est la démarche qui permet d'élaborer un modèle.

La modélisation prend en compte :

- l'objet et/ou le phénomène à représenter.
- les objectifs, c'est-à-dire l'utilisation que l'on souhaite faire du modèle.
- les données (relatives aux variables) et connaissances (relations entre variables) disponibles ou accessibles par l'expérience ou par l'observation.

La modélisation mathématique consiste à proposer une représentation dans le formalisme mathématique d'un objet ou d'un phénomène du monde réel.

Applications très variées : physique, chimie, épidémiologie, écologie, géologie, économie, biologie, sismologie ...

Un modèle a une fonction de représentation qui doit

rester un outil d'étude de la réalité : Décrire, Expliquer, Interpréter, Prévoir. Il sert parfois à tester l'écart entre l'état de la nature et les suppositions ayant servi à élaborer ce modèle.

Définition 3 :

Un système est un ensemble d'éléments qui se coordonnent pour concourir à un résultat ou de manière à former un ensemble.

1.4.2 Approches déterministes, stochastiques :

Les modèles mathématiques sont de nature soit déterministe soit stochastique.

Dans le premier cas, c'est une valeur bien précise qui est associée aux variables et aux paramètres, alors que dans le second ce sont des distributions de probabilité qui sont associées à ces grandeurs :

- Un modèle est déterministe s'il ne fait pas appel au calcul de probabilités.
- Un modèle est stochastique s'il fait appel au calcul de probabilités.

Par exemple, pour étudier l'évolution au cours du temps d'un système :

	Approche déterministe	Approche stochastique
Temps discret ($t \in \mathbb{N}$)	Equations récurrentes (ex. $u_{t+1} = f(u_t)$)	Processus en temps discrets (ex. processus (Xt) de Markov)
Temps continu ($t \in \mathbb{R}_+$)	Equations différentielles (EDO, EDP,..)	- Processus en temps continu (ex. processus gaussien) - Equations différentielles stochastiques

1.4.3 Comment modéliser ?

Modéliser, c'est créer une structure mathématique. On ne peut s'y prendre n'importe comment. Il nous semble, que dans l'activité de modélisation il y a une démarche commune en quatre étapes :

—La première consiste en une analyse des processus réels qui sont souvent complexes et des données expérimentales.

– La seconde est la construction du modèle : créer à partir d'une situation réelle, une situation mathématisable puis se ramener à un problème purement mathématique. Une première simplification est d'abord nécessaire. Par un choix d'hypothèses, on restreint la complexité du problème. En effet, une situation réelle dépend de nombreux paramètres ; on va choisir d'en conserver certains, d'en négliger d'autres, dans un premier temps au moins. Ces choix sont nécessaires car on doit toujours les mettre en regard avec les concepts et outils mathématiques disponibles. Ces derniers sont cependant en nombre limité. Ils vont nécessairement entraîner une approximation de la réalité. **Malthus** simplifie la réalité, il modélise avec des relations de récurrence qui débouchent sur des suites géométriques et arithmétiques, mais nous allons voir que cette approche de la réalité peut-être plus fine et que d'autres suites sont mieux adaptées. Mais Malthus aurait pu utiliser d'autres outils que les suites, les fonctions par exemple. Dans un grand nombre de cas – nous allons en voir un exemple – le choix d'outils différents débouche sur une modélisation différente. Le type de ressources mathématiques utilisé est fondamental et influe sur le résultat.

- choix d'hypothèses différents \Rightarrow modélisations différentes
- choix d'outils mathématiques différents \Rightarrow modélisations différentes

La troisième étape est un travail sur le modèle, dans le champ des mathématiques : Après sa création, c'est la mise en fonctionnement mathématique du modèle ainsi élaboré : pour résoudre le problème mathématique, on va produire des résultats, qui seront donc des **conjectures** à retenir. Cela se fait à la fois avec l'outil calcul et l'outil graphique. Les calculatrices et les logiciels apportent une aide précieuse.

La quatrième étape est la mise en oeuvre du modèle. On confronte les résultats obtenus avec la réalité observée. Elle est souvent laissée de côté par le mathématicien. À tort ; l'exemple de Malthus le prouve. Elle consiste en une simulation de la réalité à l'aide du modèle créé, ce qui permet d'une part une analyse critique de sa pertinence et de ses limites d'application et d'autre part d'éventuelles améliorations.

1.5 ETAPES COURANTES D'UNE MODELISATION

Etape 1.

Elaborer une vue synthétique de la situation, des connaissances a priori, des données disponibles ou accessibles.

Etape 2.

Caractériser et analyser le phénomène ou l'objet que l'on souhaite représenter. Pour cela, il est nécessaire de :

1) spécifier si le système est isolé, clos ou fermé.

2) classer les variables agissant sur le système

– Variables d'état, d'écrivant l'état du système (taille ou densité d'une population, concentration d'un produit,...)

– Variables d'action qui modifient l'état du système par action externe (par exemple température par chauffage ou refroidissement, injection ou prélèvement d'un produit, immigration ou émigration contrôlées,...)

– Variables d'observation ou observables qui permettent d'avoir des informations sur l'état du système (ces observables peuvent être directement des variables d'état).

3) Définir les relations entre les variables.

Etape 3.

Choisir et élaborer un modèle.

Exemple :

Croissance d'une population limitée en ressources, en milieu isolé (pas d'immigration ou d'émigration)

$X(t)$ = biomasse à la date t .

$s(t)$ = ressources à la date t .

La vitesse de croissance peut s'exprimer par une équation

différentielle ordinaire (EDO) $\frac{dx}{dt} = f(x, s)$

Si le rendement de la croissance est supposé constant, on a

$$x(t) - x(0) = R(s(0) - s(t)) \quad \text{où} \quad R \in \mathbb{R}_+$$

donc : $\frac{dx}{dt} = -R \frac{ds}{dt}$

Le phénomène peut donc être représenté par le système de deux EDO

$$\begin{cases} \frac{dx}{dt} = f(x, s) \\ \frac{ds}{dt} = -\frac{1}{R} f(x, s) \end{cases}$$

avec les valeurs initiales $x(0) = x_0$ et $s(0) = s_0$.

La vitesse de croissance doit vérifier les contraintes :

- a) $f(., .)$ est une fonction positive dans le premier quadrant ($x \geq 0, s \geq 0$)
- b) $f(0, s) = 0$ (il n'y a pas de génération spontanée).

Le modèle le plus simple concernant cette contrainte est $\frac{dx}{dt} = kxg(s)$ avec $k > 0$ et g une fonction réelle.

- c) $f(x, 0) = 0$ (s'il n'y a pas de ressources alors la croissance est nulle)

Le modèle le plus simple vérifiant ces trois contraintes est $\left\{ \begin{array}{l} \frac{dx}{dt} = kxs \\ \frac{ds}{dt} = -\frac{1}{R}kxs \end{array} \right. \dots\dots\dots(1)$

Comme $s = s_0 - \frac{1}{R}(x - x_0)$, on peut ramener ce système de deux EDO à une seule EDO :

$$\frac{dx}{dt} = kx \left(s_0 - \frac{1}{R}(x - x_0) \right) = \frac{kx}{R} (Rs_0 + x_0 - x)$$

donc : $\boxed{\frac{dx}{dt} = rx \left(1 - \frac{x}{\rho}(x - x_0) \right) \dots\dots\dots(2)}$

avec : $\rho = Rs_0 + x_0$ et $r = \frac{k\rho}{R}$

Remarque :

il s'agit d'élaborer le modèle le plus simple possible tout en étant le plus proche possible de la réalité.

Etape 4.

Vérifier les propriétés qualitatives du modèle.

Elles doivent être en accord avec ce que l'expérience nous permet d'observer, de mesurer, ou ce que les connaissances acquises nous disent.

Dans l'exemple de croissance de population, on aura à étudier la courbe représentative de la fonction solution de l'EDO (2).

- $x = \rho$ et $x = 0$ sont des points fixes
- Si $x \in]0, \rho[$, $\frac{dx}{dt} > 0$ et $x(t)$ croît avec t
- Si $x > \rho$, $\frac{dx}{dt} < 0$ et $x(t)$ décroît avec t .

• Etude des points d'inflexion :

$$\frac{dx^2}{dt^2} = r \left(1 - \frac{2x}{\rho}\right) \frac{dx}{dt}$$

$$dt^2 \text{ s'annule pour } \frac{dx^2}{dt^2} = 0 \text{ ou } x = \frac{\rho}{2}$$

Elle change de signe pour $x = \frac{\rho}{2}$ qui est donc un point d'inflexion.

On a une solution explicite de l'EDO :

$$x(t) = \frac{\rho}{1 + \frac{\rho - x_0}{x_0} \exp(-rt)} \dots\dots\dots(3)$$

Etape 5.

Identifier le modèle, c'est-à-dire attribuer des valeurs numériques aux paramètres du modèle à partir de données expérimentales.

Etape 6.

Valider le modèle à partir de tests statistiques.

1.6 Test d'ajustement au modèle

Il permet de déterminer si les durées de vie dans une population obéissent au modèle proposé (avec une probabilité de rejeter à tort contrôlée).

Le plus célèbre est celui dit du X^2 :

Principe : comparer fréquences observées de survie aux fréquences théoriques fournies par le modèle.

Calcul de $C = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$ où

k = nombre de classes considérées

n_i = effectif d'individus dans la classe i

p_i = probabilité de la classe i

$n = \sum_{i=1}^k n_i$ est le nombre d'individus dans l'échantillon.

Règle de décision :

Rejet du modèle si $C > X_{1-\alpha, v}^2$

Non rejet du modèle si $C \leq X_{1-\alpha, v}^2$

où $X_{1-\alpha, v}^2$ est le fractile d'ordre $1 - \alpha$ de la loi dite du X^2 à v degrés de liberté.

1.6.1 Caractéristiques d'un modèle :

- Un modèle définit une frontière entre la réalité et la perspective de l'observateur. Ce n'est pas « la réalité », mais une vue subjective de la réalité

- Bien qu'un modèle ne représente pas une réalité absolue, il en reflète des aspects importants, il en

donne une vue juste et pertinente

- Le choix du modèle a une influence profonde sur la manière de comprendre la réalité
- Tout modèle peut être exprimé à différents niveaux de précision
- Un modèle parfait d'un système complexe ne servirait à rien car il ne serait pas opérationnel

- Plusieurs modèles différents sont utiles pour approcher la même réalité
- Une modélisation ne dépend pas que du phénomène étudié, mais aussi du but poursuivi. Par exemple une boîte noire, comme en fournit la statistique empirique, pourra être utile pour la prise de décision mais non pour comprendre le fonctionnement du système

- Un bon modèle dans un contexte peut être mauvais dans un autre contexte. Exemple : un modèle météo mis au point pour prédire le futur immédiat peut être incapable de prédire le climat à long terme

Nous nous restreignons ici au modèle statistique

- Il est simple dans sa forme
- Il est basé sur un jeu de données
- Il permet d'appréhender des relations très simplifiées entre variables
- Il ne permet que de comprendre partiellement mais peut être efficace pour prédire

1.6.2 Construction d'un modèle statistique :

- Identifier la structure du modèle : sa forme (linéaire, polynomial ...), la variable qui décrit le mieux l'évolution du système (prédicteur, ou variable dépendante), les variables qui peuvent la contraindre (prédicteurs ou variables indépendantes)
 - Rassembler des données représentatives des différents états du système
 - Si on n'a aucune idée a priori de la forme du modèle, on s'aide des données pour la déterminer (représentation graphique)
 - Estimer les paramètres du modèle en minimisant les écarts entre les estimations et les observations du prédicteur (calibration)
 - Déterminer les limites du modèles (dans quelles situations il est valable) : en général il est hasardeux de prédire en dehors de l'intervalle couvert par les observations du prédicteur
 - Quantifier les incertitudes sur les paramètres et les prédictions

$$k = -2.00 + 0.34$$

$$x_1 = 5.12 + 0.26$$

$$x_2 = 1.04 + 0.18$$

Erreur standard résiduelle = 1.96

(197 degrés de liberté)

$$R_2 = 0.88$$

1.6.3 Interprétation du modèle :

- Le signe des coefficients permet de comprendre dans quel sens est corrélé le prédicteur avec le prédicteur
- La relation de cause à effet ne sera possible que si elle est établie physiquement (ou biologiquement)
- Le coefficient étant accompagné d'intervalle de confiance, on considère que le prédicteur est significativement corrélé avec le prédicteur si et seulement si le « zéro » n'appartient pas à l'intervalle (niveau de confiance à 95% par exemple)
- Un modèle doit être le plus simple possible (principe d'économie des paramètres)

1.6.4 Validation du modèle :

- Nécessité d'avoir des données n'ayant pas servi à la calibration
- Donc couper le jeu de données de calibration en deux parties (non nécessairement égales)
 - Calibrer sur la partie 1 et prédire sur la partie 2
 - Les statistiques sur la partie 2 donne les statistiques de vérification
 - Calibrer sur la partie 2 et prédire sur la partie 1
 - Les statistiques sur la partie 1 donne les statistiques de vérification
 - Si les statistiques de vérification sur les deux parties sont bonnes, la calibration finale, faite sur l'ensemble des données, est celle qui est finalement utilisée pour toute prédiction ultérieure

1.6.5 Conclusions :

- Les prédicteurs doivent être liées au prédicant
- Les données doivent être aussi diversifiées que possible
- Un bon modèle ne contient pas de prédicteur inutile (principe d'économie)
- Un bon modèle doit être vérifié sur des données indépendantes
- Il existe d'autres méthodes de validation : plusieurs méthodes valent mieux qu'une

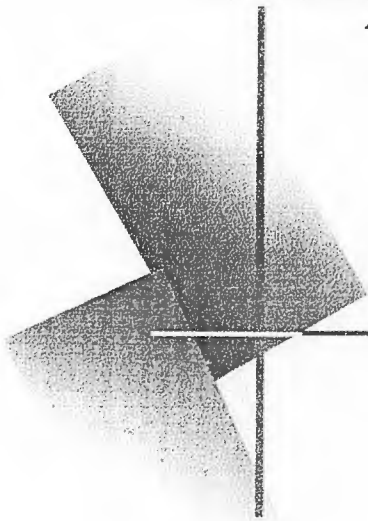
car aucune n'est parfaite

- Un modèle statistique est plus efficace pour prédire que pour expliquer
- Seuls les modèles numériques avec des équations théoriques et/ou basées sur l'expérimentation permettent explication et prédiction
- Dans tous les cas, il faut une bonne quantification des incertitudes

Chapitre II

Concept

Mathématique



Chapitre 2

Concepts mathématiques

2.1 ÉQUATIONS DIFFÉRENTIELLES :

Les équations différentielles se retrouvent dans de très nombreux domaines, en physique (par exemple mécanique), en chimie (cinétique des réactions), ou en biologie (dynamique des populations),etc.

Ce qui suit s'inspire beaucoup du cours de Jacques Istas de troisième année ((Modèles mathématiques pour l'écologie)). Les modèles démographiques les plus simples concernent une population isolée. Pour tout instant t , on note $N(t)$ l'effectif d'une population à cet instant, on suppose que cet effectif est assez important pour supposer $N(t)$ réel (et non pas entier). L'évolution de cette population est décrite par une équation différentielle :

$$dN(t)dt = \text{naissances} - \text{morts} + \text{migrations}$$

S'il n'y a pas de migrations, et si les naissances et les morts sont proportionnels à la taille de la population, on obtient une équation linéaire :

$$dN(t)dt = aN(t) - bN(t) :$$

Dans cette approche (modèle de Malthus (1766-1834), la population croît ou décroît exponentiellement vite, ce qui n'est pas réaliste, surtout à long terme. On peut alors ajouter une correction lorsque la population grossit : il existe une taille idéale, dite capacité biotique. En dessous de cette capacité, la population augmente, au dessus elle diminue.

Un modèle possible, du à Verhulst (1804-1849) et dit logistique, est le suivant :

$$DN(t)dt = rN(t)\mu_1N(t)K$$

K est la capacité biotique et les solutions de cette équation sont :

$$N(t) = N(0)KertK + N(0)(ert_i 1)$$

Cette équation admet deux points d'équilibre 0 et K. En effet si $N(0) = 0$, la population reste nulle au cours du temps. Si $N(0) = K$, la taille de la population reste stable, égale à K. De plus, si on voit que si $N(0) > 0$, alors \lim

$t_i + 1N(t) = K$. Ceci signifie que K est un point d'équilibre stable, tandis que 0 est instable : si $N(0)$ est proche de K, la solution reste proche de K ; si $N(0)$ est proche de 0, elle ne reste pas près de zéro.

On peut aussi faire entrer en jeu l'interaction avec d'autres populations. On va se limiter à deux populations avec une interaction proie- prédateur. Le premier modèle est du à Volterra (1860-1940) et est appelé modèle de Lotka-Volterra, car il a été introduit presque simultanément par Lotka comme représentation d'un système chimique exhibant un caractère oscillant. Ce modèle est basé sur les hypothèses suivantes :

- sans prédateurs, la population des proies croît exponentiellement vite (dynamique de Malthus) ;

- sans proies, le taux de morts parmi les prédateurs est proportionnel à la taille de la population ;

- le taux de disparition des proies est proportionnel au nombre de rencontres entre une proie et un prédateur, supposé lui-même proportionnel au produit des deux populations (Depiereux et Calmant 2004).

la taux de croissance des prédateurs est aussi proportionnel au nombre de rencontres entre proie et prédateur.

Si N est l'effectif des proies et P celui des prédateurs, on obtient le système d'équations :

$$dN(t)dt = N(a_i bP); dP(t)dt = P(cN_i d)$$

a, b, c et d étant quatre constantes positives. Ce système dynamique a des propriétés

très particulières. Il a deux points d'équilibre $(0; 0)$ et $(d=c; a=b)$, et dans le plan de phase, i.e. dans le plan dont les coordonnées sont N et P , les trajectoires d'un mobile données par le système, sont fermées, donc N sont périodiques .

Ce modèle a lui aussi ces limites. En effet en l'absence de prédateurs, il n'est pas réaliste de penser que la population des proies va croître indéfiniment.

À travers ces exemples, nous avons vu différents cas : équations linéaires, non linéaires, systèmes d'équations, sans savoir pourquoi ces équations avaient des solutions, ni le cas échéant, si elles admettaient une unique solution. C'est l'objet du paragraphe suivant.

2.2 Statistiques descriptives :

Le premier souci de l'expérimentateur sera de chercher à mettre de l'ordre dans ses observations, de façon à pouvoir les comprendre et les synthétiser pour pouvoir ensuite les comparer et les transmettre (Depiereux et Calmant 2004) .

2.2.1 Objective :

La population est l'ensemble des observations individuelles limitées dans l'espace et dans le temps, au sujet desquelles on désire tirer des conclusions.

La population doit être définie en fonction de l'objectif de l'étude.

Supposons que je décide de mener une enquête sur l'évolution de la consommation de tabac. Veut-on étendre les conclusions à la population mondiale? Occidentale? Belge? S'intéresse-t-on à toutes les catégories de la population? Seulement aux femmes? Cible-t-on seulement les jeunes? Les moins de 15 ans? Je peux définir que ma population représente les immigrés de 18 à 25 ans, à Bruxelles. Dans ce cas, les conclusions de l'enquête ne pourront absolument pas être étendues à une autre population.

L'échantillon est l'ensemble des observations individuelles sélectionnées dans la population.

Il est généralement impossible de réunir l'information relative à toutes les observations

individuelles comprises dans la population. L'échantillonnage est le procédé suivant lequel on sélectionne l'information. Il doit être guidé par la définition de la population.

L'échantillon doit être représentatif de la population.

Une variable est la caractéristique numérique mesurée lors de l'observation.

Par exemple, la quantité de nicotine absorbée par le fumeur sera évaluée par la concentration de nicotine dans l'urine. La variable est notée X et les différentes valeurs $x_1, x_2, x_3 \dots$ de façon générale ou 6.2, 4.3, 7.8 $\mu\text{g/l}$ d'urine dans un cas particulier. De façon générale, nous dirons avoir réalisé n observations, et une observation particulière sera notée $x_i, (i = 1, \dots, n)$.

Une variable discrète est une variable qui montre une discontinuité.

Une variable continue est une variable qui peut prendre un nombre infini de valeurs dans un intervalle donné.

Beaucoup de variables naturelles sont des variables continues : le pH d'une solution, le volume d'un organe, la concentration d'une substance dans le sang.

Expérimentalement, une variable est toujours rendue discrète par la précision de la mesure effectuée. Elles seront néanmoins considérées comme des variables continues. On dira qu'une variable continue peut prendre une infinité non dénombrable de valeurs.

2.2.2 Les séries simples :

Celles-ci se limitent à décrire une seule variable à la fois, discrète (le nombre d'enfants par famille) ou continue (concentration naturelle des eaux en Cd)

Des variables discrètes :

Une façon commode de représenter les résultats consiste à créer une distribution statistique des valeurs numériques. La distribution de fréquence reprend l'ensemble des k valeurs différentes observées, classées par ordre croissant, $x_1, x_2, \dots, x_i, \dots, x_k$, et les

fréquences correspondantes, $n_1, n_2, \dots, n_i, \dots, n_k$, la fréquence n_i étant le nombre de fois que j'observe une valeur particulière x_i dans l'échantillon.

n étant l'effectif, on a évidemment la relation :

$$\sum_{i=1}^k n_i = n \dots\dots\dots (4)$$

Cette équation implique que chaque observation peut être classée dans une et une seule des k catégories définies.

Toute statistique se réfère à des conventions : il faudra décider comment classer une femme enceinte, un enfant décédé, les enfants de familles recomposées...

En général, on définira toutes les catégories correspondant aux valeurs discrètes comprises entre les valeurs minimum et maximum observées ; si certaines valeurs ne sont pas observées, on associera à ces catégories une fréquence nulle.

Cela fait, on y voit déjà plus clair dans les résultats. Cependant, si je veux comparer

Entre deux échantillon. On aura donc intérêt à établir la distribution de fréquences relatives dans laquelle chaque fréquence est exprimée en proportion (comprise entre 0 et 1) ou en pourcentage (compris entre 0 et 100) de l'effectif.

$$n'_i = \frac{n_i}{n} \text{ ou } n'_i = \frac{n_i}{n} \times 100 \dots\dots\dots (5)$$

Enfin, les distributions de fréquences cumulées permettent de répondre facilement à des questions du type : quelle est la proportion de familles profitant d'une réduction au chemin de fer ? La fréquence cumulée est établie en additionnant les fréquences de proche en proche à partir de la première valeur n_1 . La fréquence relative cumulée est définie de la même façon. Si N'_i est la fréquence relative cumulée correspondant à la catégorie i , on peut écrire :

$$N_i = \sum_{j=1}^i n_j \text{ et } N'_i = \sum_{j=1}^i n'_j \dots\dots\dots (6)$$

$$N^k = n, N'^k = 1 \text{ ou } 100.$$

On peut ainsi calculer immédiatement la proportion de familles qui ont 3 enfants et plus :

L'abscisse correspond aux différentes valeurs discrètes prises par la variable X , et l'ordonnée représente une des distributions de fréquence.

Un tel diagramme est appelé diagramme de barres, en raison de la discontinuité de l'abscisse. Lorsqu'une fréquence relative est exprimée en ordonnée, le nombre total d'observations doit être mentionné sur le graphique.

Paramètres :

Définir le centre d'une distribution, c'est trouver un nombre caractéristique de sa position. Il existe pour cela plusieurs possibilités.

Moyenne arithmétique : La moyenne arithmétique est certainement le moyen le plus courant d'exprimer la tendance centrale d'une distribution. Pour n observations $x_1, x_2, \dots, x_i, \dots, x_n$, elle est définie comme :

$$M_x = \frac{1}{n} \sum_{i=1}^n x_i \dots \dots \dots (7)$$

Si l'on dispose seulement des valeurs regroupées en k classes, on définit $x_1, x_2, \dots, x_i, \dots, x_k$ comme étant les centres de classes, et en considérant n_i la fréquence de chaque classe, on peut estimer

$$M_x = \frac{1}{n} \sum_{i=1}^k x_i n_i \dots \dots \dots (8)$$

La moyenne arithmétique est simple à calculer, et elle présente l'avantage algébrique d'assurer que la somme des écarts à la moyenne est nulle et que la somme des carrés des écarts peut être décomposée en composantes additives .

groupes d'égale importance. S'il y a un nombre impair d'observations, M_e est une observation de la série.

2.2.3 Les séries multiples :

Nous allons à présent aborder les techniques relatives à la description de deux variables mesurées simultanément. Ces techniques sont utilisées lorsque l'expérimentateur s'intéresse à la relation qui pourrait exister entre deux variables qui interviennent dans un phénomène naturel. Ces études sont très fréquentes ; imaginons par exemple l'étude de l'évolution de la capacité respiratoire en fonction de l'exercice, de l'évolution du rythme cardiaque en fonction de l'administration d'une drogue, du poids des individus en fonction de leur taille, du taux de cholestérol sanguin en fonction du poids du corps, de la production laitière en fonction de la teneur des aliments en protéines etc...

L'expérimentateur mesure deux valeurs expérimentales pour chaque observation individuelle, ce qui produira une série statistique à deux dimensions :

Nous pouvons généraliser les techniques utilisées pour la description des observations à une dimension. Les variables continues seront regroupées en classes comme précédemment.

On dresse alors une table de fréquences présentant en tête de lignes les classes de cholestérol et en tête de colonnes les classes de poids. Chaque élément de cette table représente le nombre d'observations individuelles qui se trouvent dans la catégorie déterminée simultanément par les classes définies pour les deux variables.

S'il existe une relation entre les variables, elle apparaîtra sous la forme d'une structure diagonale dans le tableau : aux poids plus élevés correspondent généralement des taux de cholestérol plus élevés.

La table de fréquences permet d'obtenir un classement des observations expérimentales. Elle peut être représentée sous la forme d'un stéréogramme qui est une généralisation en trois dimensions de l'histogramme.

Ce type de représentation reste toutefois relativement lourd à manipuler, et ne pro-

pose pas de valeurs synthétisant l'essentiel de l'information. Nous allons entreprendre de rechercher les valeurs caractéristiques permettant de mesurer l'intensité de la relation entre deux variables, et de caractériser cette relation.

Nous envisageons cette démarche dans le cadre des variables continues.

-Covariance :

La valeur moyenne de la variabilité conjointe de X et Y est appelée covariance.

$$\boxed{cov_{xy} = S_{xy} = \frac{SPE_{XY}}{n} \dots\dots\dots(9)}$$

La "co"variance mesure si les variables varient "ensemble" ou non (co - comme dans copain , collaborateur). Comme pour la variance, il existe une autre définition, dans laquelle n est remplacé par n -1.

-Coefficient de corrélation linéaire :

La dernière étape de notre démarche va être de standardiser les variables X et Y, c'est -à -dire de rendre les valeurs indépendantes du système d'unité. Pour cela, nous allons exprimer les écarts à la moyenne en "unités écart -type", ce qui correspond à la transformation :

$$\boxed{Z_{xi} = \frac{x_i - M_x}{S_x} \text{ et } Z_{yi} = \frac{y_i - M_y}{S_y}}$$

Z étant un écart à la moyenne, le calcul de SPE se simplifie :

$$\boxed{SPE_{Z_X Z_Y} = \sum_{i=1}^n Z_{xi} Z_{yi}}$$

$$R = \frac{cov_{xy}}{S_X S_Y} = \frac{S_{xy}}{S_X S_Y} \dots\dots\dots (10)$$

C'est ce que l'on appelle R , le coefficient de corrélation linéaire de Spearman.

On peut considérer R comme étant la covariance des variables préalablement standardisées.

Pour établir que, dans le modèle linéaire, R^2 est le coefficient de détermination nous devons d'abord établir la mesure des paramètres de la droite de régression.

On peut considérer R comme étant la covariance des variables préalablement standardisées.

Pour établir que, dans le modèle linéaire, R^2 est le coefficient de détermination nous devons d'abord établir la mesure des paramètres de la droite de régression.

Relations nonn lineaires :

Ni R ni R^2 ne permettent de déterminer si une relation est linéaire ou non.

Certaines relations très étroites entre deux variables, mais non linéaires, correspondent à un coefficient de corrélation nul ou presque nul.

Ceci montre bien que la variabilité résiduelle inexplicée (ici 98.5%) n'est pas nécessairement inexplicable : seul un modèle non linéaire pourrait contribuer à l'expliquer.

Dans d'autres cas le modèle linéaire peut expliquer une partie de la variabilité de Y , ce qui explique que dans ce cas R^2 ne soit pas nul, et puisse même être très élevé.

Ceci s'explique par le fait que le modèle linéaire n'est pas très éloigné des points. La variabilité résiduelle est faible et R^2 est élevé. Le graphique montre manifestement qu'une courbe s'accorderait mieux au phénomène décrit.

Un coefficient R^2 élevé ne signifie pas nécessairement que la relation soit linéaire.

2.3 LES ANALYSES STATISTIQUES :

2.3.1 Régression linéaire :

Utilisez ce module pour créer un modèle de régression linéaire simple ou multiple dans un but explicatif ou prédictif.

La régression linéaire est sans aucun doute la méthode statistique la plus utilisée. On distingue habituellement la régression simple (une seule variable explicative) de la régression multiple (plusieurs variables explicatives) bien que le cadre conceptuel et les méthodes de calculs soient identiques (Dagnelie 2006).

Le principe de la régression linéaire est de modéliser une variable dépendante quantitative Y , au travers d'une combinaison linéaire de p variables explicatives quantitatives, X_1, X_2, \dots, X_p . Le modèle déterministe (ne prenant pas en compte d'aléa) s'écrit pour une observation i , où y_i est la valeur observée pour la variable dépendante pour l'observation i , x_{ij} est la valeur prise par la variable j pour l'observation i , et e_i est l'erreur du modèle.

Le cadre statistique et les hypothèses qui l'accompagnent ne sont pas nécessaires pour ajuster ce modèle. Par ailleurs la minimisation par la méthode des moindres carrés (on minimise la somme des erreurs quadratiques e^2_i) fournit une solution analytique exacte. Néanmoins si l'on veut pouvoir tester des hypothèses et mesurer le pouvoir explicatif des différentes variables explicatives dans le modèle, un cadre statistique est nécessaire.

Les hypothèses de la régression linéaire sont les suivantes : les erreurs e_i suivent une même loi normale $N(0, s)$ et sont indépendantes.

L'écriture du modèle complétée par cette hypothèse a pour conséquence que, dans le cadre du modèle de régression linéaire, les y_i sont des réalisations de variables aléatoires de moyenne μ_i et de variance S^2 .

Si l'on souhaite utiliser les différents tests proposés dans les résultats de la régression linéaire il est recommandé de vérifier a posteriori que les hypothèses sous-jacentes sont bien vérifiées. La normalité des résidus peut être vérifiée en analysant certains graphiques

ou en utilisant un test de normalité. L'indépendance des résidus peut être vérifiée en analysant certains graphiques ou en utilisant le test de Durbin Watson.

2.3.2 ANOVA :

Utilisez ce module pour faire de l'ANOVA (Analyse de variance) à un ou plusieurs facteurs, équilibré ou déséquilibré. Des options avancées vous permettent de choisir les contraintes sur le modèle et de tenir compte des interactions entre les facteurs. Des tests de comparaisons multiples peuvent être calculés.

L'analyse de variance utilise le même cadre conceptuel que la régression linéaire. La différence principale vient de la nature des variables explicatives : au lieu d'être quantitatives, elles sont ici qualitatives

2.3.3 ANCOVA :

L'ANCOVA (Analyse de covariance) peut être vue comme un mélange d'ANOVA et de régression linéaire puisque la variable dépendante est de même nature, le modèle est aussi un modèle linéaire, et les hypothèses sont identiques. Il en réalité est plus juste de considérer l'ANOVA et la régression linéaire comme des cas particuliers de l'ANCOVA.

2.3.4 Régression logistique :

Utilisez la régression logistique pour modéliser une variable binaire en fonction de variables explicatives quantitatives ou qualitatives (Dagnelie 2006).

La régression logistique est une méthode très utilisée car elle permet de modéliser des variables binaires ou des sommes de variables binaires. Elle est très utilisée dans le domaine médical (guérison ou non d'un patient), en sociologie et de l'épidémiologie (analyse d'enquêtes), en marketing quantitatif (achat ou non de produits ou services suite à une action) et en finance pour modélisation de risques

Le principe du modèle de la régression logistique est de relier la survenance ou la

non survenance d'un événement au niveau de variables explicatives. Par exemple, dans le domaine phytosanitaire, on cherche à évaluer à partir de quelle dose d'un agent chimique, un insecte sera neutralisé.

Le modèle de régression logistique permet d'estimer la force de l'association entre une variable qualitative deux classes (dichotomique) appelée variable dépendante et des variables qui peuvent être qualitatives ou quantitatives appelées variables explicatives ou indépendantes. La variable dépendante est la survenue ou non de l'événement étudié – maladie, acte médical, ... – et les variables explicatives ont des facteurs susceptibles d'influencer la survenue de l'événement (facteurs d'exposition ou facteurs de confusion).

Modèle de régression logistique $y = \frac{f(x)}{1 + f(x)}$ avec $f(x) \geq 0$

2.3.5 Analyse Factorielle Multiple (AFM) :

Utilisez l'Analyse Factorielle Multiple (AFM) pour analyser simultanément plusieurs tableaux de variables, et obtenir des résultats, notamment des représentations graphiques, qui permettent d'étudier la relation entre les observations, les variables et les tableaux. A l'intérieur d'un tableau les variables doivent être de même nature (quantitative ou qualitative), mais les tableaux peuvent être de différents types (Dagnelie 2006).

L'Analyse Factorielle Multiple (AFM), permet d'analyser simultanément plusieurs tableaux de variables, et d'obtenir des résultats, notamment des représentations graphiques, qui permettent d'étudier la relation entre les observations, les variables et les tableaux (Escofier et Pagès, 1984). A l'intérieur d'un tableau les variables doivent être de même nature (quantitative ou qualitative), mais les tableaux peuvent être de différents types.

L'AFM est une synthèse de l'ACP (Analyse en Composantes Principales) et l'ACM (Analyse des Correspondances Multiples) qu'elle permet de généraliser à des variables hétérogènes. La méthodologie de l'AFM se décompose en deux étapes :

2.3.6 Analyse descriptive :

Utilisez cet outil pour calculer les statistiques descriptives adaptées aux séries chronologiques.

L'une des phases essentielles de l'analyse des séries chronologiques consiste à déterminer si une valeur observée à un temps t dépend de ce qui a été observé dans le passé ou non. Si la réponse est affirmative, alors l'étape suivante essaiera de répondre à comment se manifeste cette dépendance (Dagnelie 2006).

Analyse d'effets de dose :

Utilisez cette fonction pour modéliser les effets d'une dose sur une variable réponse, en prenant éventuellement en compte un effet de mortalité naturelle.

Cet outil s'appuie sur la régression logistique (modèles Logit, Probit, Log-log complémentaire, Gompertz) pour modéliser l'impact de doses de composants chimiques (par exemple un médicament, un produit phytosanitaire) sur un phénomène binaire (guérison ou non, mort ou non).

Plus d'information sur la régression logistique est disponible dans la section de l'aide dédiée à ce sujet.

Mortalité naturelle :

Cet outil permet de prendre en compte la mortalité naturelle afin de modéliser plus précisément le phénomène étudié. En effet, si l'on considère une expérience réalisée sur des insectes, certains périront en raison de la dose injectée, d'autres en raison d'un autre phénomène. L'ensemble de ces phénomènes connexes n'est pas intéressant pour l'expérience concernant les effets de dose, mais il peut être pris en compte. Si p est la probabilité issue d'un modèle de régression logistique correspondant uniquement à l'effet de la dose, et si m est la mortalité naturelle, alors la probabilité observée pour que l'insecte succombe est :

$$P(obs) = m + (1 - m) * p$$

La formule d'Abbott (Finney, 1971) s'écrit

$$p = (P(obs) - m)/(1 - m)$$

La mortalité naturelle m peut être entrée par l'utilisateur parce que connu grâce à des expériences préalables

2.4 LES MATRICES :

2.4.1 Définitions

Définition 1 : On appelle matrice A ou $A_{p \times q}$ ou $[a_{ij}]$ un tableau rectangulaire constitué de $p * q$ éléments $(a_{ij})_{i=1,p}^{j=1,q}$

$$A = A_{p \times q} = [a_{ij}] = \begin{bmatrix} a_{11} & \cdot & \cdot & \cdot & a_{1q} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & a_{ij} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{p1} & \cdot & \cdot & \cdot & a_{pq} \end{bmatrix} ; \text{ Exemple : } \begin{bmatrix} 3 & 2 & 5 \\ 6 & -1 & 1 \\ 9 & 3 & -4 \\ 5 & 0 & 0 \end{bmatrix}$$

Le nombre de ligne p et le nombre de colonne q constituent les dimensions ou l'ordre de la matrice (Mentré 2007-2008).

Définition 2 : Une matrice qui ne comporte qu'une ligne ou une colonne est appelée vecteur (vecteur ligne ou vecteur colonne).

En particulier, une matrice qui ne possède qu'un seul élément, est considérée comme égale à cet élément $c_{11} = [c] = c$.

Définition 3 : Une matrice dont le nombre de lignes est égal au nombre ($p = q$) de colonnes est dite carrée.

Définition 4 : Une matrice carrée est dite symétrique lorsque tous les éléments occupant de positions symétriques par rapport à la diagonale descendante sont égaux.

Définition 5 : Une matrice est dite triangulaire lorsque les éléments situés d'un même coté de cette diagonale sont nuls ($a_{ij} = 0$ pour tout $i > j$ ou $i < j$, $i \neq j$).

Définition 6 : Matrice identité ou unité si $a_{ii} = 1$, et $a_{ij} = 0$ si $i \neq j$.

Définition 7 : Transposée d'une matrice

La transposée d'une matrice A_{pq} est la matrice A_{qp} obtenue en permutant les lignes et les colonnes, elle est telle que $a'_{ij} = a_{ji} \forall i, \forall j$.

En particulier : La transposée d'une matrice carrée symétrique est identique à elle-même.

Tandis que le transposé d'un vecteur ligne est un vecteur colonne et vice-versa.

Définition 8 : Déterminant d'une matrice carrée.

A toute matrice carrée A correspond un nombre $|A|$ qui peut être obtenu comme suit :

$$|a_{11}| = a_{11}.$$

$$\text{Pour une matrice d'ordre deux } \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

D'une façon générale, le déterminant d'une matrice carrée quelconque peut être calculé grâce au notion du mineur.

Définition 9 : Le mineur de l'élément a_{ij} est le déterminant que l'on obtient en éliminant la i ème ligne et la j ième colonne.

Définition 10 : Le cofacteur de a_{ij} est égal au mineur de a_{ij} multiplié par $(-1)^{i+j}$.

Définition 11 : Le déterminant $|A|$ d'une matrice A peut être calculé en effectuant la somme des produits des différents éléments d'une même ligne ou d'une même colonne par leurs cofacteurs respectifs : $|A| = \sum_{i=1}^p a_{ij}A_{ij}$ pour tout j ou $\sum_{j=1}^p a_{ij}A_{ij}$ pour tout i.

2.5 REGRESSION LINEAIRE MULTIPLE

La régression linéaire multiple est une généralisation, à p variables explicatives, de la régression linéaire simple.

Nous sommes toujours dans le cadre de la régression mathématique : nous cherchons à prédire, avec le plus de précision possible, les valeurs prises par une variable y , dite endogène, à partir d'une série de variables explicatives x_1, x_2, \dots, x_p .

Dans le cas de la régression linéaire multiple, la variable endogène et les variables exogènes sont toutes quantitatives (continues); et le modèle de prédiction est linéaire (Dagnelie 2006).

2.5.1 Equation de régression et objectifs :

Nous disposons de n observations ($i = 1, \dots, n$). L'équation de régression s'écrit

$$y_i = a_0 + a_1 \times x_{i,1} + \dots + a_p \times x_{i,p} + \theta_i$$

où θ_i est l'erreur du modèle, elle exprime, ou résume, l'information manquante dans l'explication linéaire des valeurs de y à partir des x_j (problème de spécifications, variables non prises en compte, etc.); a_0, a_1, \dots, a_p sont les coefficients (paramètres) du modèle à estimer.

La problématique reste la même que pour la régression simple :

- 1-estimer les paramètres a_j en exploitant les observations.
- 2-évaluer la précision de ces estimateurs .
- 3-mesurer le pouvoir explicatif du modèle.
- 4-évaluer l'influence des variables dans le modèle .
- 5-globalement (les p variables en bloc) et, individuellement (chaque variable) .
- 6-évaluer la qualité du modèle lors de la prédiction (intervalle de prédiction) .
- 7-détecter les observations qui peuvent influencer exagérément les résultats (points atypiques).

2.5.2 Notation matricielle :

Nous pouvons adopter une écriture condensée qui rend la lecture et la manipulation de l'ensemble plus facile. Les

équations suivantes :

$$\left\{ \begin{array}{l} y_1 = a_0 + a_1 \times x_{1,1} + \dots + a_p \times x_{1,p} + \theta_1 \\ y_2 = a_0 + a_1 \times x_{2,1} + \dots + a_p \times x_{2,p} + \theta_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ y_n = a_0 + a_1 \times x_{n,1} + \dots + a_p \times x_{n,p} + \theta_n \end{array} \right.$$

peuvent être résumées avec la notation matricielle $\boxed{Y = Xa + \theta}$

avec

Y est de dimension $(n, 1)$

X est de dimension $(n, p + 1)$

a est de dimension $(p + 1, 1)$

et $\theta >: (n, 1)$

la matrice $X = \begin{pmatrix} 1 & x_{1,1} & . & . & . & x_{1,p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & x_{n,1} & . & . & . & x_{n,p} \end{pmatrix}$

la première colonne sert à indiquer que nous procédons à une régression avec constante.

2.5.3 Hypothèses :

Les hypothèses permettent de déterminer : les propriétés des estimateurs (biais, convergence); et leurs lois de distributions (pour les estimations par intervalle et les tests d'hypothèses).

Il existe principalement deux catégories d'hypothèses :

Hypothèses stochastiques

H_1 : les X_j sont non aléatoires, $j = 1, \dots, p$

H₂ : $E(\theta_i) = 0$, le modèle est bien spécifié en moyenne

H₃ : $E(\theta_i^2) = \delta_{\theta^2}$, homoscedasticité

H₄ : $COV(\theta_i, \theta_j) = 0$ pour $i \neq j$, absence d'autocorrélation des erreurs

H₅ : $COV(X_{i,j}, \theta_i) = 0$, les erreurs sont indépendantes des variables exogènes

Hypothèses structurelles :

H₇ : absence de colinéarité entre les variables explicatives, i.e. $X'X$ est régulière, $\det(X'X) \neq 0$ et $(X'X)^{-1}$ existe (remarque : c'est la même chose, $\text{rang}(X) = \text{rang}(X'X) = p + 1$)

H₈ : $\frac{X'X}{n}$ tend vers une matrice finie non singulière lorsque $n \rightarrow +\infty$

H₉ : $n > p + 1$, le nombre d'observations est supérieur au nombre de variables +1.

Notons que s'il y avait égalité, le nombre d'équations serait égal au nombre d'inconnues a_j , la droite de régression passe par tous les points, nous sommes face à un problème d'interpolation linéaire

2.5.4 La méthode des moindres carrés ordinaires :

Estimateur des moindres carrés ordinaires (EMCO) :

Le principe des moindres carrés consiste à rechercher les valeurs des paramètres qui minimisent la somme des carrés des erreurs, à savoir $S = \sum_{i=1}^n \theta_i^2$

En adoptant l'écriture matricielle, nous minimisons donc $S = \theta' \theta$.

Ce qui revient à rechercher les solutions de . Nous disposons de $p + 1$ équations, dites équations normales, à résoudre.

La solution obtenue est l'estimateur des moindres carrés ordinaires, il s'écrit : $\hat{a} = (X'X)^{-1}X'Y$

où X' est la transposée de X

si les x_j sont centrés, $X'X$ correspond à la matrice de variance co-variance des exogènes; s'ils sont centrés et réduits, $X'X$ correspond à la matrice de corrélation

(Dagnelie 2006) .

Propriétés des estimateurs :

Si les hypothèses initiales sont respectées, cet estimateur des MCO (Moindres Carrés Ordinaires) possède d'excellentes propriétés :

il est sans biais, c.-à-d. $E(\hat{a}) = a$

il est convergent, c.-à-d. la variance des estimateurs tend vers zéro lorsque le nombre des observations n tend vers l'infini .

on peut même prouver que l'EMCO est le meilleur estimateur linéaire sans biais c.-à-d. il n'existe pas d'estimateur sans biais de a qui ait une variance plus petite

2.6 LES TESTS STATISTIQUES

2.6.1 Introduction

Un test statistique est appelé à dégager un résultat significatif au milieu d'un ensemble de données expérimentales aléatoires.

La méthodologie des tests consiste à répondre à l'aide de résultats expérimentaux à une question concernant les paramètres de la loi de probabilité des variables aléatoires.

Quatre conditions préalables au calcul d'un test doivent être réunies :

– la question doit être posée de telle sorte qu'il n'y ait que deux réponses possibles : oui et non

– on doit avoir des données chiffrées résultant d'un échantillon ou d'une expérimentation

– ces données doivent pouvoir être considérées comme la réalisation de variables aléatoires dont la forme de la loi de probabilité est connue (Anonyme) .

– la question doit concerner un ou plusieurs paramètres de cette loi.

Une fois posée cette dernière, la réponse du test est :

– soit l'acceptation de l'hypothèse, ce qui signifie que les données ne sont pas en contradiction avec l'hypothèse ;

– soit le rejet de cette hypothèse, ce qui signifie qu'il est très peu probable d'obtenir les résultats que l'on a trouvés si l'hypothèse est vraie, ou encore que les données sont en contradiction avec elle.

En un sens, le test d'hypothèse est une généralisation probabiliste du raisonnement par l'absurde, mais alors que ce dernier met en contradiction logique deux affirmations formelles,

le premier oppose une affirmation formelle (l'hypothèse) avec des résultats du monde réel (les résultats de l'expérience).

2.6.2 Tests de corrélation :

Utilisez cet outil pour calculer les coefficients de corrélation de Pearson, Spearman ou Kendall, entre au moins deux variables, et pour éventuellement déterminer si les corrélations sont significatives ou non. Des visualisations des matrices de corrélation sont aussi proposées.

Trois coefficients sont proposés pour calculer la corrélation entre des variables quantitatives continues, discrètes ou ordinales (les classes devant être représentées par des valeurs respectant l'ordre) :

Analyse Canonique des Correspondances (ACC)

Utilisez l'analyse canonique des correspondances (en anglais, Canonical Correspondence Analysis, ou CCA) pour relier l'analyse d'un tableau de contingence (typiquement un tableau de comptages, croisant sites et espèces) à des variables quantitatives ou qualitatives mesurées sur les mêmes sites.

L'analyse canonique des correspondances (en anglais, Canonical Correspondence Analysis, ou CCA) a été développée dans le cadre d'applications en écologie (Ter Braak, 1986). Néanmoins, cette méthode dont le cadre conceptuel est bien défini, peut être utilisée dans

d'autres domaines. Le géomarketing et les analyses démographiques devraient pouvoir en tirer profit.

2.6.3 Test de chi-deux (χ^2) :

Le mécanisme du test du χ^2 permet de savoir si les écarts constatés entre les distributions à comparer sont imputables ou non au hasard.

$$P(Z \geq \chi^2_\alpha) = \alpha.$$

On considère une distribution expérimentale donnée par un échantillon de taille n .

Les individus de cet échantillon sont classés et on a dénombré la fréquence absolue ou effectif de chaque classe. On note n_i l'effectif observé pour la classe $N^\circ i$. Si on connaît (ou croit connaître) la loi théorique qui suit cette distribution, on est alors capable de calculer les effectifs théoriques de chaque classe. En effet la loi théorique est connue dès lors que les probabilités attachées à chaque classe le sont. On note P_i la probabilité qu'un individu tiré au hasard appartienne à la classe $N^\circ i$. L'effectif théorique associé est alors nP_i .

2.6.4 Tests de comparaison :

Le test de comparaison de moyennes permet de résoudre ce problème en testant une hypothèse nulle formulée par H_0 : « les échantillons proviennent de populations de moyennes identiques » . Contre une hypothèse alternative H_1 : « les échantillons proviennent de deux populations de moyennes différents » ou encore H_1 : « la moyenne de la première population est supérieure à la seconde ».

On aura reconnu un test bilatéral dans le premier cas et un test unilatéral dans le second. En fait, le test de comparaison moyenne permet de comparer de populations. Il pourra conclure, par exemple, à l'efficacité d'un traitement appliqué sur les individus d'une population.

De plus, pour un traitement appliqué deux même population, on vérifiera que cette dernière est homogène par rapport à la variable étudiée .cette démarche trouve un grand nombre d'applications dans domaines agronomiques et biologique pour la sélection d'espèces pures.

Alors, si δ_1 et δ_2 sont connus (ce qui est rare), on calcule la quantité :

$$z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2}}}$$

On compare ensuite cette valeur à en adoptant la règle de décision suivant :

- Si $z < \mu_\alpha$, on accepte H_0
- Si $z \geq \mu_\alpha$, on rejette H_0 en considérant que la différence entre les deux moyennes est significative au risque $\alpha\%$.

Remarques :

- Si δ_1 et δ_2 Sont inconnus, on remplace par leurs estimations s_1 et s_2 pour calculer :

$$z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

La règle de décision est inchangée. On rappelle que $s_1 = \delta_{e1} \frac{n_1}{n_1-1}$ et $s_2 = \delta_{e2} \frac{n_2}{n_2-1}$

D'où la formule alternative pour le calcul de Z :

$$z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\delta_{e1}^2}{n_1-1} + \frac{\delta_{e2}^2}{n_2-1}}}$$

. on rappelle que cette démarche n'est valable que pour les échantillons tels que : $n_1 \geq 30$ et $n_2 \geq 30$ afin que les distributions d'échantillonnage soient réparties selon une loi normale .

. dans le cas d'un test unilatéral (ou l'hypothèse alternative H_1 est par exemple $\ll m_1 > m_2 \gg$) on compare z à v_α définie précédemment.

Test de Student : Le test de Student est un test de comparaison des moyennes dans

le cas des petits échantions (*tailles* ≤ 30).

2.7 AJUSTEMENT ET CORRELATION

2.7.1 Observation :

Lors d'une expérimentation, on peut être amené à différents moments à mesurer différents paramètres liés au phénomène ou processus étudié.

Ainsi des médecins scolaires noteront pour chaque élève d'une école, leur taille et leur poids. A chaque élève est associée une observation. L'ensemble des élèves est parfois appelé aussi une population et chaque élève est un individu.

Relation fonctionnelle :

Etant données deux variables statistiques X et Y , on dit qu'il existe une relation fonctionnelle entre de X vers Y si à chaque valeur de la variable X est associée une valeur et une seule de la variable Y .

Modèle :

S'il existe une relation fonctionnelle entre les variable X et Y et si f est la fonction donnant pour toute valeur x de X la valeur correspondante y de Y , ($Y = f(x)$) on dit que f est un modèle du phénomène étudié.

Valeur observée :

Etant donnée une variable statistique X on appelle valeur (observée de cette variable, toute valeur de cette variable relevée au cours d'une observation.

Valeur théorique :

S'il existe une relation fonctionnelle entre les variable X et Y de X vers Y et si f un modèle de cette relation ($Y = f(X)$).

A toute valeurs x_i de X est associée une valeur $y_i = f(x_i)$.

y_i est la valeur théorique ou valeur expliquée par le modèle.

Variable statistique centrée réduite :

Soit X une variable statistique et \bar{X} sa moyenne, S_X son écart-type.

On appelle variable statistique centrée la variable $X - \bar{X}$.

On appelle variable statistique centrée réduite la variable $Z = \frac{X - \bar{X}}{S_X}$

Variable explicative et variable expliquée :

S'il existe une relation fonctionnelle entre les variable X et Y de X vers Y et si f un modèle de cette relation ($Y = f(X)$) on dit que X est la variable explicative et Y est la variable expliquée.

Nuage de points :

Soit R un repère cartésien du plan.

Si au cours de n observations w_1, w_2, \dots, w_n . on relève les valeurs $X(w_i)$ et $Y(w_i)$ prises par deux variables statistiques X et Y On appelle nuage de points l'ensemble des points du plan dont les coordonnées dans le repère R sont les couples $(X(w_i), Y(w_i))$.

Point moyen :

On appelle point moyen d'un nuage de points, le point dont les coordonnées sont les moyennes arithmétiques des coordonnées des points du nuage.

Chapitre III



Applications Numériques

Chapitre 3

Applications numériques :

3.1 Application 1

3.1.1 Résumé

Cet exemple montre que les jeunes peuplements actuels ont une croissance plus rapide. Ils devraient atteindre une hauteur dominante de 41m, et pourraient donc être récoltés, 30 ans plus tôt que leurs devanciers installés au 19ème siècle. Si en outre on considère que les éclaircies en hêtraie sont actuellement plus fortes que durant la période 1850-1950, le raccourcissement des révolutions atteindra vraisemblablement 40% (90 ans contre 150 ans). La transition des vieux peuplements à croissance lente vers les jeunes à croissance rapide s'étalera sur les 50 prochaines années. Elle nécessitera une adaptation des aménagements forestiers, en particulier par une forte augmentation des surfaces mises en régénération.

3.1.2 Analyse de Régression et de Corrélation

Évolution de la hauteur dominante de deux peuplements du hêtres, le premier né en 1850 et le deuxième né en 1950

Tableau 1 : L'évolution de H_1 en fonction de Age_1 pour un peuplement replanté en 1850. :

H_1	Age_1	H_1	Age_1	H_1	Age_1	H_1	Age_1
5	14	20	53	30	86	H_1	Age_1
6	18	21	54	30.5	90	36	120
8	21	22	59	31	93	36.5	124
9	22	23	61	31.5	95	37	126
10	26	24	65	32	97	37.5	130
11	28	25	66	32.5	100	38	133
12	30	26	70	33	103	38.5	136
14	32	27	72	33.5	106	39	141
16	38	28	74	34	108	40	143
17	46	28.5	79	34.5	112	40.5	145
18	49	29	83	35	115	41	147
19	50	29.5	84	35.5	117		

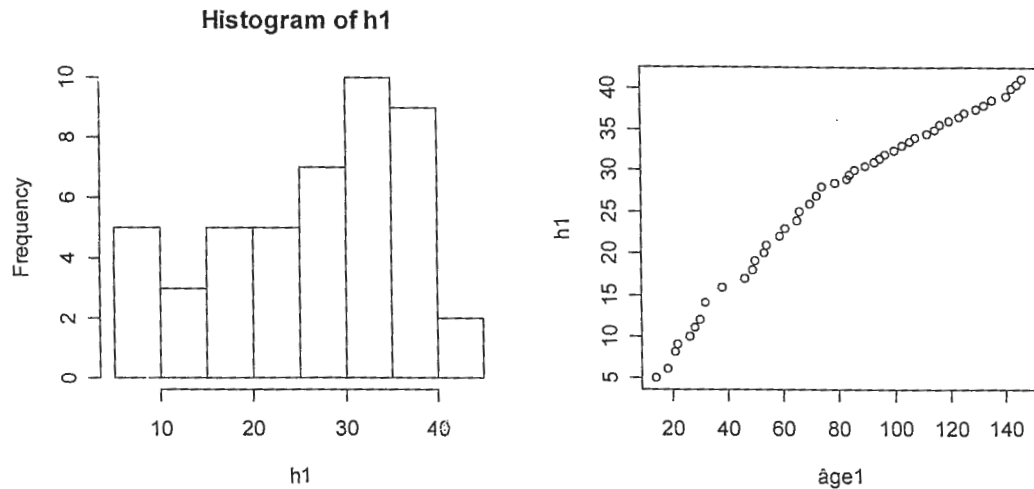


Fig 1: Histogramme de H_1 Fig 2: Nuage de points (\hat{age}_1, h_1)

Tableau 4 : Statistique descriptive de la hauteur

$P_1(1850)$	moyenne	variance	ecart-type	médiane	MIN	Q1	3Q	MAX
	26.63	107.12	210.35	29.25	5.00	19.25	34.87	41.00

$cov(H_1, Age_1)$	401.62
$var(H_1)$	107.12
$var(Age_1)$	1555.99

L'équation de la droite de régression est donnée par :

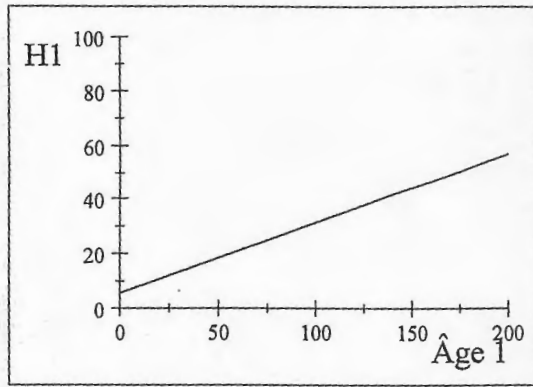
$$H_1 = a \times Age_1 + b$$

$$\text{où : } a = \frac{cov(h_1, age_1)}{var(age_1)} \quad ; \quad a = 0.26$$

$$b = \overline{H_1} - a \overline{Age_1} \quad ; \quad b = 5.61$$

Par la suite , cette équation est donnée par :

$$H_1 = 0.26 \times Age_1 + 5.61$$



$$H_1 = 0.26 \times \text{Age}_1 + 5.61$$

Fig 3 : Droite de régression linéaire de la hauteur sur l'âge

l'équation de regression de l'age sur la hauteur est donnée par :

$$a' = \frac{\text{cov}(H_1, \text{age}_1)}{\text{var}(H_1)} ; \quad a' = 3.75$$

$$b' = \overline{\text{Age}_1} - a' \overline{H_1} ; \quad b' = -18.08$$

l'équation est donnée :

$$\boxed{\text{Age}_1 = 3.75 \times H_1 - 18.08}$$

le coefficient de corrélation qui mesure l'intensité (poids) , de la régression linéaire existante entre l'age et l' hauteur est donnée par : $\sqrt{a \times a'} = 0.98$

c'est a dire : 98% de la relation entre l'age et l'hauteur est expliqué par cette equation.

Tableau 2 : L'évolution de H_2 en fonction de Age_2 pour un peuplement implantée en 1950 :

H ₂	Age ₂	H ₂	Age ₂	H ₂	Age ₂
5.5	14	24	51	35.5	86
8	16	25.5	54	36	89
10	20	27	57	37	92
11	24	28	59	37.5	95
13	26	29	62	38	99
15	29	30	65	38.5	101
16	32	31	69	39	105
17	34	32	71	40	108
19	37	33	73	40.5	111
20	40	34	76	41	114
21.5	43	34.5	79	42	116
23	48	35	82	43	120

Tableau 3 : Statistique descriptive de la hauteur

P ₂ (1950)	moyenne	variance	ecart-type	médiane	MIN	Q1	3Q	MAX
	16.32	38.71	6.22	16.5	5.50	11.50	21.12	25.50

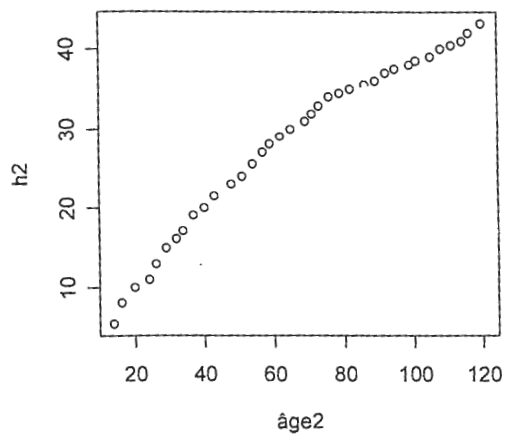
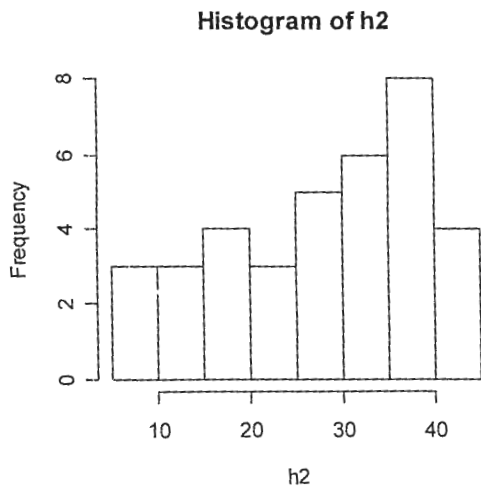


Fig 4 : Histogramme de H₂

Fig5: Nuage de points (âge2, h2)

$cov(H_2, age_2)$	401.62
$var(H_2)$	38.71
$var(Age_2)$	163.33

l'équation de la droite de régression est donnée par :

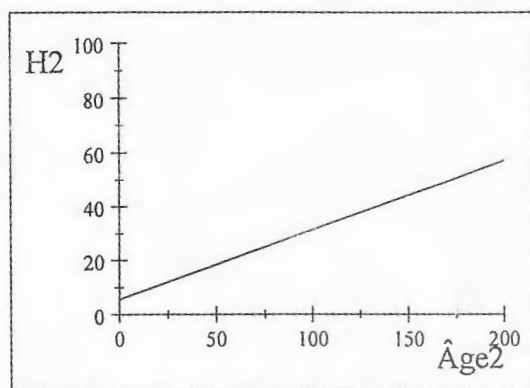
$$H_2 = a \times Age_2 + b$$

$$\text{où } a = \frac{cov(H_2, age_2)}{var(age_2)} \quad ; \quad a = 0.26$$

$$b = \overline{H_2} - a \overline{Age_2} \quad ; \quad b = 5.61$$

Par la suite , cette équation est donnée par :

$$H_2 = 0.26 \times Age_2 + 5.61$$



$$H_2 = 0.26 (Age_2) + 5.61$$

Fig 6 : Droite de régression linéaire de la hauteur sur l'Age

l'équation de régression de l'age sur la hauteur est donnée par :

$$a' = \frac{cov(H_2, age_2)}{var(H_2)} \quad ; \quad a' = 3.75$$

$$b' = \overline{Age_2} - a' \overline{H_2} \quad ; \quad b' = -18.08$$

l'équation est donnée :

$$\text{Age}_2 = 3.75 \times H_2 - 18.08$$

le coefficient de corrélation qui mesure l'intensité (poids) , de la régression linéaire existante entre l'age et l' hauteur est donnée par : $\sqrt{a \times a'} = 0.98$

c'est -à-dire :98%de la relation entre l'Age et l'hauteur est expliqué par cette equation.

Tableau 4 : l'évolution de H_2 en fonction du $\log(\text{Age}_2)$ pour un peuplement inplantée en1950 :

H_2	5.5	8.0	10.0	11.0	13.0	15.0	16.0
$\log(\text{Age}_2)$	2.64	2.77	3.10	3.18	3.26	3.37	3.43

H_2	17.0	19.0	20.0	21.5	23.0	24.0	25.5
$\log(\text{Age}_2)$	3.53	3.61	3.69	3.76	3.83	3.93	3.99

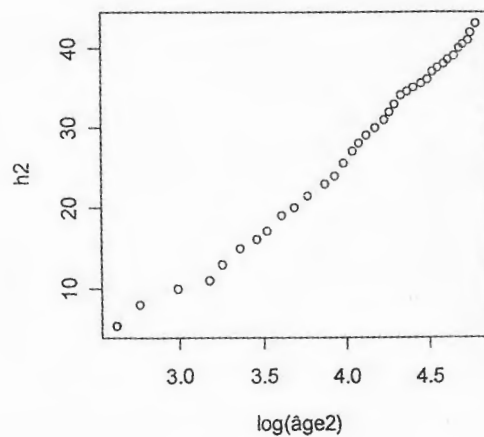


Fig 7 : Nuage de points de h_2 en fonction du $\log(\hat{\text{age}}_2)$

$cov(H_2, \log(Age_2))$	2.61
$var(H_2)$	38.71
$var(\log(Age_2))$	0.18

l'équation de la droite de régression est donnée par :

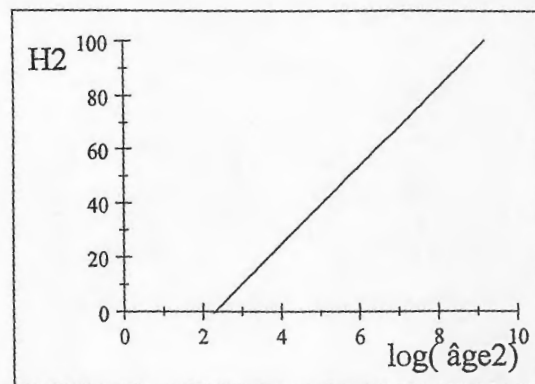
$$H_2 = a \times \log(Age_2) + b.$$

$$\text{ou} : a = \frac{cov(H_2, \log(Age_2))}{var(\log(Age_2))} ; a = 14.61$$

$$b = \overline{H_2} - a \overline{\log(Age_2)} ; b = -33.80$$

Par la suite , cette équation est donnée par :

$$H_2 = 14.61 \times \log(Age_2) - 33.80$$



$$H_2 = 14.61 \times \log(Age_2) - 33.80$$

Fig 8 : Droite de régression linéaire de la hauteur sur log(âge)

l'équation de regression de l'âge sur la hauteur est donnée par :

$$a' = \frac{cov(H_2, \log(Age_2))}{var(H_2)} ; a' = 0.07$$

$$b' = \overline{\log(Age_2)} - a' \overline{H_2} ; b' = 2.33$$

l'équation est donnée :

$$\log(\text{Age}_2) = 0.07 \times H_2 + 2.33$$

le coefficient de corrélation qui mesure l'intensité (poids) , de la régression linéaire existante entre l'age et l' hauteur est donnée par : $\sqrt{a \times a'} = 0.98$

c'est -à -dire : 98%de la relation entre l'âge et la hauteur est expliquée par cette equation.

Tableau 5 : Statistique descriptive de la l'Age

	moyenne	variance	ecart-type	médiane	MIN	Q1	3Q	MAX
$P_1(1850)$	81.76	1555.99	39.52	83.50	14.00	50.75	114.25	147.00
$P_2(1950)$	33.35	163.33	12.78	32.50	14.00	24.50	42.25	54.00
$P_2'(1950)$	3.43	0.18	0.42	3.48	2.64	3.20	3.74	3.99

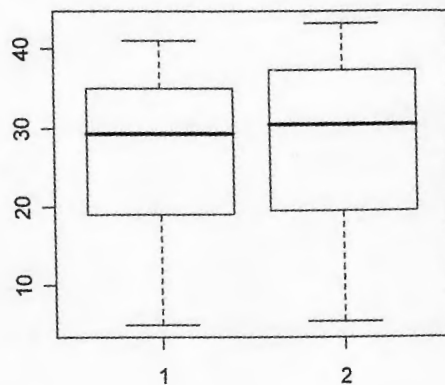


Fig 9 : Diagramme en Boîtes de h1 et h2

3.1.3 Analyse inférentielle

Tests d'ajustement

tableau 6 : Tables des valeurs théoriques et observées pour le p_2

$H_2(obs)$	5.5	8	10	11	13	15	16
$Age_2(obs)$	14	16	20	24	26	29	31
$H_2(theo)$	-23.14	-19.06	-10.90	-2.74	1.34	7.46	11.54
$H_2(obs)$	17	19	20	21.5	23	24	25.5
$Age_2(obs)$	34	37	40	43	48	51	54
$H_2(theo)$	17.66	23.78	29.9	36.02	46.22	52.34	58.46

$$\chi_{calculée(5\%)}^2 = \sum_1^{14} \frac{(H_{obs} - H_{theo})^2}{H_{theo}} = 192.10$$

formulation des hypothèses :

H_0 : hypothèse nulle = la différence entre les valeurs observées et théoriques n'est pas significative

H_1 : hypothèse alternative (contraire) la différence est significative.

l'acceptation de H_0 équivalant à l'adéquation du modèle ; c'est-à-dire que le modèle proposée est adéquat.

le premier modèle $\chi_{tab}^2 = 24.7 \ll \chi_{cal}^2 = 192.10$ se qui conduit au rejet de H_0

tableau 7 : Tables des valeurs théoriques et observées pour le p_1

$H_1(\text{obs})$	$\text{Age}_1(\text{obs})$	$H_1(\text{theo})$
5	14	9.24
6	18	10.28
8	21	11.06
9	22	11.32
10	26	12.36
11	28	12.88
12	30	13.40
14	32	13.92
15	38	15.48
17	46	17.56
18	49	18.34
19	50	18.60

$H_1(\text{obs})$	$\text{Age}_1(\text{obs})$	$H_1(\text{theo})$
20	53	19.38
21	54	19.64
22	59	20.94
23	61	21.46
24	65	22.50
25	66	22.76
26	70	23.80
27	72	24.32
28	74	24.84
28.5	79	26.14
29	83	27.17
29.5	84	27.47

$H_1(\text{obs})$	$\text{Age}_1(\text{obs})$	$H_1(\text{theo})$
30	86	27.96
30.5	90	29
31	93	29.78
31.5	95	30.3
32	97	30.82
32.5	100	31.6
33	103	32.38
33.5	106	33.16
34	108	33.68
34.5	112	34.72
35	115	35.5
35.5	117	36.02

$H_1(\text{obs})$	$\text{Age}_1(\text{obs})$	$H_1(\text{theo})$
36	120	36.8
36.5	124	37.84
37	126	38.36
37.5	130	39.4
38	133	40.18
38.5	136	40.96
39	141	42.26
40	143	42.72
40.5	145	43.30
41	147	43.82

$$\chi^2 = \sum_1^{46} \frac{(H_{obs} - H_{theo})^2}{H_{theo}} = 9.65$$

formulation des hypothése :

H_0 : hypothése nulle = la différence entre les valeur observées et théoriques n'est pas significative

H_1 : hypothése alternative (contraire) la différence est significative.

l'acceptation de H_0 equivalant a l'adequation du modèle; c'est-a-dire que le modèle proposé est adéquat.

le deuxième modèle $\chi_{46;0.05}^2(\text{théorique}) = 61.65 \gg \chi^2(\text{observée}) = 9.65$

On accepte H_0 . c est a dire il n y a pas de difference significative entre les valeurs theoriques et observees

tableau 8 : Tables des valeurs théoriques et observées pour le p_2

$H_2(\text{obs})$	5.5	8	10	11	13	15	16
$\log(\text{Age}_2)$	2.64	2.77	3.10	3.18	3.26	3.37	3.43
$H_2(\text{theo})$	4.77	6.67	11.49	12.66	13.83	15.43	16.31
$H_2(\text{obs})$	17	19	20	21.5	23	24	25.5
$\log(\text{Age}_2)$	3.53	3.61	3.69	3.76	3.87	3.93	3.99
$H_2(\text{theo})$	17.77	18.94	20.11	21.13	22.74	23.62	24.74

$$\chi^2 = \sum_1^{14} \frac{(H_2\text{obs} - H_2\text{theo})^2}{H_2\text{theo}} = 0.95$$

formulation des hypothése :

H_0 : hypothése nulle = la différence entre les valeur observées et théoriques n'est pas significative

H_1 : hypothése alternative (contraire) la différence est significative.

l'acceptation de H_0 equivalant a l'adequation du modèle; c'est-a-dire que le modèle proposée est adéquat.

$$\chi^2_{13;0.05}(\text{théorique}) = 23.68 \gg \chi^2(\text{observée}) = 0.95$$

On accepte H_0 . c est a dire il n y a pas de difference significative entre les valeurs theoriques et observées

Test de comparaison de Student

tableau 9 : Table des observations de la hauteur des deux peuplements pendant 55 ans

H1	05	06	08	09	10	11	12	14	16	17
H1	25.0	26.0	27.0	28.0	28.5	29.0	29.5	30.0	30.5	31.0
H1	35.0	35.5	36.0	36.5	37.0	37.5	38.0	38.5	39.0	40.0
H2	5.5	8.0	10.0	11.0	13.0	15.0	16.0	17.0	19.0	20.0
H2	30.0	31.0	32.0	33.0	34.0	34.5	35.0	35.5	36.0	37.0
H2	42.0	43.0	-	-	-	-	-	-	-	-
H1-H2	-0.5	-2	-2	-2	-3	-4	-4	-3	-3	-3
H1-H2	-5	-5	-5	-5	-5.5	-5.5	-5.5	-5.5	-5.5	-6
H1-H2	-7	-7.5	-	-	-	-	-	-	-	-

18.0	19.0	20.0	21.0	22.0	23.0	24.0
31.5	32.0	32.5	33.0	33.5	34.0	34.5
40.1	41.0	-	-	-	-	-
21.5	23.0	24.0	25.5	27.0	28.0	29.0
37.5	38.0	38.5	39.0	40.0	40.5	41.0
-	-	-	-	-	-	-
-3.5	-4.0	-4.0	-4.5	-5.0	-5.0	-5.0
-6.0	-6.0	-6.0	-6.0	-6.5	-6.5	-6.5

$$T = \frac{|\bar{x}_1 - \bar{x}_2|}{\sigma^2 \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 0.60 \quad / \quad \sigma^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1+n_2}$$

Dans la table de la loi normale. la valeur theorique qui corespond au risque de 5% est 1.64 \gg à la valeur calculer 0.60.

c'est-a-dire on ne peut pas conclur à une différence significative entre les deux peuplements

3.1.4 Conclusion

Malgré la forte corrélation existante entre l'évolution de la hauteur et l'Age, les tests statistiques que nous ayons employé ne permettent pas la conclusion d'une différence significative entre les deux peupement.

Si l'on veut être certain de trouver le meilleur modele qui explique le comportement de la variable dépendante en fonction des variables indépendantes deux autres procédures existent.

Partant du modèle complet, on élimine progressivement les variables les moins significatives. La variable la moins significative est celle dont la suppression provoque la plus petite baisse de la statistique de Fisher. l'approche en avant : on part d'un modèle avec une seule variable explicative. La variable retenue initialement est la plus significative (celle qui a le plus grand coefficient de corrélation avec la variable explicative). On ajoute ensuite une seconde variable (celle qui permet le plus fort accroissement du test de Fisher).

Il est bon de rappeler, dans l'ensemble, nos connaissances sont beaucoup plus limitées en ce qui concerne la robustesse des méthodes d'analyse à plusieurs variables, c'est-à-dire leur non sensibilité ou non respect de leurs conditions d'application.

Enfin, les difficultés d'exécution des calcules. Sauf peut-être pour des cas aussi simples, ou aussi simplifiés, l'ordinateur est ici un outil indispensable.

3.2 Application 2 :

3.2.1 Résumé

Ce paragraphe est consacré à l'étude d'un exemple illustrant la mise en œuvre de l'ACR polynomiale sur un ensemble des données réelles. Nous avons choisi des données écologiques recueillies à 20 sites d'étange de Thau (cote méditerranéenne de la France) dans le cadre du programme écothau (amanieu et al. (1)). Les données, présentées au tableau ci-dessous, sont tirées de Legendre et Legendre

Dans ce tableau, les deux variables-réponse Bna et Ma représentent des quantités de deux types de bactéries hétérotrophes aérobies dans les sites examinés, après transformation logarithmique des abondances; elles forment la matrice des variables dépendantes Y . La matrice des variables explicatives X contient deux variables chimiques, NH_4 et la production bactérienne, ainsi qu'une variable spatiale qui représente les coordonnées des sites le long du grand axe géographique de THAU.

L'ACR linéaire et polynomiale ont été calculées pour les données du tableau. Les résultats ci-dessous ont été obtenus : deux axes canoniques ont été trouvés par l'ACR linéaire et polynomiale. La méthode polynomiale explique 67.62% de la variation totale de la matrice des variables-réponse Y alors que la méthode linéaire n'explique que 18.70% de la variation totale de Y . Les résultats des tests de signification des deux modèles sont présentés plus bas; nous verrons que seule le modèle polynomiale est adéquat pour expliquer les variables-réponse Bna et Ma .

3.2.2 Analyse de régression Multiple

Tableau 10 : Données recueillies à 20 sites de l'étange de Thau le 25 octobre 1988.

<i>sites</i>	<i>Bna</i>	<i>MA</i>	<i>NH₄</i>	<i>Prod.bact.</i>	<i>var.spatiale</i>
N^0	Y_1	Y_2	x_1	x_2	x_3
1	4.615	10.003	0.307	0.274	-9.4173
2	5.226	9.999	0.207	0.213	-7.1865
3	5.081	9.636	0.140	0.134	-5.8174
4	5.278	8.331	1.371	0.177	-6.8322
5	5.756	8.929	1.447	0.091	-4.6014
6	5.328	8.839	0.668	0.272	-4.2471
7	4.263	7.384	0.300	0.460	-1.8632
8	5.442	8.023	0.329	0.253	-0.4940
9	5.328	8.294	0.207	0.235	0.8751
10	4.663	7.883	0.223	0.362	-0.1398
11	6.775	9.941	0.788	0.824	-1.1546
12	5.442	8.657	1.112	0.419	0.2145
13	5.421	8.117	1.273	0.398	4.9824
14	5.602	8.117	0.956	0.172	3.9676
15	5.442	8.487	0.708	0.141	2.4602
16	5.303	7.955	0.637	0.360	6.3515
17	5.602	10.545	0.519	0.261	5.8441
18	5.505	9.687	0.247	0.450	4.8293
19	6.019	8.700	1.664	0.287	4.6762
20	5.464	10.240	0.182	0.510	6.5527

Matrice1 (y_1, x_1, x_2, x_3) :

	y_1	x_1	x_2	x_3
4.615	0.307	0.274	-9.4173	
5.226	0.207	0.213	-7.1865	
5.081	0.140	0.134	-5.8174	
5.278	1.371	0.177	-6.8322	
5.756	1.447	0.091	-4.6014	
5.328	0.668	0.460	-1.8632	
4.263	0.300	0.272	-4.2471	
5.442	0.329	0.253	-0.4940	
5.328	0.207	0.235	0.8751	
4.663	0.223	0.362	-0.1398	
6.775	0.788	0.824	-1.1546	
5.442	1.112	0.419	0.2145	
5.421	1.273	0.398	4.9824	
5.602	0.956	0.172	3.9676	
5.442	0.708	0.141	2.4602	
5.303	0.637	0.360	6.3515	
5.602	0.519	0.261	5.8441	
6.505	0.247	0.450	4.8293	
6.019	1.664	0.287	4.6762	
5.464	0.182	0.510	6.5527	

Matrice de Corrélations entre les variables $y_1; x_1; x_2$ et x_3 :

$$MC = \begin{pmatrix} & y_1 & x_1 & x_2 & x_3 \\ y_1 & 1.00 & 0.35 & 0.35 & 0.38 \\ x_1 & 0.35 & 1.00 & -0.12 & 0.10 \\ x_2 & 0.35 & -0.12 & 1.00 & 0.29 \\ x_3 & 0.38 & 0.10 & 0.29 & 1.00 \end{pmatrix}$$

D'après la matrice de corrélation nous navons remarque acune variables dominante

c'est-à-dire : x_1, x_2, x_3 expliquent approximativement le même taux de variation (35%, 35% et 38%), d'autre part nous remarquons que les deux variables x_1, x_2 sont faiblement corrélées négativement c'est-à-dire l'évolution de l'une des deux défavorise l'autre ; la variable x_3 est positivement corrélée avec x_2 et à un degré moins avec x_1 ; la matrice de corrélation nous donne les poids (intensités) des relations existantes entre les différentes variables (à expliquer et explicatives)

$X = \text{matrice}(id, x_1, x_2, x_3)$

$id = (1, 1)$

id	x_1	x_2	x_3
1	0.307	0.274	-9.4173
1	0.207	0.213	-7.1865
1	0.140	0.134	-5.8174
1	1.371	0.177	-6.8322
1	1.447	0.091	-4.6014
1	0.668	0.272	-4.2471
1	0.300	0.460	-1.8632
1	0.329	0.253	-0.4940
1	0.207	0.235	0.8751
1	0.223	0.362	-0.1398
1	0.788	0.824	-1.1546
1	1.112	0.419	0.2145
1	1.273	0.398	4.9824
1	0.956	0.172	3.9676
1	0.708	0.141	2.4602
1	0.637	0.360	6.3515
1	0.519	0.261	5.8441
1	0.247	0.450	4.8293
1	1.664	0.287	4.6762
1	0.182	0.510	6.5527

★*IDENTITE*

la ligne des commandes suivante nous permet de calculer le coefficient de régression multiple

`solve(t(x)*%(x))*%(t(x))*%y1`

constante = 4.79

$x_1 = 0.43$

$x_2 = 1.11$

$x_3 = 0.03$

$Y_1 = 4.79 + 0.43x_1 + 1.11x_2 + 0.03x_3$

$\text{Log}(Bna) = 4.79 + 0.43\log(NH_4) + 1.11\log(\text{prod.bact}) + 0.03\log(\text{var.spaciale})$

matrice2 (y_2, x_1, x_2, x_3)

y_2	x_1	x_2	x_3
10.003	0.307	0.274	-9.4173
9.999	0.207	0.213	-7.1865
9.636	0.140	0.134	-5.8174
8.331	1.371	0.177	-6.8322
8.929	1.447	0.091	-4.6014
8.839	0.668	0.272	-4.2471
7.784	0.300	0.460	-1.8632
8.023	0.329	0.253	-0.4940
8.294	0.207	0.235	0.8751
7.883	0.223	0.362	-0.1398
9.741	0.788	0.824	-1.1546
8.657	1.112	0.419	0.2145
8.117	1.273	0.398	4.9824
8.117	0.956	0.172	3.9676
8.487	0.708	0.141	2.4602
7.955	0.637	0.360	6.3515
10.545	0.519	0.261	5.8441
9.687	0.247	0.450	4.8293
8.700	1.664	0.287	4.6762
10.240	0.182	0.510	6.5527

Matrice de corrélation multiple :

$$\begin{pmatrix} & y_2 & x_1 & x_2 & x_3 \\ y_2 & 1.00 & -0.29 & 0.16 & -0.12 \\ x_1 & -0.29 & 1.00 & -0.12 & 0.10 \\ x_2 & 0.16 & -0.12 & 1.00 & 0.29 \\ x_3 & -0.12 & 0.10 & 0.29 & 1.00 \end{pmatrix}$$

D'après la matrice de corrélation nous navons remarquée aucune variables dominante c'est-a-dire : x_1, x_2, x_3 expliquent approximativement le même taux de variation (29%,16%,12%).d'autre part nous remarquons que les deux variables x_1, x_2 sont faiblement corrélés négativement c'est-a-dire l'évolution de l'une des deux défavorisent l'autre la variable x_3 , est positivement corrélés avec x_2 et un degré moins avec x_1 .la matrice de corrélation nous donne les poids (intensités)des relation existante entre les différentes variables (a expliquée et explicative)

En utilisant le langage R :

```
solve(t(x) %*% (x)) %*% (t(x)) %*% y2
```

constante = 8.93

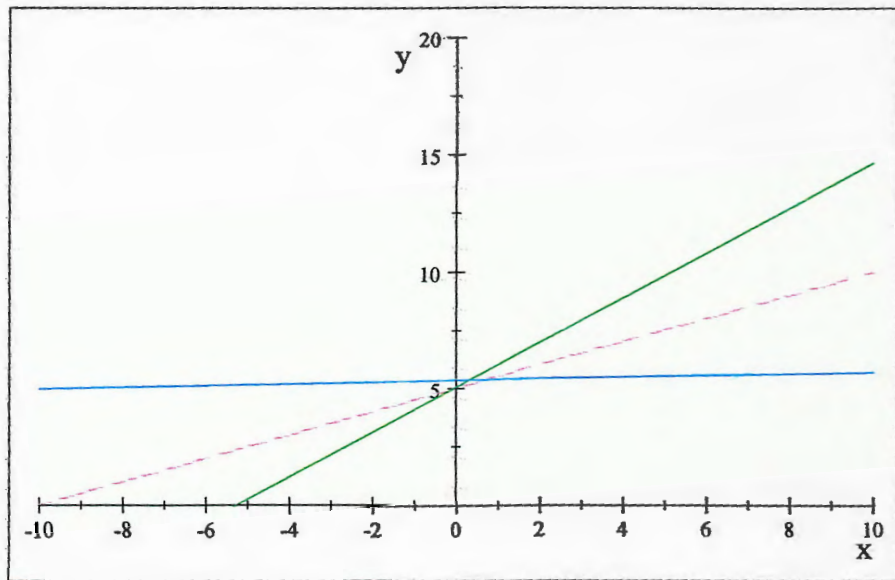
$x_1 = -0.46$

$x_2 = 0.87$

$x_3 = -0.025$

```
Y2 = 8.93 - 0.46x1 + 0.87x2 - 0.025x3
```

```
Log(Bna) = 8.93 - 0.46 log(NH4) + 0.87 log(prod.bact) - 0.025 log(var.spaciale)
```

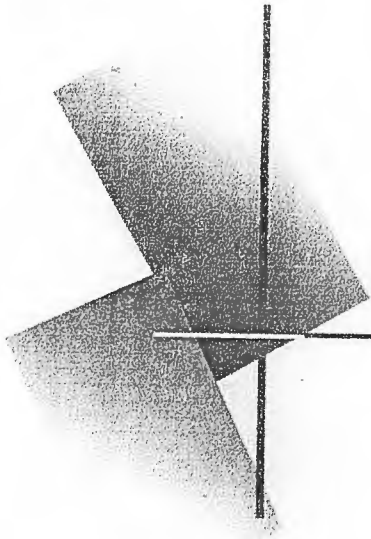


- Bna = 5.04 + 0.5 NH₄.
- - - Bna = 5.08 + 0.96 prod .bact.
- Bna = 5.38 + 0.033 var spaciae.

Fig10 : les régressions linéaires

Conclusion :

Un modèle multiple comportant es trois variable explicatives x_1, x_2 méritent d'être proposé car nous voyons les trois variables interagissent avec le même poids. Dans le cas ou l'une des trois variables est dominante au sens de corrélation le modèle proposé ne semble pas réaliste, et un modèle polynomiale en fonction de la variable dominante sera mieux explicatif.



Conclusion

Conclusion

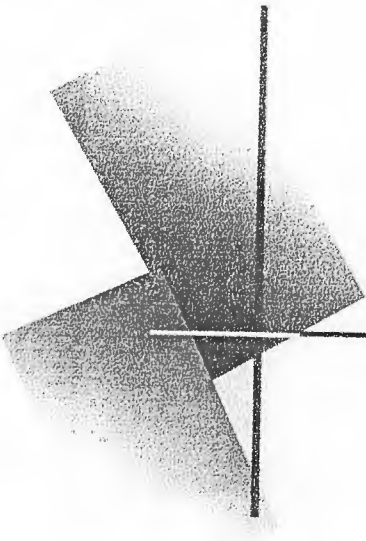
Malgré la forte corrélation existante entre l'évolution de la hauteur et l'Age, les tests statistiques que nous ayons employé ne permettent pas la conclusion d'une différence significative entre les deux peuplement.

Si l'on veut être certain de trouver le meilleur modèle qui explique le comportement de la variable dépendante en fonction des variables indépendantes deux autres procédures existent.

Partant du modèle complet, on élimine progressivement les variables les moins significatives. La variable la moins significative est celle dont la suppression provoque la plus petite baisse de la statistique de Fisher. L'approche en avant : on part d'un modèle avec une seule variable explicative. La variable retenue initialement est la plus significative (celle qui a le plus grand coefficient de corrélation avec la variable explicative). On ajoute ensuite une seconde variable (celle qui permet le plus fort accroissement du test de Fisher).

Il est bon de rappeler, dans l'ensemble, nos connaissances sont beaucoup plus limitées en ce qui concerne la robustesse des méthodes d'analyse à plusieurs variables, c'est-à-dire leur non sensibilité ou non respect de leurs conditions d'application.

Enfin, les difficultés d'exécution des calculs. Sauf peut-être pour des cas aussi simples, ou aussi simplifiés, l'ordinateur est ici un outil indispensable.



Références

Bibliographie

- [1] Alexandre Popier, Olivier Wintenberger. 2006-2007.
- [2] A. Mackenzi, A.S. Ball. Juillet 2000 . L'essentiel en écologie. Paris. 81 -172
- [3] Bernard Saugier ,Claude Lobry; Jérôme Harmand, Françoise Burel ,Pierre Auger ,Yann Martineau
- [4] Calmant Philippe. cours biostatistique.2004.18-20.
- [5] Cours Université de Nice SV1. Département de Mathématiques Appliquées à la Biologie . 2005-2006
- [6] Dagnelie. statistique théorique et appliquée .2006.(tomme 2).
- [7] François Goreaud . Mastér écologie biodiversité et évolution – paris 6 –ENS-INAPG. module : écologie théorique et statistique de l'écologie . cour .2004.
- [8] France Mentré.cours 6 biostatistiques.université de Paris. 2007-2008.
- [9] Jean-Pierre Villa. UMR analyse de système et Biométrie INRA-ENSA ,M .2004.
- [10] Pierre Auger. université Claud Bernard Lyon 1.(oct. 2002) . 47-50.
- [11] Robert Barbault . Ecologie générale : structure et fonctionnement de la biosphère . Paris. 5eme édition. 2000.
- [12] Roger Dajoz .précis d'écologie. juin 2006 .127-148.
- [13] Roger Pratel, Olivier Gimenez . centre d'écologie fonctionnelle et évolutive – CNRS . 2004.
- [14] Vivien Rossi, laboratoire de biostatistique Institut universitaire de recherche clinique Montpellier

[15] Séminaire IREM-DMI du 29/02/08

Sites web

[16] www.devinci.fr

[17] www.edreams.fr

[18] www.geoaznr.unice.fr

[19] www.gardengold.fr

[20] www.unice/zetitique

[21] <http://www.chez.com/laccessoireantiquites/>

[22] www.multimania.com/coligny/

[23] <http://kourou.cirad.fr/>

[24] <http://www.onf.fr/>

[25] <http://www.plustech.fi/plustech.nsf/?Open>

[26] [htm://www.lisc.clcrmont.comagrcf.Fr/modelisation](http://www.lisc.clcrmont.comagrcf.Fr/modelisation)



Annexe



Tableau 11.7 – Loi normale centrée réduite

Soit la variable, à valeurs réelles, $X \sim \mathcal{N}(0, 1)$, de densité ρ . La table donne, pour différentes valeurs de u positives :

$$F(u) = \mathbb{P}(X < u) = \int_{-\infty}^u \rho(x) dx$$

Exemple : $F(1.35) = \mathbb{P}(X < 1.35) = 0.9115 = 0.5 + \frac{1}{2}\theta(1.35)$ (voir table 11.6)

u	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

Pour $u < 0$, on a $\mathbb{P}(X < u) = 1 - F(-u)$

Tableau 11.10 – Lois du chi-deux

Si ν est le nombre de degrés de liberté d'une variable χ^2 de densité rho , si u est un nombre positif et si on pose :

$$F(u) = \mathbb{P}(T < u) = P$$

la table donne u pour différentes valeurs de ν et de P .

$\nu \backslash P$	0.01	0.025	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.975	0.99
1	0.00	0.00	0.00	0.01	0.06	0.14	0.27	0.45	0.70	1.07	1.64	2.70	3.84	5.02	6.63
2	0.02	0.05	0.10	0.21	0.44	0.71	1.02	1.38	1.83	2.40	3.21	4.60	5.99	7.37	9.21
3	0.11	0.21	0.35	0.58	1.00	1.42	1.86	2.36	2.94	3.66	4.64	6.25	7.81	9.34	11.34
4	0.29	0.48	0.71	1.06	1.64	2.19	2.75	3.35	4.04	4.87	5.98	7.77	9.48	11.14	13.27
5	0.55	0.83	1.14	1.61	2.34	2.99	3.65	4.35	5.13	6.06	7.28	9.23	11.07	12.83	15.08
6	0.87	1.23	1.63	2.20	3.07	3.82	4.57	5.34	6.21	7.23	8.55	10.64	12.59	14.44	16.81
7	1.23	1.68	2.16	2.83	3.82	4.67	5.49	6.34	7.28	8.38	9.80	12.01	14.06	16.01	18.47
8	1.64	2.17	2.73	3.48	4.59	5.52	6.42	7.34	8.35	9.52	11.03	13.36	15.50	17.53	20.09
9	2.08	2.70	3.32	4.16	5.38	6.39	7.35	8.34	9.41	10.65	12.24	14.68	16.91	19.02	21.66
10	2.55	3.24	3.94	4.86	6.17	7.26	8.29	9.34	10.47	11.78	13.44	15.98	18.30	20.48	23.20
11	3.05	3.81	4.57	5.57	6.98	8.14	9.23	10.34	11.52	12.89	14.63	17.27	19.67	21.92	24.72
12	3.57	4.40	5.22	6.30	7.80	9.03	10.18	11.34	12.58	14.01	15.81	18.54	21.02	23.33	26.21
13	4.10	5.00	5.89	7.04	8.63	9.92	11.12	12.33	13.63	15.11	16.98	19.81	22.36	24.73	27.68
14	4.66	5.62	6.57	7.78	9.46	10.82	12.07	13.33	14.68	16.22	18.15	21.06	23.68	26.11	29.14
15	5.22	6.26	7.26	8.54	10.30	11.72	13.02	14.33	15.73	17.32	19.31	22.30	24.99	27.48	30.57
16	5.81	6.90	7.96	9.31	11.15	12.62	13.98	15.33	16.77	18.41	20.46	23.54	26.29	28.84	31.99
17	6.40	7.56	8.67	10.08	12.00	13.53	14.93	16.33	17.82	19.51	21.61	24.76	27.58	30.19	33.40
18	7.01	8.23	9.39	10.86	12.85	14.43	15.89	17.33	18.86	20.60	22.75	25.98	28.86	31.52	34.80
19	7.63	8.90	10.11	11.65	13.71	15.35	16.85	18.33	19.91	21.68	23.90	27.20	30.14	32.85	36.19
20	8.26	9.59	10.85	12.44	14.57	16.26	17.80	19.33	20.95	22.77	25.03	28.41	31.41	34.16	37.56
21	8.89	10.28	11.59	13.23	15.44	17.18	18.76	20.33	21.99	23.85	26.17	29.61	32.67	35.47	38.93
22	9.54	10.98	12.33	14.04	16.31	18.10	19.72	21.33	23.03	24.93	27.30	30.81	33.92	36.78	40.28
23	10.19	11.68	13.09	14.84	17.18	19.02	20.69	22.33	24.06	26.01	28.42	32.00	35.17	38.07	41.63
24	10.85	12.40	13.84	15.65	18.06	19.94	21.65	23.33	25.10	27.09	29.55	33.19	36.41	39.36	42.97
25	11.52	13.11	14.61	16.47	18.93	20.86	22.61	24.33	26.14	28.17	30.67	34.38	37.65	40.64	44.31
26	12.19	13.84	15.37	17.29	19.82	21.79	23.57	25.33	27.17	29.24	31.79	35.56	38.88	41.92	45.64
27	12.87	14.57	16.15	18.11	20.70	22.71	24.54	26.33	28.21	30.31	32.91	36.74	40.11	43.19	46.96
28	13.56	15.30	16.92	18.93	21.58	23.64	25.50	27.33	29.24	31.39	34.02	37.91	41.33	44.46	48.27
29	14.25	16.04	17.70	19.76	22.47	24.57	26.47	28.33	30.28	32.46	35.13	39.08	42.55	45.72	49.58
30	14.95	16.79	18.49	20.59	23.36	25.50	27.44	29.33	31.31	33.53	36.25	40.25	43.77	46.97	50.89
31	15.65	17.53	19.28	21.43	24.25	26.43	28.40	30.33	32.34	34.59	37.35	41.42	44.98	48.23	52.19
32	16.36	18.29	20.07	22.27	25.14	27.37	29.37	31.33	33.38	35.66	38.46	42.58	46.19	49.48	53.48
33	17.07	19.04	20.86	23.11	26.04	28.30	30.34	32.33	34.41	36.73	39.57	43.74	47.39	50.72	54.77
34	17.78	19.80	21.66	23.95	26.93	29.24	31.31	33.33	35.44	37.79	40.67	44.90	48.60	51.96	56.06
35	18.50	20.56	22.46	24.79	27.83	30.17	32.28	34.33	36.47	38.85	41.77	46.05	49.80	53.20	57.34
36	19.23	21.33	23.26	25.64	28.73	31.11	33.25	35.33	37.50	39.92	42.87	47.21	50.99	54.43	58.61
37	19.96	22.10	24.07	26.49	29.63	32.05	34.22	36.33	38.53	40.98	43.97	48.36	52.19	55.66	59.89
38	20.69	22.87	24.88	27.34	30.53	32.99	35.19	37.33	39.56	42.04	45.07	49.51	53.38	56.89	61.16
39	21.42	23.65	25.69	28.19	31.44	33.93	36.16	38.33	40.59	43.10	46.17	50.65	54.57	58.12	62.42
40	22.16	24.43	26.50	29.05	32.34	34.87	37.13	39.33	41.62	44.16	47.26	51.80	55.75	59.34	63.69
41	22.90	25.21	27.32	29.90	33.25	35.81	38.10	40.33	42.65	45.22	48.36	52.94	56.94	60.56	64.95
42	23.65	25.99	28.14	30.76	34.15	36.75	39.07	41.33	43.67	46.28	49.45	54.09	58.12	61.77	66.20
43	24.39	26.78	28.96	31.62	35.06	37.69	40.04	42.33	44.70	47.33	50.54	55.23	59.30	62.99	67.45
44	25.14	27.57	29.78	32.48	35.97	38.64	41.02	43.33	45.73	48.39	51.63	56.36	60.48	64.20	68.70
45	25.90	28.36	30.61	33.35	36.88	39.58	41.99	44.33	46.76	49.45	52.72	57.50	61.65	65.41	69.95
46	26.65	29.16	31.43	34.21	37.79	40.52	42.96	45.33	47.78	50.50	53.81	58.64	62.82	66.61	71.20
47	27.41	29.95	32.26	35.08	38.70	41.47	43.94	46.33	48.81	51.56	54.90	59.77	64.00	67.82	72.44
48	28.17	30.75	33.09	35.94	39.62	42.42	44.91	47.33	49.84	52.61	55.99	60.90	65.17	69.02	73.68
49	28.94	31.55	33.93	36.81	40.53	43.36	45.88	48.33	50.86	53.66	57.07	62.03	66.33	70.22	74.91
50	29.70	32.35	34.76	37.68	41.44	44.31	46.86	49.33	51.89	54.72	58.16	63.16	67.50	71.42	76.15
60	37.48	40.48	43.18	46.45	50.64	53.80	56.62	59.33	62.13	65.22	68.97	74.39	79.08	83.29	88.37
70	45.44	48.75	51.73	55.32	59.89	63.34	66.39	69.33	72.35	75.68	79.71	85.52	90.53	95.02	100.42
80	53.54	57.15	60.39	64.27	69.20	72.91	76.18	79.33	82.56	86.11	90.40	96.57	101.87	106.62	112.32
90	61.75	65.64	69.12	73.29	78.55	82.51	85.99	89.33	92.76	96.52	101.05	107.56	113.14	118.13	124.11
100	70.06	74.22	77.92	82.35	87.94	92.12	95.80	99.33	102.94	106.90	111.66	118.49	124.34	129.56	135.80
200	156.43	162.72	168.27	174.83	183.00	189.04	194.31	199.33	204.43	209.98	216.60	226.02	233.99	241.05	249.44

Tableau 11.11 – Lois de Fisher-Snedecor

Soit F une variable dont la loi est celle de Fisher Snedecor, à ν_1 et ν_2 degrés de liberté, de densité ρ . La table donne, pour divers couples (ν_1, ν_2) les valeurs u pour lesquelles $\mathbb{P}(F < u) = G(u) = 0.95$

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	8	12	24	>25
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.08	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
>120	3.84	2.99	2.60	2.37	2.21	2.10	1.94	1.75	1.52	1.00

Tableau 11.9 – Lois de Student

Soit T une variable de Student à ν degrés de liberté, de densité ρ . Si u est un nombre positif et si on pose :

$$F(u) = \mathbb{P}(T < u) = P$$

la table donne u pour différentes valeurs de ν et de P .

ν^P	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	0.975	0.99	0.995
1	0.1584	0.3249	0.5095	0.7265	1.0000	1.3764	1.9626	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.1421	0.2887	0.4447	0.6172	0.8165	1.0607	1.3862	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.1366	0.2767	0.4242	0.5844	0.7649	0.9785	1.2498	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.1338	0.2707	0.4142	0.5686	0.7407	0.9410	1.1896	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.1322	0.2672	0.4082	0.5594	0.7267	0.9195	1.1558	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.1311	0.2648	0.4043	0.5534	0.7176	0.9057	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.1303	0.2632	0.4015	0.5491	0.7111	0.8960	1.1192	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.1297	0.2619	0.3995	0.5459	0.7064	0.8889	1.1081	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.1293	0.2610	0.3979	0.5435	0.7027	0.8834	1.0997	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.1289	0.2602	0.3966	0.5415	0.6998	0.8791	1.0931	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.1286	0.2596	0.3956	0.5399	0.6974	0.8755	1.0877	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.1283	0.2590	0.3947	0.5386	0.6955	0.8726	1.0832	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.1281	0.2586	0.3940	0.5375	0.6938	0.8702	1.0795	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.1280	0.2582	0.3933	0.5366	0.6924	0.8681	1.0763	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.1278	0.2579	0.3928	0.5357	0.6912	0.8662	1.0735	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.1277	0.2576	0.3923	0.5350	0.6901	0.8647	1.0711	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.1276	0.2573	0.3919	0.5344	0.6892	0.8633	1.0690	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.1274	0.2571	0.3915	0.5338	0.6884	0.8620	1.0672	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.1274	0.2569	0.3912	0.5333	0.6876	0.8610	1.0655	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.1273	0.2567	0.3909	0.5329	0.6870	0.8600	1.0640	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.1272	0.2566	0.3906	0.5325	0.6864	0.8591	1.0627	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.1271	0.2564	0.3904	0.5321	0.6858	0.8583	1.0614	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.1271	0.2563	0.3902	0.5317	0.6853	0.8575	1.0603	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.1270	0.2562	0.3900	0.5314	0.6848	0.8569	1.0593	1.3178	1.7109	2.0639	2.4922	2.7970
25	0.1269	0.2561	0.3898	0.5312	0.6844	0.8562	1.0584	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.1269	0.2560	0.3896	0.5309	0.6840	0.8557	1.0575	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.1268	0.2559	0.3894	0.5306	0.6837	0.8551	1.0567	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.1268	0.2558	0.3893	0.5304	0.6834	0.8546	1.0560	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.1268	0.2557	0.3892	0.5302	0.6830	0.8542	1.0553	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.1267	0.2556	0.3890	0.5300	0.6828	0.8538	1.0547	1.3104	1.6973	2.0423	2.4573	2.7500
40	0.1265	0.2550	0.3881	0.5286	0.6807	0.8507	1.0500	1.3031	1.6839	2.0211	2.4233	2.7045
50	0.1263	0.2547	0.3875	0.5278	0.6794	0.8489	1.0473	1.2987	1.6759	2.0086	2.4033	2.6778
60	0.1262	0.2545	0.3872	0.5272	0.6786	0.8477	1.0455	1.2958	1.6706	2.0003	2.3901	2.6603
70	0.1261	0.2543	0.3869	0.5268	0.6780	0.8468	1.0442	1.2938	1.6669	1.9944	2.3808	2.6479
80	0.1261	0.2542	0.3867	0.5265	0.6776	0.8461	1.0432	1.2922	1.6641	1.9901	2.3739	2.6387
90	0.1260	0.2541	0.3866	0.5263	0.6772	0.8456	1.0424	1.2910	1.6620	1.9867	2.3685	2.6316
100	0.1260	0.2540	0.3864	0.5261	0.6770	0.8452	1.0418	1.2901	1.6602	1.9840	2.3642	2.6259
200	0.1258	0.2537	0.3859	0.5252	0.6757	0.8434	1.0391	1.2858	1.6525	1.9719	2.3451	2.6006

Résumé

La recherche en Biologie passe par une démarche expérimentale permettant le recueil des données de terrain par exemple en écologie ou dans le domaine des sciences de l'environnement. Les données de terrain sont ensuite analysées et traitées.

On peut constater aussi que dans ces dernières années, la part de la modélisation mathématique en Biologie n'a fait qu'augmenter. Aujourd'hui, la plupart des articles publiés par les grandes revues de Biologie présentent des modèles mathématiques. C'est particulièrement vrai dans le domaine de l'écologie et des sciences de l'environnement, où la modélisation mathématique apparaît comme une étape incontournable de la recherche en Biologie.

Summary

The research in biology goes through an experimental approach to the collection of field data for example in ecology or in the field of environmental sciences. Field data are then analyzed and processed.

It can be seen that in recent years, the share of mathematical modeling in biology has been increasing. Today, most articles published by major journals Biology present mathematical models. This is particularly true in the field of ecology and environmental sciences, or mathematical modeling is seen as an inevitable stage of research in biology

ملخص

البحث في البيولوجيا تمر من خلال نهج تجريبي لجمع البيانات الميدانية على سبيل المثال في البيئة أو في مجال العلوم البيئية. بعد ذلك البيانات الميدانية وتحليلها ومعالجتها. ويمكن ان يلاحظ انه في السنوات الأخيرة ، فان حصة النمذجة الرياضية في الأحياء قد تزايد. اليوم، معظم المواد التي نشرتها مجلة علم الأحياء الحالي الرئيسية النماذج الرياضية. وينطبق هذا بصفة خاصة في مجال البيئة وعلوم البيئة ، أو النمذجة الرياضية ، يعتبر نتيجة حتمية لمرحلة البحث في علم الأحياء.
