

# Statistique et Économétrie

Luciole Batola

**Luciole BATOLA**  
*Maître-Assistante à  
l'Université Paris-XIII*

# **Statistique et Économétrie**

Cours-Exemples traités-Exercices  
Licence et maîtrise ès Sciences  
Économiques

MASSON  
Paris New York Barcelone Milan  
Mexico Sao Paulo  
1983

Tous droits de traduction, d'adaptation et de reproduction par tous procédés,  
réservés pour tous pays.

La loi du 11 mars 1957 n'autorisant, aux termes des alinéas 2 et 3 de l'article 41, d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause, est illicite » (alinéa 1<sup>er</sup> de l'article 40).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles 425 et suivants du Code pénal.

© Masson, Paris, 1983

ISBN : 2-225-79299-2

---

MASSON S.A.  
MASSON PUBLISHING U.S.A. Inc.  
TORRAY-MASSON S.A.  
MASSON ITALIA EDITORI S.p.A.  
MASSON EDITORES  
EDITORIA MASSON DO BRASIL Ltda

120 Bd Saint-Germain, 75280 Paris Cedex 06  
133 East 58th Street, New York, N.Y. 10022  
Balma, 151, Barcelona 8  
Via Giovanni Pascoli 55, 20133 Milano  
Dakota 383 Colonia Napoles, Mexico 18 DF  
Rua Dr Cesario Motta Jr, 61, 01121 Sao Paulo S.P.

## Préface

Le cours de STATISTIQUE ET ECONOMETRIE de Madame Batola couvre largement le programme de la Licence ès Sciences Economiques dans ce domaine, puisque, outre des rappels nécessaires de l'Enseignement de Statistique de deuxième année, elle traite une partie du programme d'Econométrie de Maîtrise. Il a été élaboré au cours de plusieurs années d'enseignement effectif à l'Université Paris XIII et comporte de nombreux exemples entièrement traités.

Au-delà de son indiscutable valeur pédagogique, attestée par la clarté et la rigueur de l'exposition ainsi que par le souci des illustrations économiques, cet ouvrage se recommande aussi par ses implications méthodologiques. Un modèle mathématique, ici statique et comportant une partie aléatoire, considéré comme représentation d'un modèle économique préalable, fournit du "réel", d'accès difficile ou impossible, une image nécessairement imparfaite. Mais l'imperfection même de l'image est à la fois un inconvénient évident et un avantage si l'on sait ne retenir que les éléments les plus pertinents de la réalité. Il y a là une sorte de dualité entre la finesse de la représentation, intéressante si l'on veut décrire, et sa simplicité, utile si l'on recherche l'intelligibilité.

Enfin, le soin pris par l'auteur de situer dans le temps les mathématiciens, probabilistes et statisticiens rencontrés au cours des chapitres, témoigne d'un intérêt pour l'aspect historique de la découverte scientifique que l'on souhaiterait rencontrer plus souvent.

R. Vallée'

Robert Vallée  
Septembre 1982

## Table des matières

INTRODUCTION. Ce qu'est l'économétrie .....	1
Le rôle des mathématiques .....	3
Le plan du travail .....	4
PREMIERE.PARTIE : LES TESTS	
CHAPITRE 0. Indications générales sur les tests .....	7
I. Hypothèses .....	7
II. Tests .....	8
III. Risques .....	10
CHAPITRE I. Test paramétrique de J. Neyman et E. Pearson .....	13
I. Test du rapport de vraisemblances dans le cas d'hypothèses simples .....	13
II. Test d'une hypothèse nulle simple contre une hypothèse alternative composite .....	28
III. Exemples de tests de Neyman et Pearson utilisant les lois normales, de Student, de Fisher et du $X^2$ .....	35
Exercices .....	57
CHAPITRE II. Test paramétrique de Karl Pearson utilisant la v.a. $X^2$ .....	70
I. Méthode de Karl Pearson .....	70
II. Propriétés du test de K. Pearson .....	75
III. Test des valeurs de plusieurs probabilités ou pourcentages associés à un même système complet d'événements .....	78
IV. Test de l'indépendance de deux systèmes complets d'événements associés à une même épreuve ou test de l'indépendance de deux caractères dans une population .....	82
Exercices .....	91

<i>CHAPITRE III.</i> Tests de confiance .....	96
I. Test du $\chi^2$ de Karl Pearson .....	96
II. Test de Kolmogorov .....	106
Exercices .....	113

DEUXIEME PARTIE : MODELES LINEAIRES ALEATOIRES

<i>CHAPITRE IV.</i> Modèle classique linéaire à une équation .....	117
I. Exemples de linéarisation de modèles.	
Notions de régressant et de régresseurs .....	117
II. Définition du modèle classique linéaire .....	119
III. Théorème de Gauss-Markov .....	122
IV. Propriétés de l'estimateur $\hat{A}$ .....	126
V. Propriétés de l'estimateur de Var $\hat{A}$ .....	130
VI. Mesures des liaisons entre les différentes variables d'un modèle .....	132
VII. Prévission .....	137
VIII. Test des coefficients du modèle .....	139
Exercices .....	150
<i>CHAPITRE V.</i> Exemples de quelques modèles linéaires non classiques .....	158
I. Modèle linéaire généralisé à une équation. Théorème d'Aitken.	158
II. Modèle à régresseurs aléatoires relevant de la méthode de la variable instrumentale .....	166
III. Modèle linéaire aléatoire indépendant .....	172
IV. Modèle linéaire aléatoire non asymptotiquement corrélé avec les perturbations .....	174
V. Modèle linéaire autorégressif d'ordre $k$ .....	177
VI. Modèle linéaire à contraintes égalités .....	179
VII. Modèle à retards échelonnés avec la spécification de Koyck .....	182
<i>CHAPITRE VI.</i> Modèles linéaires à plusieurs équations .....	186
I. Modèle classique linéaire pluridimensionnel .....	186
II. Modèle structurel classique linéaire à $m$ équations .....	197
III. Modèle réduit associé au modèle structurel classique linéaire .....	201
IV. Identification du modèle structurel .....	205
V. Estimation d'une équation juste identifiable ; méthode des moindres carrés indirects .....	213
VI. Estimation d'une équation sur identifiable ; méthode des doubles moindres carrés .....	217

VII. Estimation d'un modèle récursif .....	226
Exercices .....	230
CHAPITRE VII. Notions d'analyse de variance .....	238
I. Formulation mathématique du problème dans le cas d'une classification simple .....	239
II. Exemples simples d'analyse de variance .....	243
III. Exemple d'analyse de variance dans le cas d'une classifi- cation double .....	244
Notations .....	248
ANNEXE A : Rappels de résultats d'algèbre matricielle .....	249
ANNEXE B : Rappels de résultats de théorie des probabilités .....	254
Principales distributions de probabilité .....	256
ANNEXE C : Rappels de statistique .....	261
Table de la fonction de répartition de la loi normale réduite .....	264
Table des grandes valeurs de la fonction de répartition de la loi normale réduite .....	265
Table du $d$ de Durbin-Watson .....	266
Table de $F$ .....	268
Table des valeurs de la distribution de Poisson .....	270
Table de $T$ .....	271
Table du $\chi^2$ .....	272
Index alphabétique des matières .....	273

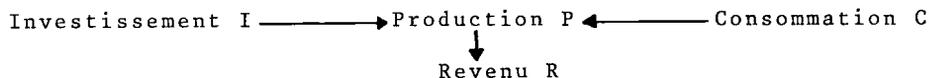
# Introduction

L'*économétrie au sens strict* traite des *modèles aléatoires* tandis que l'*économétrie au sens large* traite en plus de la théorie mathématique de la décision. On ne s'intéresse ici qu'à l'*économétrie au sens strict*.

## 1. CE QU'EST L'ECONOMETRIE

Toute *théorie économique* est constituée d'un ensemble de propositions déduites les unes des autres à partir d'hypothèses dont l'énoncé fait intervenir des grandeurs économiques parfaitement définies : marchandises, prix, consommation, revenu, épargne, investissement, etc ...

L'*économiste* établit des *modèles économiques* appelés *modèles réels*, qui donnent, sous forme de relations causales entre les principales grandeurs économiques, des images simplifiées d'aspect de la réalité, conformément à la théorie économique en vigueur au lieu et à l'époque de l'étude.



Exemple du modèle keynésien (1) simplifié.

---

(1) John Maynard Keynes (1883-1946) : économiste britannique

L'économètre fait correspondre à chaque modèle *réel*, au moins un *modèle mathématique* appelé modèle structurel, où les grandeurs et les processus du modèle réel sont remplacés par des symboles et des opérations. Un modèle mathématique, lorsqu'il est aléatoire se présente comme l'ensemble de relations

$$\begin{aligned} f_1(Y_1, \dots, Y_m, X_1, \dots, X_k, e_1) &= 0 \\ \cdot & \\ \cdot & \\ \cdot & \\ f_m(Y_1, \dots, Y_m, X_1, \dots, X_k, e_m) &= 0 \end{aligned}$$

où  $Y_1, \dots, Y_m$  représentent les *variables endogènes* (encore appelées *variables expliquées*) qui agissent les uns sur les autres et/ou dépendent des *variables exogènes*  $X_1, \dots, X_k$  (encore appelées *variables explicatives*) dont la valeur doit être connue dès l'écriture du modèle (parmi les variables exogènes figurent les données relatives à l'environnement international et la politique économique). A ces variables endogènes et exogènes vient s'adjoindre, pour la  $j^{\text{ème}}$  relation du modèle, une variable aléatoire  $e_j$  (et cela pour tout  $j$ ) appelée *perturbation* (ou erreur) ; elle intervient dans la rédaction des équations du modèle pour synthétiser l'effet de toutes les variables explicatives omises et/ou de toutes les erreurs de mesure faites éventuellement sur les variables explicatives ou expliquées.

Le modèle est *complet* s'il comprend autant de relations que de variables expliquées ( $m = m'$ ). Les relations  $f_1 \dots f_m$  du modèle complet comportent des coefficients  $a_1, \dots, a_m, b_1, \dots, b_k$  que l'économètre aura à *estimer* à partir de séries statistiques se rapportant aux variables endogènes et exogènes, dont il aura à tester la signification et la qualité de l'estimation, et dont il se servira pour prévoir les valeurs des variables expliquées puis tester l'adéquation entre les prévisions et les observations faites sur le terrain.

Le modèle est *linéaire* si ses coefficients  $a_i$  et  $b_j$  ( $i = 1, \dots, m$  et  $j = 1, \dots, k$ ) n'interviennent que par leur puissance un, et *linéarisé* lorsqu'il est rendu linéaire après transformation mathématique (comme passage au logarithme, élévation à une puissance, etc ...). L'étude la plus célèbre de modèle linéarisé est due à Lawrence Klein (prix Nobel d'Economie 1980) et Arthur Goldberger (Professeur à l'Université du Wisconsin), en 1959, et concerne l'économie globale des U.S.A.

Lorsqu'aucune des relations  $f_1, \dots, f_m$  ne comporte de variable aléatoire, le modèle est dit *déterministe*.

$$\begin{array}{ll} C = aR + c & C : \text{Consommation} \quad R : \text{Revenu} \\ R = bP & P : \text{Production} \\ P = C + I & I : \text{Investissement} \end{array}$$

Exemple de modèle *déterministe* correspondant au modèle keynésien simplifié où I est l'unique variable exogène.

$$\begin{array}{l} C = aR + c + e_1 \\ R = bP + e_2 \\ P = C + I \end{array}$$

Exemple de modèle *aléatoire* correspondant au modèle keynésien simplifié.

L'étude des modèles aléatoires ne se conçoit pas sans la statistique qui est une branche du Calcul des Probabilités lui même partie intégrante de la Mathématique.

## 2. LE ROLE DES MATHÉMATIQUES

On s'adresse aux Mathématiques pour étudier les modèles économiques parce que les modèles mathématiques se sont révélés quasiment infaillibles notamment en Astronomie (c'est à la suite des calculs mathématiques de Le Verrier (1) en 1846 que la planète Neptune a été découverte la même année, et à la suite des calculs de Lowell (2) en 1915 que la planète Pluton a été découverte en 1930) et en Astrophysique (les astrophysiciens Schwatzman et Maeder de l'Université PARIS XI ont estimé en Juin 1980 un modèle du bombardement de la terre par des flux de neutrinos en provenance du centre solaire, rendant *très bien* compte du flux d'hélium 3 rencontré dans le vent solaire provenant de la surface solaire).

Or, comme les phénomènes astrophysiques, les phénomènes économiques ne peuvent que s'observer et non être reproduits en laboratoire au cours d'expériences contrôlées. On fonde

---

(1) Urbain Le Verrier (1811-1877) : astronome français.

(2) Percival Lowell (1855-1916) : astronome américain.

donc beaucoup d'espoir sur une utilisation judicieuse de modèles *valables* qui permettront de perfectionner la connaissance des phénomènes économiques, de départager les théories économiques en mettant en évidence celle (s) qui explique (nt) le moins mal la réalité ou qui donne (nt) les moins mauvaises prévisions. Pour qu'un modèle soit valable, il faut mais il ne suffit pas que les grandeurs économiques auxquelles il se rapporte soient observées pendant longtemps. Lorsqu'en 1950 on construisait des modèles, on ne disposait ni de séries statistiques suffisamment longues ni de moyens informatiques suffisants. En 1968 on a fabriqué le modèle DECA dont les relations se sont révélées insuffisamment estimées. En Juillet 1973 on a créé le modèle macroéconomique STAR (Schéma Théorique d'Accumulation et de Répartition) qui étudie l'économie française durant la période 1953-1972. On fabrique également des modèles microéconomiques qui servent à décrire et prévoir le comportement d'un agent économique (une entreprise par exemple). Dans tous les cas l'économiste *indique* les relations causales que l'économètre *mesure* conférant ainsi à l'économie son statut de science.

### 3. LE PLAN DU TRAVAIL

Le présent cours de Statistique et Econométrie, accompagné d'exemples traités et d'exercices, s'attache à donner en première partie les outils nécessaires à la construction des tests statistiques, et à exposer en seconde partie la théorie des modèles aléatoires linéaires classiques munis de leurs tests de validité. Ceci fait l'objet des chapitres 0 à IV, recouvrant ainsi le programme généralement vu en année de Licence ès Sciences Economiques. Dans le chapitre V sont traités quelques exemples de modèles non classiques à une équation généralement vus en quatrième année ; on aura intérêt à consulter les rappels de Statistique avant la lecture de ce chapitre. On aborde ensuite dans le chapitre VI l'étude des modèles linéaires à plusieurs équations généralement faite en quatrième année et au-delà. Le chapitre VII donne un aperçu de ce que l'on appelle communément "analyse de variance" et qui permet de faire la liaison théorique entre la notion de modèle linéaire classique et l'analyse statistique de données.

Des exemples et des illustrations numériques traités de manière détaillée figurent dans le courant de chaque chapitre. Pour rendre plus aisée la compréhension de la partie "Econométrie" de ce manuel (chapitres IV à VII) l'annexe A donne des rappels de résultats d'Algèbre matricielle.

L'annexe B donne des rappels de résultats de théorie des Probabilités, suivis des propriétés des principales distributions de probabilité utilisées dans la partie "Statistique" du manuel (chapitres I à III). L'annexe C donne des rappels de résultats de Statistique suivis des principales tables statistiques utilisées lors de la construction des tests.

Pour rendre plus facile la lecture de ce manuel, on se reportera utilement à l'exposé des notations (précédant immédiatement les annexes), et à l'index alphabétique (suivant les annexes). On distinguera plus facilement les titres des paragraphes, écrits en haut des pages impaires, des titres des chapitres écrits en haut des pages paires.

Je remercie par avance tous ceux qui voudront bien me signaler les erreurs en tous genres qui figurent dans le présent manuel.

Première Partie

# **Statistique**

## **Les Tests**

# Indications générales sur les tests

## I. LES HYPOTHESES

### 1. Echantillon

On considère une épreuve  $E$  à laquelle est associé un ensemble  $\Omega = \{\omega_1, \dots, \omega_i, \dots\}$  de résultats possibles formant un *système complet d'événements*  $\omega_i \cap \omega_j = \emptyset$  et  $\sum \omega_i = \Omega$ . On associe à  $E$  une variable aléatoire  $X$  ( $X$  est une application de  $\Omega$  dans  $\mathbb{R}$ ) de loi de probabilité, discrète ou continue, notée  $f(x)$ . On effectue  $n$  *épreuves indépendantes*  $E_1, \dots, E_2, \dots, E_n$  *identiques* à  $E$ . On appelle  $X_i$  la v.a. correspondant à  $E_i$  ( $i = 1, \dots, n$ ).  $M = (X_1, \dots, X_i, \dots, X_n)$  est l'échantillon aléatoire de taille  $n$  de  $X$ . Si  $x_1, \dots, x_n$  sont les valeurs obtenues pour  $X$  lors des  $n$  épreuves  $E_i$ ,  $(x_1, \dots, x_n)$  est l'échantillon expérimental ou observé de taille  $n$  de  $X$ . La vraisemblance de  $M$  ou loi de probabilité de  $M$  a pour expression (aussi notée  $L(M)$ ):

$$V(M) = \prod_{i=1}^n f(x_i)$$

que la loi de probabilité  $f$  soit discrète ou continue.

## 2. Hypothèses statistiques

Une hypothèse statistique porte sur la loi de probabilité d'une v.a.X ou de plusieurs variables aléatoires.

Elle est dite *paramétrique* si, concernant une v.a.X dont la loi de probabilité dépend de  $k$  paramètres ( $k \geq 1$ ) :  $f(x, \theta_1, \dots, \theta_k)$ , elle peut être exprimée comme relation entre ces paramètres, par exemple comme égalité d'un paramètre avec un nombre réel donné, ou comme égalité de plusieurs paramètres entre eux.

Une hypothèse paramétrique est *simple* si elle concerne une valeur *unique* de *chacun* des paramètres de la loi  $f$  de  $X$  ; elle est dite *composite* dans le cas contraire.

Exemple d'hypothèse paramétrique simple :  $X$  de loi  $N(0, 1)$ .

Exemple d'hypothèse paramétrique composite :  $X$  de loi  $N(0, v)$  où  $v$  est positive ;  $X$  de loi  $N(0, v)$  où  $v > 1$  ;  $X$  de loi  $N(m, 1)$  où  $m > 0$ .  $X$  de loi  $N(m, 1)$  où  $m \neq 0$  ;  $X$  de loi  $N(m, v)$  etc ...

En général, on appelle *hypothèse nulle* et l'on note  $H_0$ , l'hypothèse qui affirme l'*absence* d'un phénomène dont on veut prouver l'existence (qu'on la souhaite ou qu'on la redoute) ;  $H_0$  peut être simple ou composite. On appelle *hypothèse alternative* et l'on note  $H_1$  l'hypothèse telle que  $H_0$  et  $H_1$  s'excluent l'une l'autre.

## II. TESTS

### 1. Test statistique

C'est une procédure de décision d'accepter l'hypothèse statistique  $H_0$  (par conséquent de rejeter  $H_1$ ) ou de rejeter  $H_0$  (par conséquent d'accepter  $H_1$ ) à partir d'un échantillon expérimental de taille  $n$  de  $X$ . Le test permet seulement de se prononcer pour ou contre  $H_0$  connaissant  $H_1$ , ce *n'est pas une démonstration*. La décision de rejeter ou de garder  $H_0$  dépend de la définition de  $H_1$ .

Exemple d'une v.a.X de loi  $N(m, v)$  :

Supposons que l'on veuille tester  $H_0 : m \in ]-1, 0 [$  contre  $H_1 : m \in ]0, 1 [$  et que l'on ait obtenu, à partir d'un échantillon de

taille  $n$  de  $X$ , une estimation de  $m$  égale à  $0,3$  :  $\tilde{m} = 0,3$  ; il paraît naturel de rejeter  $H_0$ . Par contre si l'on teste  $H_0$  contre la nouvelle hypothèse alternative  $H_1 : m \in ]1, 2[$ , il paraît alors naturel de garder  $H_0$ .

## 2. Test paramétrique

Un test statistique est dit paramétrique lorsque les hypothèses  $H_0$  et  $H_1$  entre lesquelles le test doit permettre de trancher sont paramétriques ; il est dit non paramétrique dans le cas contraire.

## 3. Test de confiance

Un test statistique est dit test de confiance lorsque  $H_0$  ne concerne pas la (ou les) valeur (s) d'un ou de plusieurs paramètres de la loi de  $X$  mais la fonction de répartition  $F(x)$  de la v.a.  $X$ , autrement dit la *nature* de la loi de  $X$ .

Si l'hypothèse  $H_0$  *spécifie entièrement*  $F(x)$ , elle est dite simple. Si elle ne spécifie pas  $F(x)$  entièrement,  $H_0$  est dite composite.

Exemple d'hypothèse  $H_0$  simple :

$X \in N(0,4)$  ? la v.a.  $X$  suit-elle une loi *normale* de moyenne  $0$ , de variance  $4$  ?

Exemples d'hypothèse  $H_0$  composite :

$X$  de loi  $N(m,s)$  ;  $X$  de loi  $N(0,s)$  ;  $X$  de loi  $N(m,4)$   
etc ...

Les hypothèses nulles données ici en exemple diffèrent de celles données au § I.2. en ce que, ici, on *s'interroge* sur la normalité de la loi de  $X$  en précisant ou non tous ses paramètres (selon que  $H_0$  est simple ou non), alors qu'au § I.2. on suppose *a priori* la loi de  $X$  normale, et l'on s'interroge sur la valeur de ses paramètres.

## 4. Ensemble de rejet $W$ de $H_0$

Pour pouvoir prendre une décision concernant une hypothèse  $H_0$ , il faut disposer d'un échantillon observé  $(x_1, \dots, x_n)$  de la v.a.  $X$  auquel on fait correspondre un point  $M$  de l'espace  $E$  des

échantillons ;  $E \subset \mathbb{R}^n$ . Après étude de la loi de  $M$  (vraisemblance  $V(M)$  de l'échantillon aléatoire Cf § I.1), on fixe l'ensemble  $W$  des échantillons  $(x_1, \dots, x_n)$  pour lesquels on décide de rejeter  $H_0$  ;  $W$  est appelé *ensemble de rejet* de  $H_0$  ;  $W \subset E$ . Le complémentaire de  $W$  dans  $E$ ,  $\bar{W}$ , est appelé *ensemble d'acceptation* de  $H_0$ . La décision de rejeter ou d'accepter  $H_0$  est prise si le résultat expérimental obtenu appartient à  $W$  ou à  $\bar{W}$ .

### 5. Test unilatéral. Test bilatéral

Le test est dit *unilatéral* si l'ensemble de rejet  $W$  de  $H_0$  est d'un seul tenant ; l'hypothèse  $H_1$  est dite unilatérale.

Le test est *bilatéral* si  $W$  est en deux parties ;  $H_1$  est alors dite bilatérale, etc ...

### 6. Statistique ou fonction discriminante du test

C'est une fonction de  $(x_1, \dots, x_n)$  permettant de fixer  $W$  après étude de  $V(M)$ .

## III. RISQUES

En prenant la décision d'accepter ou de rejeter  $H_0$ , il y a deux façons de se tromper, ou encore deux espèces d'erreurs :

- rejeter  $H_0$  à tort            ou erreur de première espèce
- accepter  $H_0$  à tort        ou erreur de seconde espèce.

### 1. Risque de première espèce

C'est la probabilité, notée  $\alpha$ , de l'erreur de première espèce. Le risque de première espèce est encore appelé *seuil de confiance* du test.

$$\alpha = P(W \mid H_0)$$