

**Reproduction sur d'autres sites interdite mais lien vers le document accepté :**

[www.irdes.fr/documentation/syntheses/statistiques-et-econometrie.pdf](http://www.irdes.fr/documentation/syntheses/statistiques-et-econometrie.pdf)

[www.irdes.fr/documentation/syntheses/statistiques-et-econometrie.epub](http://www.irdes.fr/documentation/syntheses/statistiques-et-econometrie.epub)

**Toutes nos synthèses sont disponibles à cette adresse :**

[www.irdes.fr/documentation/syntheses-et-dossiers-bibliographiques.html](http://www.irdes.fr/documentation/syntheses-et-dossiers-bibliographiques.html)

## Statistique et Econométrie : Ouvrages et sites de références

Septembre 2016

### Sommaire

Méthodes statistiques .....	2
Probabilités – Mathématique .....	7
Econométrie.....	9
Modèles multiniveaux.....	15
Autres méthodes.....	18
Analyses de données.....	21
Data mining.....	23
La statistique avec SAS.....	24
La statistique avec Stata .....	26
La statistique avec R.....	29

## Méthodes statistiques

Calot, G. (1973). Cours de statistique descriptive, Paris : Dunod

Grais, B. (1979). Statistique descriptive. Techniques statistiques 1, Paris : Dunod

Grais, B. (1977). Méthodes statistiques. Techniques statistiques 2, Paris : Dunod

Gremy, J. P. (2003). Introduction à la lecture des tableaux statistiques, Auxerre : Sciences Humaines éditions  
<http://www.scienceshumaines.com/textesInedits/Gremy.pdf>

Ce manuel propose quelques principes empiriques destinés à faciliter la lecture des tableaux statistiques que l'on rencontre dans les sciences sociales. À cette fin, il présente plusieurs types différents de tableaux tirés de la littérature sociologique, et montre comment il est possible d'en extraire les informations essentielles sans utiliser d'autres outils mathématiques que les quatre opérations arithmétiques. Les exemples sont présentés par ordre de complexité croissante. Ils sont l'occasion de rappeler les notions de base indispensables pour lire et analyser les données quantitatives (niveaux de mesure, liaison entre variables, interaction, etc.), mais en les considérant seulement du point de vue des services qu'ils peuvent rendre au sociologue ; tout formalisme mathématique en a donc été exclu. C'est pour introduire ces notions d'une manière progressive que ce manuel accorde une place qui pourrait sembler trop importante aux tableaux les plus simples que l'on puisse trouver : les distributions à une seule dimension (tiré de l'intro).

Harper, W. M. (1988). Statistics, Londres : Pitman Publishing

Levy, M. L. (1979). Comprendre les statistiques, Paris : Seuil

Nakache, J. P. and J. Confais (2003). Statistique explicative appliquée, Paris : Editions Technip

Ce guide pratique présente trois méthodes de classement couramment utilisées et implantées dans plusieurs logiciels statistiques : l'analyse discriminante linéaire et quadratique ; le modèle logistique binaire et multinomial ; la segmentation par arbre de régression et de discrimination. L'ouvrage insiste particulièrement sur les illustrations de ces trois méthodes, réalisées avec les logiciels SAS® et SPAD®, et sur l'interprétation des résultats.

Novi, M. (1998). Pourcentages et tableaux statistiques, Paris : Presses Universitaires de France

Cet ouvrage traite des pourcentages en analyse statistique. Il se consacre aux méthodes descriptives et les présente par complexité croissante, le seul outil requis demeurant la règle de trois. L'étude et la mesure de l'association font l'essentiel de ce livre qui s'adresse à tous publics, mais en particulier aux étudiants des disciplines pour lesquelles l'usage des échelles nominales est intensif.

Py, B. (2007). La statistique sans formule mathématique. Comprendre la logique et maîtriser les outils, Paris : Pearson Education France

Cet ouvrage est un outil de sensibilisation à la statistique. Il cherche à expliquer les bases de la culture statistique, ses outils et ses méthodes : les enquêtes, les tableaux, les moyennes et variances, les indices, les séries chronologiques et la corrélation.

Saporta, G. (1990). Probabilités - Analyse des données et statistiques, Paris : éditions Technip

Ce livre présente les méthodes les plus courantes de la statistique moderne sous forme d'un manuel de cours. Son originalité est de fournir en un seul volume l'essentiel des connaissances utiles en calcul des probabilités, analyse des données multidimensionnelles (analyses factorielles, classification) et statistique décisionnelle (tests, estimation, régression, discrimination).

Tassi, P. (1985). Méthodes statistiques, Paris : Economica

Hart, A. (2002). Making sense of statistics in healthcare, Abingdon : Radcliffe Medical Press

This book is unique in being written for people who want to be able to make sense of published studies, or embark on their own studies, without getting bogged down by the details of how to use specific methods. It gives insight into how statisticians view data and research, and into what they do. This book is virtually formula free, and it does not describe the mechanics of techniques but is packed with ideas and examples. It describes how statisticians use data and provides a thorough explanation of statistical concepts in the context of real studies to enable the reader to make informed judgements. Invaluable for those involved in research, it shows how to combine formal results with common and clinical sense.

Elliott, P., et al. (2000). Spatial epidemiology : methods and applications, Oxford : Oxford University Press

Cet ouvrage est une nouvelle édition de l'ouvrage de référence : Spatial epidemiology : methods and applications. Il analyse de façon claire et concise la variation spatiale des maladies, et étudie l'impact des risques environnementaux sur la santé au niveau des petits territoires.

Vignerot, E. (1997). Géographie et statistique, Paris : Presses Universitaires de France

L'objectif de cet ouvrage est de se limiter à l'exposé synthétique d'une méthode et de techniques susceptible d'ancrer la géographie dans le monde des sciences. Il n'aborde pas l'analyse statistique des données géographiques sous l'angle de l'analyse spatiale ou uniquement de l'analyse de données. Pour cela l'ouvrage est organisé autour des principales questions qui se posent au géographe dans l'analyse des données qu'il a recueillies et qui peuvent se présenter sous forme de tableaux de valeurs.

Zaninetti, J. M. (2005). Statistique spatiale : méthodes et applications géométriques, Paris : Lavoisier ; Paris : Hermès

Cet ouvrage propose une synthèse des méthodes spécifiques de la statistique spatiale et permet d'analyser l'information géographique dans différents domaines d'études. Il est conçu dans une perspective interdisciplinaire, allant des sciences de l'environnement aux sciences sociales en associant la recherche fondamentale et la recherche appliquée.

Ancelle, T. and G. Brucker (2011). Statistique épidémiologique, Paris : Maloine

Cet ouvrage se propose de rendre attractives et compréhensibles les disciplines de la statistique et de l'épidémiologie pour les étudiants en sciences de la santé, mais aussi pour tous les professionnels de santé (médecins, pharmaciens, biologistes, infirmières, professions paramédicales, techniciens biologistes, vétérinaires). Il met l'accent sur la compréhension des principes fondamentaux grâce auxquels il devient possible, à partir de nombreux exemples et exercices, d'utiliser les tests statistiques les plus appropriés pour une recherche ou pour la conduite d'une enquête épidémiologique. La première partie étudie les outils servant à décrire les données. La deuxième aborde les méthodes d'estimation d'un paramètre inconnu à partir d'un échantillon. La troisième concerne l'emploi des tests statistiques. Elle comporte de nombreux tableaux pratiques d'aide au choix d'un test en fonction de la nature des problèmes, des paramètres à comparer et des conditions d'application ; cette partie est complétée par une série de « fiches pratiques » des principaux tests usuels. La quatrième partie est orientée vers les concepts statistiques utilisés en épidémiologie de terrain (4e de couverture).

Bouyer, J. (1999). Méthodes statistiques : médecine - biologie : exercices corrigés, Paris : Editions ESTEM

L'apprentissage des méthodes statistiques commence bien sûr par la compréhension des principes théoriques qui les gouvernent ; c'est l'objet des cours et des livres de méthodes. Mais il passe aussi par une phase d'exercices pour mettre en oeuvre ces méthodes. Ce livre est consacré à cette partie du travail qui, pour paraître fastidieuse, n'en est pas moins indispensable.

Beaucage, C. and Bonnier-Vigier. Y. (1996). Epidémiologie appliquée. Une initiation à la lecture critique de la littérature en sciences de la santé, Paris : Gaëtan Morin éditeur Europe

Cet ouvrage a pour objectif d'initier, l'utilisateur à la lecture critique de la littérature en sciences de la santé, lecture qui nécessite des connaissances en biostatistique et en épidémiologie. Le manuel comporte quinze chapitres : cinq de ces chapitres traitent de l'analyse critique de la littérature scientifique, neuf d'épidémiologie et de biostatistique et un chapitre de prévention clinique.

Morton, R. F., et al. (1996). Epidémiologie et biostatistique, Paris : Doin ; Initiatives santé

Cet ouvrage est la quatrième édition d'un ouvrage devenu un best-seller , aussi bien aux Etats-Unis que dans les pays francophones. Il propose une vision claire et concise de deux disciplines de la santé que sont l'épidémiologie et la biostatistique. Son originalité est avant tout pédagogique. Une succession de courts chapitres, suivis d'exercices et de questions tests, permettent au lecteur de vérifier les progrès accomplis tout au long de sa lecture. Cette nouvelle édition est enrichie d'un chapitre d'aide à la lecture des articles scientifiques utilisant les méthodes de l'épidémiologie et de la biostatistique. Une attention particulière est en outre portée à certaines notions de base, parfois difficiles, tels que les interactions entre variables, les modèles épidémiologiques de régression, les analyses de survie, etc.

Schwartz, D. (1969). Méthodes statistiques à l'usage des médecins et des biologistes, Paris : Flammarion

Schwartz, D. (1995). Méthodes statistiques à l'usage des médecins et des biologistes, Paris : Médecine Sciences Flammarion

Cette édition diffère des précédentes par une nouvelle rédaction de la troisième partie. En premier lieu, il est apparu que la distinction entre régression et corrélation n'était pas assez claire. Par ailleurs, pour les formules exprimant la variance liée, les fractions de variance expliquée et non expliquée et la mesure de l'intensité de la liaison, la condition de validité indiquée était une distribution binormale, alors qu'en fait, il suffit qu'une des régressions soit linéaire. Enfin, plusieurs modifications mineures ont été apportées.

Schwartz, D., et al. (1994). Statistique médicale et biologique, Paris : Médecine Sciences Flammarion

L'emploi de la méthode statistique s'impose chaque jour davantage en médecine et en biologie. C'est que cette forme de raisonnement s'avère nécessaire à l'analyse des problèmes où la variabilité des données est importante, comme c'est le cas dans les sciences de la vie. Cet ouvrage est plus particulièrement destiné aux futurs médecins. Les raisonnements fondamentaux de la méthode statistique y sont donc exposés sur des exemples inspirés de certains problèmes médicaux.

Falissard, B. (1998). Comprendre et utiliser les statistiques dans les sciences de la vie, Paris : Masson

Que ce soit en biologie, en médecine ou dans les sciences de la vie en général, il est indispensable de recourir à des méthodes objectives pour confronter et valider des résultats expérimentaux. Cet ouvrage traite dans une première partie des méthodes univariées (représentations graphiques, estimation, tests

d'hypothèse). La deuxième partie s'intéresse aux techniques multivariées, développant notamment les concepts d'ajustement (régression linéaire, analyse de variance, régression logistique, analyse discriminante...), ou encore ceux de modèle ou de représentation graphique multidimensionnelle. Exploitant largement l'outil informatique, cet ouvrage abonde d'exemples tirés de situations réelles.

Valleron, A. J. (1998). Introduction à la biostatistique, Paris : Masson

L'objectif de cet ouvrage de synthèse est de fournir les bases biostatistiques nécessaires à la compréhension de la littérature biomédicale moderne, de présenter des applications importantes de la biostatistique dans le domaine de la biomédecine, en particulier de l'épidémiologie et de la clinique. Après une familiarisation avec la notion de variabilité, sont donnés les bases mathématiques des calculs de probabilité et de nombreux exemples d'application, fournissant une initiation aux méthodes de mesure du risque en épidémiologie et d'évaluation diagnostique. Sont ensuite présentés les éléments de théorie de l'estimation et des tests statistiques utilisés dans la recherche. Puis sont décrits des tests classiques de comparaison et d'indépendance, l'analyse des données de survie dans les contextes de la démographie et de l'épidémiologie clinique, la présentation de la méthodologie des essais thérapeutiques contrôlés. L'aspect pratique est présent tout au long de l'ouvrage.

Porter, T. M. (1996). Trust in numbers : the pursuit of objectivity in science and public life, Princeton : Princeton University Press

This investigation of the overwhelming appeal of quantification in the modern world discusses the development of cultural meanings of objectivity over two centuries. How are we to account for the current prestige and power of quantitative methods? The usual answer is that quantification is seen as desirable in social and economic investigation as a result of its successes in the study of nature. Theodore Porter is not content with this. Why should the kind of success achieved in the study of stars, molecules, or cells be an attractive model for research on human societies? he asks. And, indeed, how should we understand the pervasiveness of quantification in the sciences of nature? In his view, we should look in the reverse direction: comprehending the attractions of quantification in business, government, and social research will teach us something new about its role in psychology, physics, and medicine. Drawing on a wide range of examples from the laboratory and from the worlds of accounting, insurance, cost-benefit analysis, and civil engineering, Porter shows that it is "exactly wrong" to interpret the drive for quantitative rigor as inherent somehow in the activity of science except where political and social pressures force compromise. Instead, quantification grows from attempts to develop a strategy of impersonality in response to pressures from outside. Objectivity derives its impetus from cultural contexts, quantification becoming most important where elites are weak, where private negotiation is suspect, and where trust is in short supply.

Lebart, L., et al. (1994). Statistique textuelle, Paris : Dunod

La statistique textuelle, en plein développement, est à la croisée de plusieurs disciplines : la statistique classique, la linguistique, l'analyse du discours, l'informatique, le traitement des enquêtes. En effet, chercheurs et praticiens ont aujourd'hui à faire face à un double développement, d'une part celui des textes provenant des enquêtes, des entretiens, des archives, des bases documentaires, d'autre part, celui des outils informatiques de saisie et de gestion de textes. La statistique textuelle se veut précisément un outil destiné à parfaire l'analyse, la description, la comparaison, en un mot, le traitement des textes. Ce livre, illustré d'exemples nombreux, présente les concepts de base et les fondements des méthodes de la statistique textuelle. Il combine une approche pédagogique des outils et un exposé sur l'état de l'art de cette discipline.

Schwartz, D. (1994). Le jeu de la science et du hasard : la statistique et le vivant, Paris : Médecine Sciences Flammarion

La statistique n'est pas accumulation de chiffres. C'est un mode de passé original, indispensable pour bien comprendre et agir dans le domaine du vivant, de la vie, où le Hasard est roi. Un roi à qui tout est permis, mais qui n'a ni intelligence ni mémoire. Grâce à quoi le statisticien est en mesure de lui faire la loi et de prévoir l'imprévisible, sous réserve d'un risque d'erreur contrôlé. Parfois il faut quémander son aide, au roi Hasard, comme dans le tirage au sort des échantillons. Un duel serré, passionnant à suivre, ce jeu de la Science et du Hasard. Le mode de pensée statistique est exposé dans ce livre de manière simple ; sans faire appel aux mathématiques. En dehors de cinq fiches, aucune formule n'y figure. Le lecteur y trouvera le sens exact d'expression qui lui sont familières, comme la fourchette, l'échantillon représentatif, la différence significative. Il comprendra aussi le pourquoi d'innombrables erreurs dues à la méconnaissance du mode de pensée statistique et peut-être découvrira-t-il que la statistique n'est pas nécessairement ennuyeuse.

Déchiffrer les statistiques économiques et sociales, Paris : Dunod

Les statistiques sont omniprésentes (dans les études, les journaux...). Il est indispensable à tout étudiant qui souhaite comprendre le monde dans lequel il vit d'apprivoiser tous ces chiffres. L'objectif de ce livre est de mettre à la portée de tous les connaissances de base nécessaires pour comprendre les statistiques, se faire une opinion sur leur sens, leur pertinence, leur fiabilité... et leurs limites. Ce livre propose une analyse des chiffres qui font débat, un mode d'emploi des statistiques accessible à tous, une information claire et précise sur la manière dont sont construites les statistiques.

Tribout, B. (2007). Statistique pour économistes et gestionnaires, Paris : Pearson Education France

Toutes les étapes de la réalisation d'une étude statistique sont détaillées : présentation, résumé, mesure de l'évolution et croisement des données, estimation, tests d'hypothèses, analyse de la variance et étude de la régression. Les problématiques de fond, éclairées par des situations concrètes, sont clairement posées. Ainsi, chaque chapitre débute par un cas réel d'entreprise. Ce cas, qui est un défi relevé au fur et à mesure du chapitre, met en évidence à la fois les enjeux de la prise de décision dans l'entreprise et l'intérêt des concepts présentés. De très nombreux exercices, construits à partir de données réelles issues de l'actualité économique, sont entièrement résolus et commentés. Tout comme le défi, ces exercices permettent à la fois d'ancrer les connaissances des étudiants dans un contexte professionnel et de les motiver à franchir l'obstacle de la théorie. Le fil conducteur de l'ouvrage tient ainsi en une question : « comment traiter les données pour en extraire de l'information utile à la prise de décision ? ». Le livre respecte une courbe d'apprentissage progressive mais n'hésite pas à s'attarder sur des points théoriques délicats ou à proposer une approche originale de certains concepts. Par ailleurs, une rubrique spécifique est utilisée pour (ré)introduire aux moments opportuns, de manière très simple, des pré-requis indispensables, mais souvent mal maîtrisés par les étudiants.

Fourastie, J. and S. Levy (1993). Statistiques appliquées à l'économie, Paris : Masson

Cet ouvrage de cours de "statistique" est le résultat de recherches pédagogiques menées au Conservatoire National des Arts et de Métiers depuis 30 ans. Conçu pour des économistes et des comptables, il est orienté sur les concepts et les méthodes statistiques directement utilisables en économie et démographie : caractéristiques de tendance centrale et de dispersion séries chronologiques, corrélation linéaires, pensions, indices.

Urbain, M. Jorth J. S. (1994). Computer intensive statistical methods : Validation model selection and bootstrap, Londres : Chapman & Mall

Lavallee, P. d. and L. P. d. Rivest (2006). Méthodes d'enquêtes et sondages : pratiques européenne et nord-américaine, Paris : Dunod

Ce livre regroupe les méthodes statistiques appliquées aux enquêtes et sondages, présentées lors du

colloque francophone sur les sondages qui s'est tenu à Québec en 2005.

Grange, D., et al. (1993). Traitements statistiques des enquêtes, Paris : Dunod

Le traitement statistique des enquêtes se situe à un carrefour interdisciplinaire : les données issues de la socio-économie, la démographie, l'épidémiologie, le marketing, les sciences politiques y convergent pour alimenter des procédures et des logiciels. Les résultats sont ensuite livrés à l'interprétation de spécialistes qui manient des concepts et parlent des langages différents... Et pourtant, les outils utilisés sont, eux, étonnamment similaires, bien que rarement présentés simultanément dans ce contexte spécifique : l'enquête. C'est pourquoi les auteurs de ce recueil expliquent à la fois les mécanismes de ces procédures et logiciels, et le métier de ceux qui les mettent en oeuvre, avec ses règles et aussi ses tours de main, ses subtilités. Cet ouvrage s'adresse à ceux qui, pour leurs recherches, leurs études, leurs enseignements, sont confrontés aux enquêtes par sondage.

Duthion, R. and M. Guillemet (1990). 5 % de risque : méthodes statistiques, Rennes : E.N.S.P.

Ce manuel de statistique a été élaboré à l'intention des étudiants de l'Ecole Nationale de la Santé Publique (ENSP), Directeurs d'Hôpital, Directeurs d'Etablissements Sociaux etc... Il convient toutefois à un public plus vaste, professionnel ou étudiant, soucieux de mettre en oeuvre des techniques statistiques élémentaires de description et d'interprétation sur échantillon. Les méthodes d'estimation et de tests d'hypothèses sont exposés simplement. Elles s'appuient essentiellement sur une approche intuitive de ces mécanismes. Elles sont illustrées de nombreux exercices corrigés de difficulté progressive. L'approche retenue dans ce manuel est essentiellement pragmatique.

Cehessat, R. (1976). Exercices commentés de statistique et informatique appliquées, Paris : Dunod

Lebart, L. and J. P. Fenelon (1971). Statistique et informatique appliquée, Paris : Dunod

Lebart, L., et al. (1977). Techniques de la description statistique : méthodes et logiciels pour l'analyse des grands tableaux, Paris : Dunod

Lebart, L., et al. (1982). Traitement des données statistiques : méthodes et programmes, Paris : Dunod

## Probabilités – Mathématique

Caumel, Y. (2011). Probabilités et processus stochastiques, Paris : Springer

Ce livre a pour objectif de fournir au lecteur les bases théoriques nécessaires à la maîtrise des concepts et des méthodes utilisées en théorie des probabilités, telle qu'elle s'est développée au dix-septième siècle par l'étude des jeux de hasard, pour aboutir aujourd'hui à la théorisation de phénomènes aussi complexes et différents que les processus de diffusion en physique ou l'évolution des marchés financiers. Après un exposé introductif à la théorie probabiliste dont les liens avec l'analyse fonctionnelle et harmonique sont soulignés, l'auteur présente en détail une sélection de processus aléatoires classiques de type markoviens à temps entiers et continus, poissonniens, stationnaires, etc., et leurs diverses applications dans des contextes tels que le traitement du signal, la gestion des stocks, la modélisation des files d'attente, et d'autres encore. Le livre se conclut par une présentation détaillée du mouvement brownien et de sa genèse. Cent cinquante exercices (pour la plupart corrigés), ainsi qu'un ensemble de notes historiques ou épistémologiques permettant d'illustrer la dynamique et le contexte de découverte des théories évoquées, viennent compléter cet ouvrage (4e de couverture).

Saporta, G. (2006). Probabilités, analyse des données et statistique, Paris : Editions technip

La démarche statistique n'est pas seulement une auxiliaire des sciences destinée à valider ou non des modèles préétablis, c'est aussi une méthodologie indispensable pour extraire des connaissances à partir de données et un élément essentiel pour la prise de décision. La très large diffusion d'outils informatiques peut donner l'illusion de la facilité à ceux qui n'en connaissent pas les limites, alors que la statistique est plus que jamais un mode de pensée fondamental pour maîtriser la complexité, l'aléatoire et les risques, en donnant la prudence scientifique nécessaire. Ce manuel présente l'ensemble des connaissances utiles pour pouvoir pratiquer la statistique. Il est destiné à un vaste public (étudiants, chercheurs, praticiens de toutes disciplines) possédant le niveau d'algèbre et d'analyse d'un premier cycle universitaire scientifique ou économique. Cette nouvelle édition est une révision complète, avec des ajouts, de l'édition de 1990 et comporte de nombreux développements sur des méthodes récentes. Les 21 chapitres sont structurés en cinq parties : outils probabilistes, analyse exploratoire, statistique inférentielle, modèles prédictifs et recueil de données. On y trouve l'essentiel de la théorie des probabilités, les différentes méthodes d'analyse exploratoire des données (analyses factorielles et classification), la statistique "classique" avec l'estimation et les tests mais aussi les méthodes basées sur la simulation, la régression linéaire et logistique ainsi que des techniques non linéaires, la théorie des sondages et la construction de plans d'expériences.

Cocoza-Thivent, C. (1997). Processus stochastiques et fiabilité des systèmes, Paris : Springer

Ce livre, construit à partir des cours de DESS et de DEA de l'auteur, peut également intéresser des étudiants de maîtrise et des ingénieurs. Son fil directeur est la fiabilité et son but est de montrer concrètement ce que peut apporter l'étude des processus stochastiques dans ce domaine. Chemin faisant, cela permet d'aborder, dans des cas relativement simples, des techniques variées utilisées dans l'étude des processus stochastiques, tout en conservant l'esprit des démonstrations générales. Certaines parties sont d'inspiration très appliquée et peuvent être abordées par toute personne (étudiant, ingénieur) ayant des connaissances en probabilités- statistiques. D'autres font appel à des concepts plus pointus et offrent des ouvertures sur la recherche, certaines applications présentées ont d'ailleurs été obtenues récemment. Le plus souvent un thème donne lieu à deux chapitres: le premier présente les outils mathématiques, le second les applications en fiabilité (4e de couverture).

Fourastie, J. (1991). Mathématiques appliquées à l'économie, Paris : Dunod

Cet ouvrage, illustré de nombreux exemples concrets, présente les méthodes mathématiques utilisées en économie (calcul des probabilités, algèbre, analyse, théorie des graphes).

Archinard, G. and B. Guerrien (1988). Analyse mathématique pour économistes : cours et exercices corrigés, Paris : Economica

Principaux résultats d'analyse mathématique utiles en Sciences Economiques. Ce document présente des instruments mathématiques importants pour les économistes ainsi qu'un certain nombre de résultats de base de la microéconomie traditionnelle. A la fin de chaque chapitre, des exercices avec corrigés détaillés sont proposés pour faciliter l'assimilation des énoncés et résultats généraux.

Ciarlet, P. G., et al. (1996). Introduction à l'analyse numérique matricielle et à l'optimisation, Paris : Masson Editeur

Cet ouvrage est un classique de l'analyse numérique. Sous une forme très pédagogique, il propose l'essentiel sur les méthodes modernes de résolution des systèmes linéaires, sur le calcul des valeurs et vecteurs propres, et sur les méthodes d'optimisation.

Monfort, A. (1982). Cours de statistique mathématique, Paris : Economica

Monfort, A. (1980). Cours de probabilités, Paris : Economica  
Irdes - Pôle documentation - Marie-Odile Safon – Relecture : Anissa Afrite  
[www.irdes.fr/documentation/syntheses-et-dossiers-bibliographiques.html](http://www.irdes.fr/documentation/syntheses-et-dossiers-bibliographiques.html)  
[www.irdes.fr/documentation/syntheses/statistiques-et-econometrie.pdf](http://www.irdes.fr/documentation/syntheses/statistiques-et-econometrie.pdf)  
[www.irdes.fr/documentation/syntheses/statistiques-et-econometrie.epub](http://www.irdes.fr/documentation/syntheses/statistiques-et-econometrie.epub)



## Econométrie

Bourbonnais, R. and M. Terraza (2010). Analyse des séries temporelles : applications à l'économie et à la gestion. Manuel et exercices corrigés, Paris : Dunod

Cet ouvrage développe les méthodes d'analyse des séries temporelles : méthodes standard de traitement (régression, désaisonnalisation, lissage exponentiel) et techniques plus modernes (analyse spectrale, étude de stationnarisation, modèles ARIMA, modèles ARCH). Ces techniques s'appliquent à diverses disciplines, telles la prévision macroéconomique, la finance, le marketing... L'alternance systématique du cours et des exercices permet une mise en pratique rapide des connaissances avec l'utilisation de logiciels. Un site internet compagnon permet au lecteur de télécharger les séries statistiques utilisées et les programmes de traitement. Cette troisième édition propose de nouveaux exercices ainsi que les développements les plus récents de la discipline (4e de couverture).

Gourieroux, C. and A. Monfort (1983). Cours de series temporelles, Paris : Economica

Baltagi, B. H. (2005). Econometric analysis of panel data, Chichester : John Wiley & Sons

If you need to know how to perform estimation and inference on panel data from an econometric standpoint, then *Econometric Analysis of Panel Data* is the book to read. Besides being a leading graduate textbook, this book is the standard reference because it contains all the details you need to understand and implement the standard models. In addition, it provides a very good introduction to the newer and more-advanced techniques. This book provides an excellent introduction for the student or the applied researcher because of its attention to detail and its use of examples. The detail is especially useful in the many sections that grow out of Baltagi's own work. In these sections, readers gain a deep enough understanding of the models to implement them in a programming language like Stata. In other sections, such as the chapter on limited dependent variables, Baltagi combines a good introduction to the mechanics with an excellent introduction to the literature, which readers can follow up on if more details are required. This second edition covers many topics not touched on in the first edition. The new chapter on nonstationary panels stands out in this respect. Other topics, in particular limited dependent variables, receive much more attention.

Hsiao, C. (2005). Analysis of panel data, Cambridge : Cambridge University Press

Panel data models have become increasingly popular among applied researchers due to their heightened capacity for capturing the complexity of human behavior as compared to cross-sectional or time series data models. As a consequence, richer panel data sets also have become increasingly available. This second edition is a substantial revision of the highly successful first edition of 1986. Recent advances in panel data research are presented in a rigorous and accessible manner and are carefully integrated with the older material. The thorough discussion of theory and the judicious use of empirical examples make this book useful to graduate students and advanced researchers in economics, business, sociology, political science, etc. Other specific revisions include the introduction of the notion of strict exogeneity with estimators presented in a generalized method of moments framework, the notion of incidental parameters, more intuitive explanations of pairwise trimming, and discussion of sample selection dynamic panel models.

Matyas, L. (éd.) and P. Sevestre (éd.) (2008). The econometrics of panel data : fundamentals and recent developments in theory and practice, Berlin : Springer Verlag

This completely restructured, updated third edition of *The Econometrics of Panel Data*, first published in 1992, provides a general overview of the econometrics of panel data, both from a theoretical and from an applied viewpoint. Since the pioneering papers by Kuh, Mundlak, Hoch and Balestra and Nerlove, the pooling of cross section and time series data has become an increasingly popular way of quantifying economic relationships. Each series provides information lacking in the other, so a combination of both leads to more accurate and reliable results than would be achievable by one type of series alone. This third, enhanced edition provides a complete and up to date presentation of theoretical developments as well as surveys about how econometric tools are used to study firms and household's behaviors. It contains eleven entirely new chapters while the others have been largely revised to account for recent developments in the field (4e de couverture).

Sevestre, P. (2002). Econométrie des données de panel, Paris : Dunod

Le recours croissant à l'utilisation de données de panel - données constituées d'observations répétées sur un ensemble d'individus - est l'un des aspects marquants de l'évolution de l'économie appliquée au cours de ces vingt-cinq dernières années. L'objectif de cet ouvrage est de faire un tour d'horizon complet des bases fondamentales de l'économétrie des données de panel et de donner aux étudiants des outils et des méthodes leur permettant une application concrète. Les thèmes abordés sont les suivants : introduction à l'économétrie de données de panels, le modèle à effets fixes, le modèle à erreurs composées, le modèle à effets individuels corrélés, les modèles dynamiques, les erreurs de mesure et la simultanéité, les modèles à coefficients aléatoires et à coefficients composés, les modèles à variable dépendante qualitative.

Wooldridge, J. M. (2001). Econometric analysis of cross section and panel data, Cambridge : The MIT Press

Cet ouvrage de référence sur l'économétrie porte principalement sur les méthodes de coupe transversale et sur les données de panel.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data, Cambridge : The MIT Press

Cet ouvrage de référence sur l'économétrie porte principalement sur les méthodes de coupe transversale et sur les données de panel.

Angrist, J. D. and J. S. Pischke (2009). Mostly harmless econometrics : an empiricist's companion, Princeton : Princeton University

The core methods in today's econometric toolkit are linear regression for statistical control, instrumental variables methods for the analysis of natural experiments, and differences-in-differences methods that exploit policy changes. In the modern experimentalist paradigm, these techniques address clear causal questions such as: Do smaller classes increase learning? Should wife batterers be arrested? How much does education raise wages? Mostly Harmless Econometrics shows how the basic tools of applied econometrics allow the data to speak. In addition to econometric essentials, Mostly Harmless Econometrics covers important new extensions--regression-discontinuity designs and quantile regression--as well as how to get standard errors right. The authors explain why fancier econometric techniques are typically unnecessary and even dangerous. The applied econometric methods emphasized in this book are easy to use and relevant for many areas of contemporary social science (D'après l'éditeur).

Behaghel, L. (2012). Lire l'économétrie, Paris : Editions de la Découverte

Discipline au croisement de l'économie et de la statistique, l'économétrie reste pour beaucoup mystérieuse, alors qu'elle est de plus en plus souvent invoquée dans les débats politiques. L'objectif de ce guide est de rendre la lecture des textes économétriques accessible à des non-spécialistes. Le parcours, inspiré d'un enseignement auprès d'étudiants non économistes, comporte trois entrées. L'entrée par l'histoire montre les controverses suscitées par la discipline: l'économétrie permet-elle, conformément à

son projet initial, d'expliquer les phénomènes économiques, de tester les théories et de faire des prédictions? L'entrée par la méthodologie présente les principaux outils en privilégiant l'intuition sur la technique. Enfin et surtout, l'entrée par la lecture guidée d'articles scientifiques illustre l'apport de l'économétrie dans deux débats contemporains: la générosité du RMI est-elle une des causes du chômage? Les baisses de charges patronales sur les bas salaires permettent-elles de créer davantage d'emplois ? (4e de couverture).

Bourbonnais, R. (2000). Econométrie : manuel et exercices corrigés, Paris : Dunod

Cette troisième édition de l'ouvrage insiste particulièrement sur les concepts de l'économétrie moderne et présente de façon extrêmement pédagogique : - les domaines classiques de l'économétrie (modèle linéaire général, autocorrélation des erreurs, hétéroscédasticité, etc.) ; - les différents tests statistiques issus de l'économétrie ; - une introduction à l'analyse des séries temporelles (tests de Dickey-Fuller, méthodologie de Box-Jenkins) ; - la modélisation à plusieurs équations et les modèles VAR ; - la cointégration et le modèle à correction d'erreur. Les données présentées et les programmes de résolution des exercices sont disponibles par téléchargement à partir d'un site internet à l'adresse : <http://www.dauphine.fr/cip/pages/bourbonnais/>.

Call, S. T. and W. L. Holahan (1983). Microeconometrics, Belmont : Wadsworth Publishing Company

Cameron, A. C. and P. K. Trivedi (2005). Microeconometrics : methods and applications, Cambridge : Cambridge University Press

This book provides the most comprehensive treatment to date of microeconometrics, the analysis of individual-level data on the economic behavior of individuals or firms using regression methods for cross section and panel data. The book is oriented to the practitioner. A basic understanding of the linear regression model with matrix algebra is assumed. The text can be used for a microeconometrics course, typically a second-year economics PhD course; for data-oriented applied microeconometrics field courses; and as a reference work for graduate students and applied researchers who wish to fill in gaps in their toolkit. Distinguishing features of the book include emphasis on nonlinear models and robust inference, simulation-based estimation, and problems of complex survey data. The book makes frequent use of numerical examples based on generated data to illustrate the key models and methods. More substantially, it systematically integrates into the text empirical illustrations based on seven large and exceptionally rich data sets.

Cameron, A. C. and P. K. Trivedi (2009). Microeconometrics using Stata, Texas : Stata Press

This book is an outstanding introduction to microeconometrics and how to do microeconomic research using Stata. Aimed at students and researchers, it covers topics left out of microeconometrics textbooks and omitted from basic introductions to Stata. Cameron and Trivedi provide the most complete and up-to-date survey of microeconomic methods available in Stata. Early in the book, Cameron and Trivedi introduce simulation methods and then use them to illustrate features of the estimators and tests described in the rest of the book. While simulation methods are important tools for econometricians, they are not covered in standard textbooks. By introducing simulation methods, the authors arm students and researchers with techniques they can use in future work. The authors address each topic with an in-depth Stata example, and they reference their 2005 textbook, *Microeconometrics: Methods and Applications*, where appropriate. They also show how to use Stata's programming features to implement methods for which Stata does not have a specific command. Although the book is not specifically about Stata programming, it does show how to solve many programming problems. These techniques are essential in applied microeconometrics because there will always be new, specialized methods beyond what has already been incorporated into a software package. Cameron and Trivedi's choice of topics perfectly reflects the current practice of modern microeconometrics. After introducing the reader to Stata, the authors introduce linear regression, simulation, and generalized least-squares methods. The section on

cross-sectional techniques is thorough, with up-to-date treatments of instrumental-variables methods for linear models and of quantile-regression methods. The next section of the book covers estimators for the parameters of linear panel-data models. The authors' choice of topics is unique: after addressing the standard random-effects and fixed-effects methods, the authors also describe mixed linear models? a method used in many areas outside of econometrics.

Chow, G. C. (1983). Econometrics, Singapore : Mac Graw Hill

Dormont, B. (2007). Introduction à l'économétrie, Paris : Montchrestien

Cette Introduction à l'économétrie est destinée aux étudiants des masters d'économie quantitative et d'économétrie. Elle peut aussi intéresser les élèves des grandes écoles, les doctorants et les chercheurs désireux d'acquérir des bases en économétrie ou de rafraîchir leurs connaissances. Cet ouvrage a pour ambition de faire comprendre les principes des méthodes économétriques : les résultats théoriques sont démontrés en détail et les notions introduites illustrées par des applications. L'estimation du modèle linéaire de base est tout d'abord étudiée, avec une présentation des tests usuels. Les hypothèses initiales sont ensuite progressivement relâchées pour considérer des modèles plus réalistes, avec des perturbations hétéroscédastiques ou corrélées entre elles, ou avec des variables explicatives non exogènes. L'ouvrage se termine avec une présentation du traitement économétrique des données de panel. Les méthodes applicables à ces données peuvent être abordées simplement et permettent d'améliorer fortement l'analyse empirique (4e de couverture).

Fourgeaud, C. and B. Lenclud (1978). Econométrie, Paris : Presses Universitaires de France

Gourieroux, C. and A. Monfort (1989). Statistique et modèles économétriques, volume 1 : notions générales, estimation, prévision, algorithmes, Paris : Economica

Gourieroux, C. and A. Monfort (1989). Statistique et modèles économétriques, volume 2 : tests, régions de confiance, choix de modèles, théorie asymptotique, rappels de probabilités et d'algèbre linéaire, Paris : Economica

Greene, W. H. (1991). Econometric analysis, New-York : Mc Millon publishing Company  
<http://www.decitre.fr/livres/econometrie-9782744073373.html>

Gujarati, D. N. and B. Bernier (2012). Econométrie, Bruxelles : De Boeck

Très à jour sur le plan théorique, cet ouvrage est plus qu'une simple introduction à l'économétrie. Sont présentés tous les concepts et techniques fondamentaux, y compris ceux qui, considérés à tort comme moins importants, sont trop souvent omis dans la littérature mais se révèlent d'une indéniable utilité pratique. Ouvrage éminemment pédagogique, il ne requiert qu'un niveau de base en mathématique et en statistique. Des annexes à chaque chapitre permettent au lecteur qui le souhaite de trouver, d'une part des démonstrations de théorèmes et formules, et, d'autre part, une présentation de l'approche matricielle de la régression. Ce manuel - et c'est là un atout de poids - est illustré de nombreux exercices, problèmes et exemples tirés de données réelles. Certaines de ces données sont utilisées dans différents modèles, ce qui permet de montrer comment choisir la meilleure spécification. Il s'adresse principalement aux étudiants en sciences économiques et en gestion, mais également aux étudiants de plusieurs autres disciplines : sciences politiques, relations internationales, agriculture et sciences de la santé, ainsi qu'aux praticiens de l'économétrie (chargés d'études et statisticiens en entreprise) (4e de couverture).

Wooldridge, J. M. (2009). Introductory Econometrics : a modern approach, Mason : South-Western Cengage Learning

The modern approach of this book recognizes that econometrics has moved from a specialized mathematical description of economics to an applied interpretation based on empirical research

techniques. It bridges the gap between the mechanics of econometrics and modern applications of econometrics by employing a systematic approach motivated by the major problems facing applied researchers today. Throughout the content, the emphasis on examples gives a concrete reality to economic relationships and allows treatment of interesting policy questions in a realistic and accessible framework (4e de couverture.).

Wooldridge, J. M. (2015). [Introduction à l'économétrie : une nouvelle approche](#), Bruxelles : De Boeck

En recourant à de nombreuses applications empiriques, ce manuel d'introduction réussit à exploiter et à simplifier la présentation de l'économétrie sans renoncer aux exigences de rigueur et de cohérence requises au niveau universitaire. Les méthodes économétriques sont présentées avec l'objectif de répondre à des questions pratiques liées à l'analyse du comportement des agents économiques, l'évaluation de politiques publiques ou la réalisation de prévisions. Devenu une référence dans le monde anglo-saxon, cet ouvrage permet de comprendre et d'interpréter les hypothèses d'un modèle à la lumière de nombreuses applications empiriques. L'ouvrage distingue clairement le type de données analysées. Non seulement il couvre les données en coupe transversale et les séries chronologiques, mais il aborde également les données de panel dont l'utilisation est devenue très fréquente aujourd'hui. Ce livre offre également une introduction aux modèles à variable dépendante limitée qui sont d'une grande utilité en économie appliquée et en gestion. Chaque chapitre contient un large éventail d'exercices, dont un grand nombre repose sur l'utilisation de bases de données économiques disponibles sur le web. Le lecteur peut ainsi reproduire les nombreux exemples empiriques développés dans les chapitres de l'ouvrage et maîtriser toutes les étapes de la modélisation économétrique. Cet ouvrage intéressera non seulement les étudiants et professeurs de premier cycle universitaire, mais également les étudiants de Master et les praticiens de l'économie.

Jones, A. (2007). [Applied econometrics for health economists : a practical guide](#), Abingdon : Radcliffe Publishing Ltd

Applied Econometrics for Health Economists introduces readers to the appropriate econometric techniques for use with different forms of survey data, known collectively as microeconometrics. The book provides a complete illustration of the steps involved in doing microeconomic research. The only study to deal with practical analysis of qualitative and categorical variables, it also emphasises applied work, illustrating the use of relevant computer software applied to large-scale survey datasets. This is a comprehensive reference guide ? it contains a glossary of terms, a technical appendix, software appendix, references, and suggestions for further reading. It is concise and easy to read ? technical details are avoided in the main text and key terms are highlighted.

Jones, A. (2001). [Applied econometrics for health economists - a practical guide](#), Londres : OHE

Etant donné l'utilisation de plus en plus importante de données d'enquêtes à un niveau individuel en économie de la santé, il est important de comprendre les techniques économétriques qui sont disponibles et utilisables par les chercheurs. Il est également important de connaître les limites et pièges associés à chaque technique. Le but de cet ouvrage est de présenter les techniques économétriques à utiliser selon les données d'enquêtes - connues sous le vocable " microéconométriques". Le document insiste sur les travaux appliqués, pour montrer l'utilisation de logiciels pertinents sur de grands fichiers d'enquêtes, et pour illustrer les étapes d'une recherche micro-économétrique : formuler les problèmes empiriques, sélectionner la méthode économétrique appropriée ; faire attention aux méthodes d'estimation pour des variables qualitatives et limitées qui en sont dépendantes et aux logiciels à utiliser, construire des fichiers de données utilisables et connaître les limites de l'enquête, interpréter les résultats de l'analyse de façon compréhensible. Au sommaire : introduction ; l'enquête santé et de condition de vie ; variables dichotomiques (binaires) dépendantes ; le modèle Probit ordonné ; modèles multinomiaux ; modèle logistique (probit) dichotomique ; régression sur données quantitatives ; modèle de durée ; modèles sur données de panel. Un glossaire complète ce document. Le résumé et les thèmes couverts par cet ouvrage

sont disponibles sur le site internet de l'OHE :

<http://www.ohe.org/all%20publications.htm#Health%20Statistics>

Jones, A. (éd.) and O. O'Donnell (2002). Econometric analysis of health data, Chichester : John Wiley & Sons

Given extensive use of individual level data in Health Economics, it has become increasingly important to understand the microeconomic techniques available to applied researchers. The purpose of this book is to give readers convenient access to a collection of recent contributions that contain innovative applications of microeconomic methods to data on health and health care. Contributions are selected from papers presented at the European Workshops on Econometrics and Health Economics and published in Health Economics. Topics covered include : Latent Variables ; Unobservable heterogeneity and selection problems ; Count data and survival analysis ; Flexible and semiparametric estimators for limited dependent variables ; Classical and simulation methods for panel data ; Publication marks the tenth anniversary of the Workshop series.

Gourieroux, C. (1984). Econométrie des variables qualitatives, Paris : Economica

Lollivier, S. (2006). Econométrie avancée des variables qualitatives, Paris : Economica

L'économétrie des variables qualitatives s'est considérablement développée depuis le début des années 1980 et l'émergence des modèles de base (probit, logit). Les causes sont diverses : besoin d'une meilleure compréhension des comportements individuels, capacités de calcul accrues. L'ouvrage vise à faire le point sur les développements récents : données longitudinales et endogénéité des variables explicatives. De fait, un bon nombre des techniques classiques sur variables continues peut être étendu aux variables qualitatives, au prix néanmoins d'hypothèses paramétriques plus restrictives et d'une complexité accrue des calculs.

Maddala, G. S. (1999). Limited-dependent and qualitative variables in econometrics, Cambridge : Cambridge University Press

Kumbhakar, S. C. and Knox Lovell C.A. (2004). Stochastic frontier analysis, Cambridge : Cambridge University Press

This book develops economic techniques for the estimation of production, cost and profit frontiers, and for the estimation of the technical and economic efficiency with which producers approach these frontiers. Since these frontiers envelop rather than intersect the data, and since the authors continue to maintain the traditional econometric belief in the presence of external forces contributing to random statistical noise (4e de couverture).

Greene, W. H. and D. A. Hensher (2010). Modeling ordered choices. A primer, Cambridge : Cambridge University Press

It is increasingly common for analysts to seek out the opinions of individuals and organizations using attitudinal scales such as degree of satisfaction or importance attached to an issue. Examples include levels of obesity, seriousness of a health condition, attitudes towards service levels, opinions on products, voting intentions, and the degree of clarity of contracts. Ordered choice models provide a relevant methodology for capturing the sources of influence that explain the choice made amongst a set of ordered alternatives. The methods have evolved to a level of sophistication that can allow for heterogeneity in the threshold parameters, in the explanatory variables (through random parameters), and in the decomposition of the residual variance. This book brings together contributions in ordered choice modeling from a number of disciplines, synthesizing developments over the last fifty years, and suggests useful extensions to account for the wide range of sources of influence on choice.

Pudney, S. (1989). Modelling individual choice : the econometrics of corners, links and holes, Cambridge Oxford : Blackwell

Winkelmann, R. (2008). Econometric analysis of count data, Berlin : Springer Verlag

The book provides graduate students and researchers with an up-to-date survey of statistical and econometric techniques for the analysis of count data, with a focus on conditional distribution models. Proper count data probability models allow for rich inferences, both with respect to the stochastic count process that generated the data, and with respect to predicting the distribution of outcomes. The book starts with a presentation of the benchmark Poisson regression model. Alternative models address unobserved heterogeneity, state dependence, selectivity, endogeneity, underreporting, and clustered sampling. Testing and estimation is discussed from frequentist and Bayesian perspectives. Finally, applications are reviewed in fields such as economics, marketing, sociology, demography, and health sciences. The fifth edition contains several new topics, including copula functions, Poisson regression for non-counts, additional semi-parametric methods, and discrete factor models. Other sections have been reorganized, rewritten, and extended (4e de couverture).

## Modèles multiniveaux

Lott, J. L. (éd.) and J. S. Antony (éd.) (2012). Multilevel modeling techniques and applications in institutional research, San Francisco : John Wiley & Sons

Multilevel modeling is an increasingly popular multivariate technique that is widely applied in the social sciences. Increasingly, practitioners are making instructional decisions based on results from their multivariate analyses, which often come from nested data that lend themselves to multilevel modeling techniques. As data-driven decision making becomes more critical to colleges and universities, multilevel modeling is a tool that will lead to more efficient estimates and enhance understanding of complex relationships. This volume illustrates both the theoretical underpinnings and practical applications of multilevel modeling in IR. It introduces the fundamental concepts of multilevel modeling techniques in a conceptual and technical manner. Providing a range of examples of nested models that are based on linear and categorical outcomes, it then offers important suggestions about presenting results of multilevel models through charts and graphs (4e de couverture).

Wang, J., et al. (2012). Multilevel models. Applications using SAS, Berlin : De Gruyter

Interest in multilevel statistical models for social science and public health studies has been aroused dramatically since the mid-1980s. New multilevel modeling techniques are giving researchers tools for analyzing data that have a hierarchical or clustered structure. Multilevel models are now applied to a wide range of studies in sociology, population studies, education studies, psychology, economics, epidemiology and public health. This book covers a broad range of topics about multilevel modeling. The goal of the authors is to help those who are interested in analysis of multilevel data to understand the basic concepts, theoretical frameworks and application methods of multilevel modelling. The book is focusing on the methods and application of various multilevel models, using the internationally widely used statistical software, SAS (d'après la 4<sup>ème</sup> de couv.)

Hesketh, S. and A. Skrondal (2008). Multilevel and longitudinal modeling using Stata, Texas : Stata Press

Multilevel and Longitudinal Modeling Using Stata, by Sophia Rabe-Hesketh and Anders Skrondal, looks specifically at Stata's treatment of generalized linear mixed models, also known as multilevel or hierarchical models. These models are "mixed" because they allow fixed and random effects, and they are "generalized" because they are appropriate for continuous Gaussian responses as well as binary, count,

and other types of limited dependent variables. Beginning with the comparatively simple random-intercept linear model without covariates, Rabe-Hesketh and Skrondal develop the mixed model from principles, thereby familiarizing the reader with terminology, summarizing and relating the widely used estimating strategies, and providing historical perspective. Once the authors have established the mixed-model foundation, they smoothly generalize to random-intercept models with covariates and then to random-coefficient models. The middle chapters of the book apply the concepts for Gaussian models to models for binary responses (e.g., logit and probit), ordinal responses (e.g., ordered logit and ordered probit), and count responses (e.g., Poisson). The authors then consider models with multiple levels of random variation and models with crossed (nonnested) random effects. In its examples and end-of-chapter exercises, the book contains real datasets and data from the medical, social, and behavioral sciences literature. The book has several applications of generalized mixed models performed in Stata. Rabe-Hesketh and Skrondal developed gllamm, a Stata program that can fit many latent-variable models, of which the generalized linear mixed model is a special case. With the release of version 9, Stata introduced the xtmixed command for fitting linear (Gaussian) mixed models. Stata users can use gllamm and xtmixed in conjunction with the rest of the xt suite of commands to perform comparative mixed-model analyses for various response families. The type of models fit by these commands sometimes overlap; when this happens, the authors highlight the differences in syntax, data organization, and output for the two (or more) commands that can be used to fit the same model. The authors also point out the relative strengths and weaknesses of each command when used to fit the same model, based on considerations such as computational speed, accuracy, and available predictions and postestimation statistics. The book delineates the relationship between gllamm and xtmixed, clearly showing how they complement one another.

Gelman, A. and J. Hill (2007). Data analysis using regression and multilevel/hierarchical models, New York : Cambridge University Press

Data Analysis Using Regression and Multilevel/Hierarchical Models is a comprehensive manual for the applied researcher who wants to perform data analysis using linear and nonlinear regression and multilevel models. The book introduces a wide variety of models, whilst at the same time instructing the reader in how to fit these models using available software packages. The book illustrates the concepts by working through scores of real data examples that have arisen from the authors' own applied research, with programming codes provided for each one. Topics covered include causal inference, including regression, poststratification, matching, regression discontinuity, and instrumental variables, as well as multilevel logistic regression and missing-data imputation. Practical tips regarding building, fitting, and understanding are provided throughout. Author resource page: <http://www.stat.columbia.edu/~gelman/arm/>

Courgeau, D. (2004). Du groupe à l'individu : synthèse multiniveau, Paris : INED

Ce manuel présente une vue historique de l'évolution de la pensée démographique, de ses origines au XVIIe siècle à nos jours, pour montrer comment l'approche multiniveau résout certaines des contradictions apparues au cours du temps et permet une synthèse des différentes approches suivies. Une comparaison plus générale avec les autres sciences sociales y est aussi tentée. L'introduction pose le problème de fond que l'ensemble du volume va chercher à résoudre : l'opposition entre société et individu. Une première partie explore la façon dont ces approches ont été suivies depuis l'origine de la démographie : approche transversale, approche longitudinale, biographies, analyse multiniveau. Une seconde partie examine plus en détail les divers modèles qu'une approche multiniveau peut utiliser, en indiquant leur formalisation mathématique en fonction notamment du type de variable en jeu, les problèmes que l'on peut rencontrer en les utilisant et les questions qui restent à résoudre. La conclusion cherche à dépasser cette approche multiniveau en proposant un regard nouveau sur l'utilisation des points de vue objectiviste, subjectiviste et logiciste en probabilités pour traiter des modèles hiérarchiques en sciences sociales. Elle propose également une théorie plus complète des sciences humaines, qui permet de comprendre, non seulement comment un comportement observé à un niveau d'agrégation



peut s'expliquer en faisant intervenir simultanément divers niveaux, mais plus généralement l'évolution au cours du temps et dans l'espace de ces différents niveaux.

Gauvin, L. and C. Dassa (2004). "L'analyse multiniveaux : avancées récentes et retombées anticipées pour l'étude des inégalités sociales et de santé." Santé Société Et Solidarité : Revue De L'observatoire Franco-Québécois(2): 187-195.

L'essor récent des études sur les inégalités sociales et de santé permet de mettre en relief la nécessité d'exploiter certaines avancées méthodologiques et statistiques afin de contribuer à l'avancement des connaissances dans ce domaine. Cet article vise à promouvoir l'utilisation de l'analyse multi-niveaux dans l'étude des inégalités sociales et de santé. Les objectifs spécifiques visent à décrire un éventail de nouvelles applications pour ce type d'analyse et d'expliquer comment celles-ci pourraient contribuer à l'étude des processus sous-jacents aux inégalités sociales et de santé. A cette fin, les auteurs décrivent des modèles d'analyse multi-niveaux complexes. Cette description est accompagnée d'une recension d'écrits récents qui illustrent certaines applications avancées de l'analyse multi-niveaux. En conclusion, les auteurs discutent des retombées prévisibles d'une utilisation judicieuse de l'analyse multiniveaux (résumé d'auteur).

Skrondal, A. and S. Rabe-Hesketh (2004). Generalized latent variable modeling : multilevel, longitudinal, and structural equation models, Boca Raton : Chapman & Hall/CRC

Skrondal (epidemiology, Norwegian Institute of Public Health) and Rabe-Hesketh (educational statistics, U. of California-Berkeley) unify and extend latent variable modeling, considering not only the types mentioned in the subtitle, but also their kith and kin such as generalized linear mixed models, random coefficient models, item response models, and frailty models. Believing that progress is hampered by localized disciplinary jargon, they seek to synthesize approaches from different disciplines and to translate between the languages of statistics, biometrics, psychometrics, and econometrics. They identify the chapters that focus on methodology and those that emphasize interpretation and analysis, so readers can choose what they want.

Snijders, T. (2002). "Analyse multiniveaux : recherche en cours." Bulletin De Methodologie Sociologique (Bms)(73): 31-41.

Cet article sur l'analyse multiniveaux est basé sur un cours d'une journée donné par l'auteur à Lille et à Paris pour des sociologues francophones, qui commençaient tout juste à se familiariser avec cette méthode. Il démontre que l'analyse multiniveaux est un ensemble de techniques statistiques conçues pour l'analyse de données : de structure emboîtée, donc de plusieurs niveaux ordonnés de nature hiérarchique, chaque niveau correspondant à un type d'entité d'analyse, utilisant des modèles statistiques à plusieurs sources de variance. Les interférences statistiques dans l'analyse multiniveaux portent surtout sur plusieurs ensembles d'entités d'analyse, correspondant aux différentes sources de variation.

Leyland, A. H. (éd.) and H., Goldstein (éd.)(2001). Multilevel modelling of health statistics, Chichester : John Wiley & Sons

Multilevel modelling facilitates the analysis of hierarchical data where observations may be nested within higher levels of classification. In health care research, for example, a study may be undertaken to determine the variability of patient outcomes where these also vary by hospital or health care region. Inference can then be made on the efficacy of health care practices. This book provides the reader with the analytical techniques required to study such data sets.

Snijders, T. A. B. and R. J. Bosker (1999). Multilevel analysis : an introduction to basic and advanced multilevel modeling, Londres : Sage Publications

The main methods, techniques and issues for carrying out multilevel modeling and analysis are covered in this book. The book is an applied introduction to the topic, providing a clear conceptual understanding of the issues involved in multilevel analysis and will be a useful reference tool. Information on designing multilevel studies, sampling, testing and model specification and interpretation of models is provided. A comprehensive guide to the software available is included. Multilevel Analysis is the ideal guide for researchers and applied statisticians in the social sciences, including education, but will also interest researchers in economics, and biological, medical and health disciplines.

Rice, N. and A. Jones (1997). "Multilevel models and health economics." Health Economics 6(6): 561-575.

Multilevel analyses have become an accepted statistical technique in the field of education where over the past decade or so the methods have been developed to explore the relationships between pupil characteristics and the characteristics of the schools they attend. More recently, widespread use has extended to other social sciences and health research. However, to date, little use has been made of these techniques within the health economics literature. This paper presents an introductory account of multilevel models and describes some of the areas of health economics research that may benefit from their use.

De, Leeuw J. and I. G. G. Kreft (1995). "Questioning multilevel models." Journal of Educational and Behavioral Statistics 20(2): 171-189.

Practical problems with multilevel techniques are discussed. These problems relate to terminology, computer programs employing different algorithms, and interpretations of the coefficients in either one or two steps. The usefulness of hierarchical linear models (HLMs) in common situations in educational research is explored. While elegant, HLMs are not always necessary. (SLD)

Raudenbush, S. W. and A. S. Bryk (2002). Hierarchical linear models : applications and data analysis methods, Thousand Oaks : Sage publications

Raudenbush, S., et al. (2001). HLM 5 : hierarchical linear and nonlinear modeling, Lincolnwood : Scientific Software International

## Autres méthodes

Sericola, B. (2013). Chaînes de Markov : théorie, algorithmes et applications, Paris : Lavoisier ; Paris : Hermès Science

Les chaînes de Markov sont des modèles probabilistes utilisés dans des domaines variés comme la logistique, l'informatique, la fiabilité, les télécommunications, ou encore la biologie et la physique-chimie. On les retrouve également dans la finance, l'économie et les sciences sociales. Cet ouvrage présente une étude approfondie des chaînes de Markov à temps discret et à temps continu avec des applications détaillées aux processus de naissance et mort et aux files d'attente. Ces applications sont illustrées par des algorithmes généraux de calcul de probabilités d'état et de distribution de temps de passage. Le développement de ces algorithmes repose sur l'utilisation de la technique d'uniformisation des chaînes de Markov qui est présentée de manière théorique et intuitive. Ce livre s'adresse aux ingénieurs et chercheurs ayant besoin de modèles probabilistes pour évaluer et prédire le comportement des systèmes qu'ils étudient ou qu'ils développent (4e de couverture).

Nuel, G. and B. Prum (2007). Analyse statistique des séquences biologiques : modélisation markovienne, alignements et motifs, Paris : Lavoisier ; Paris : Hermès Science

L'ouvrage Analyse statistique des séquences biologiques est le résultat de quinze années de recherche et présente de manière synthétique et didactique une approche nouvelle fondée sur les automates finis. L'analyse statistique des séquences génomiques - un des points forts de la recherche en bioinformatique ou en biostatistique en France - développe des outils permettant la modélisation des séquences (chaînes de Markov, chaînes de Markov cachées) ainsi que l'étude des occurrences de motifs (par une approche novatrice fondée sur les automates finis déterministes). Elle permet également la recherche automatique de gènes et d'autres signaux biologiques (annotation) et compare des séquences provenant d'espèces différentes (alignement), en particulier pour reconstruire l'histoire de leur évolution (phylogénie). Une présentation du domaine biologique et des divers modèles markoviens utilisés permet une lecture autonome. Les méthodes sont illustrées par de nombreux exemples et figures, les algorithmes qui leur correspondent sont systématiquement détaillés et l'on donne en référence les sites web et programmes qui permettent leur mise en œuvre pratique (4e de couverture).

Nuel, G. and B. Prum (2007). Analyse statistique des séquences biologiques : modélisation markovienne, alignements et motifs, Paris : Lavoisier ; Paris : Hermès Science

L'ouvrage Analyse statistique des séquences biologiques est le résultat de quinze années de recherche et présente de manière synthétique et didactique une approche nouvelle fondée sur les automates finis. L'analyse statistique des séquences génomiques - un des points forts de la recherche en bioinformatique ou en biostatistique en France - développe des outils permettant la modélisation des séquences (chaînes de Markov, chaînes de Markov cachées) ainsi que l'étude des occurrences de motifs (par une approche novatrice fondée sur les automates finis déterministes). Elle permet également la recherche automatique de gènes et d'autres signaux biologiques (annotation) et compare des séquences provenant d'espèces différentes (alignement), en particulier pour reconstruire l'histoire de leur évolution (phylogénie). Une présentation du domaine biologique et des divers modèles markoviens utilisés permet une lecture autonome. Les méthodes sont illustrées par de nombreux exemples et figures, les algorithmes qui leur correspondent sont systématiquement détaillés et l'on donne en référence les sites web et programmes qui permettent leur mise en œuvre pratique (4e de couverture).

Berchtold, A. (2005). Chaînes de Markov et modèles de transition : applications aux sciences sociales, Paris : Hermès

Cet ouvrage est consacré à l'étude des chaînes de Markov et des modèles de transition. Il présente de façon claire et précise un ensemble de méthodes d'analyse novatrices ayant un très large champ d'application en sciences sociales et dans tous les domaines travaillant avec des données catégorielles. Prenant comme point de départ l'analyse des tables de contingence, l'auteur les transforme de façon à exploiter au maximum l'information qu'elles renferment. Des mesures de qualité et de pouvoir prédictif des données sont présentées, ainsi que des méthodes permettant une analyse dynamique des relations entre variables.

Pearl, J. (2000). Causality : models, reasoning, and inference, Cambridge : Cambridge University Press  
<http://www.cambridge.org/catalogue/catalogue.asp?isbn=9780521773621>

Causality offers the first comprehensive coverage of causal analysis in many sciences, including recent advances using graphical methods. Pearl presents a unified account of the probabilistic, manipulative, counterfactual and structural approaches to causation, and devises simple mathematical tools for analyzing the relationships between causal connections, statistical associations, actions and observations. The book will open the way for including causal analysis in the standard curriculum of statistics, artificial intelligence, business, epidemiology, social science and economics.

Zelterman, D. (1999). Models for discrete data, Oxford : Oxford University Press

Cox, D. R. and N. Wermuth (1998). Multivariate dependencies : models, analysis and interpretation, Boca Raton : Chapman & Hall/CRC

Large observational studies involving research questions that require the measurement of several features on each individual arise in many fields including the social and medical sciences. Multivariate dependencies set out both the general concepts and the more technical statistical issues are described in outline and four studies are discussed in some detail. The use of graphical representations of dependencies and independencies among the features under study is stressed, both incorporate available knowledge at the planning stage of an analysis and to summarize aspects important for interpretation after detailed statistical analysis is complete.

Tenenhaus, M. (1998). La régression PLS : théorie et pratique, Paris : Editions Technip

La régression PLS (Partial Least Squares) est une méthode d'analyse des données spécifiquement construite pour décrire les relations très complexes entre des variables d'entrées-sorties X et Y, en l'absence d'un modèle théorique. Sur le plan théorique, cet ouvrage a trois objectifs : situer la régression PLS parmi les méthodes d'association et de prédiction en analyse des données : analyse canonique, analyse factorielle inter-batteries, analyse des redondances, algorithme NIPALS, algorithme SIMPLS et approche PLS ; décrire l'algorithme de régression PLS dans sa forme originale telle qu'elle est programmée dans des logiciels comme SIMCA ou The Unscrambler ; présenter en détail les principales propriétés mathématiques de la régression PLS, car leur connaissance est essentielle pour une bonne utilisation de la méthode.

Robert, C. (1996). Méthodes de Monte Carlo par chaînes de Markov, Paris : Economica

Ce livre présente les éléments de base de la simulation de lois de probabilité et leur utilisation en statistique.

Efron, B. and R. J. Tibshirani (1993). An introduction to the bootstrap, New York : Chapman & Hill

Ce livre constitue le premier ouvrage de synthèse réalisé sur le "bootstrap", méthode de régression statistique basée sur informatique.

Robert, C. (1992). L'analyse statistique bayésienne, Paris : Economica

Ce livre présente l'approche bayésienne, qui constitue une branche importante de la statistique paramétrique. Les méthodes bayésiennes sont étudiées sous l'angle de la théorie de la décision. Les derniers chapitres abordent des développements théoriques récents (équivalence, classes complètes, modèles hiérarchiques et empiriques...) ainsi qu'un exposé des techniques de calcul des estimateurs de Bayes. De nombreux exercices et exemples viennent compléter chaque chapitre de cet ouvrage.

Newsom, J. T. é., et al. (2012). Longitudinal data analysis : a practical guide for researchers in aging, health, and social sciences, New York : Routledge

This book provides accessible treatment to state-of-the-art approaches to analyzing longitudinal studies. Comprehensive coverage of the most popular analysis tools allows readers to pick and choose the techniques that best fit their research. The analyses are illustrated with examples from major longitudinal data sets including practical information about their content and design. Illustrations from popular software packages offer tips on how to interpret the results. Each chapter features suggested readings for additional study and a list of articles that further illustrate how to implement the analysis and report the results. Syntax examples for several software packages for each of the chapter examples are provided at [www.psyppress.com/longitudinal-data-analysis](http://www.psyppress.com/longitudinal-data-analysis). Although many of the examples address health or social science questions related to aging, readers from other disciplines will find the analyses relevant to their work. In addition to demonstrating statistical analysis of longitudinal data, the book shows how to interpret and analyze the results within the context of the research design. The methods covered in this

book are applicable to a range of applied problems including short- to long-term longitudinal studies using a range of sample sizes. The book provides non-technical, practical introductions to the concepts and issues relevant to longitudinal analysis. Topics include use of publicly available data sets, weighting and adjusting for complex sampling designs with longitudinal studies, missing data and attrition, measurement issues related to longitudinal research, the use of ANOVA and regression for average change over time, mediation analysis, growth curve models, basic and advanced structural equation models, and survival analysis. An ideal supplement for graduate level courses on data analysis and/or longitudinal modeling taught in psychology, gerontology, public health, human development, family studies, medicine, sociology, social work, and other behavioral, social, and health sciences, this multidisciplinary book will also appeal to researchers in these fields (4e de couverture).

## Analyses de données

Morgan, S. L. and C. Winship (2007). Counterfactuals and causal inference : methods and principles for social research, New York : Cambridge University Press

Did mandatory busing programs in the 1970s increase the school achievement of disadvantaged minority youth? Does obtaining a college degree increase an individual's labor market earnings? Did the use of the butterfly ballot in some Florida counties in the 2000 presidential election cost Al Gore votes? If so, was the number of miscast votes sufficiently large to have altered the election outcome? At their core, these types of questions are simple cause-and-effect questions. Simple cause-and-effect questions are the motivation for much empirical work in the social sciences. This book presents a model and set of methods for causal effect estimation that social scientists can use to address causal questions such as these. The essential features of the counterfactual model of causality for observational data analysis are presented with examples from sociology, political science, and economics (4e de couverture).

Hill, C., et al. (2004). Analyse statistique des données de survie, Paris : Médecine Sciences Flammarion

Les expressions modèle de Cox, test du logrank, estimation de Kaplan-Meier, survie actuarielle sont de plus en plus souvent présentées dans la littérature médicale internationale. Cette terminologie se rapporte à l'analyse statistique des données de survie, ou plus généralement des données liées au temps : détail de cicatrisation ou délai jusqu'à une rechute par exemple. Ce livre s'adresse aussi bien aux cliniciens ou chercheurs qui désirent mieux comprendre les analyses de survie présentées dans la littérature qu'aux biostatisticiens, médecins ou non, qui ont à analyser les données de survie. Les méthodes présentées sont décrites de façon relativement complète ce qui nécessite un certain formalisme mathématique. Cependant, chaque méthode est illustrée par un exemple numérique concret développé jusqu'à l'interprétation des résultats reposant en général sur des données publiées.

Singer, J. D. and J. B. Willet (2003). Applied Longitudinal Data Analysis : Modeling Change and Event Occurrence, New York : Oxford University Press

Cet ouvrage est un manuel de référence à l'usage des chercheurs, qui utilisent les méthodes quantitatives pour l'analyse de données longitudinales dans le domaine social et comportemental. S'appuyant sur des exemples concrets, il présente notamment les modèles multi-niveaux applicables aux changements individuels, ainsi que les modèles de mortalité applicables aux événements.

Diggle, P. J., et al. (2002). Analysis of longitudinal data, New York : Oxford University Press

The new edition of this important text has been completely revised and expanded to become the most up-to-date and thorough professional reference text in this fast-moving and important area of biostatistics. Two new chapters have been added on fully parametric models for discrete repeated

measures data and on statistical models for time-dependent predictors where there may be feedback between the predictor and response variables. It also contains the many useful features of the previous edition such as, design issues, exploratory methods of analysis, linear models for continuous data, and models and methods for handling data and missing values.

Volle, M. (1997). Analyse des données, Paris : Economica

L'analyse des données recouvre un ensemble de méthodes de statistique descriptive (ou statistique exploratoire) dont le développement est assez récent. Le présent ouvrage vise à en fournir une description aussi pédagogique et cohérente que possible. Il décrit les deux grands volets de l'analyse de données (analyse factorielle et classification automatique) et souligne les relations existant entre eux, notamment en les rapportant tous deux à la théorie de l'information. Cet ouvrage donne une description des pratiques renforcée par la présentation d'exemples, et montre les propriétés logiques des méthodes ; l'expérience montre, en effet, que l'utilisation correcte de l'analyse des données suppose à la fois une claire connaissance de ses aspects logiques, et de la maîtrise de quelques trucs de métier pratiques.

Foucart, T. (1985). Analyse factorielle : programmation sur micro-ordinateur, Paris : Masson

Lebart, L., et al. (1995). Statistique exploratoire multidimensionnelle, Paris : Dunod

Destiné à un public assez large, appuyé sur de nombreux exemples, l'ouvrage présente les concepts de base et les fondements des méthodes exploratoires de base (analyses factorielles, classification), et rend compte aussi des travaux les plus récents. Les auteurs insistent sur la complémentarité de ces méthodes et sur leur insertion dans l'arsenal des méthodes statistiques plus classiques. L'articulation avec les techniques explicatives et prévisionnelles, la validité et la stabilité des résultats sont deux points sensibles largement développés. L'ouvrage peut-être lu à plusieurs niveaux : celui du praticien, celui de l'utilisateur exigeant, enfin celui du chercheur en méthodologie statistique.

Dazy, F. and Le Barzic J. F. (1996). L'analyse des données évolutives. Méthodes et applications, Paris : Editions Technip

Cet ouvrage du GERI se propose d'assurer une meilleure diffusion des connaissances dans le domaine de l'analyse des données évolutives, en rapprochant les différentes méthodes existantes parmi les plus utilisées (STATIS, AFM, DACP, et ASCM). La première partie de cet ouvrage est consacrée à l'analyse des données classiques (ACP, AFC), la seconde partie porte sur les méthodes d'analyse de données évolutives, la troisième partie fait le point des méthodes de classification tout particulièrement en analyse de données évolutives.

Escofier, B. and J. Pages (1993). Analyses factorielles simples et multiples - Objectifs, méthodes et interprétation, Paris : DUNOD

Cet ouvrage fait le point sur les méthodes factorielles, qui représentent les plus vastes possibilités d'application. Aux méthodes classiques d'analyse d'un seul tableau on ajoute les méthodes récentes d'analyse de tableaux multiples.

Nakache, J. P. and J. Confais (2005). Approche pragmatique de la classification, Paris : Editions Technip

La classification est une branche de l'analyse statistique multidimensionnelle descriptive qui a fait l'objet de très nombreuses publications. Elle connaît, ces dernières années, un renouvellement et un développement considérables avec la multiplication de bases de données de plus en plus importantes, nécessitant une exploration fouillée (data mining) avant d'adopter un modèle probabiliste suggéré par les résultats. Les techniques de classification font appel à une démarche algorithmique et non à des techniques mathématiques complexes : les classes, obtenues après des opérations simples et répétitives,

sont souvent faciles à décrire et à caractériser. Ce manuel pratique présente un large éventail de méthodes de classification, des plus classiques aux plus récentes, regroupées en trois types : classification hiérarchique ascendante et descendante ; construction d'une partition unique fondée sur différentes notions : similarité, densité, modèle probabiliste ou neuronal, graphique ; classification d'un ensemble de variables. La plupart des méthodes implémentées dans plusieurs logiciels statistiques est largement illustrée et l'interprétation des résultats occupe une place importante. Cet ouvrage intéressera les praticiens confrontés, dans leurs travaux, à des données multidimensionnelles importantes et exerçant dans de nombreux domaines : médecine, sciences sociales, industrie, marketing, psychologie, météorologie, documentation, etc. Il s'adressera également aux enseignants, chercheurs, ingénieurs, étudiants et pourra servir de support de cours dans les universités et les grandes écoles.

Robert, C. (1992). L'analyse statistique bayésienne, Paris : Economica

Ce livre présente l'approche bayésienne, qui constitue une branche importante de la statistique paramétrique. Les méthodes bayésiennes sont étudiées sous l'angle de la théorie de la décision. Les derniers chapitres abordent des développements théoriques récents (équivalence, classes complètes, modèles hiérarchiques et empiriques...) ainsi qu'un exposé des techniques de calcul des estimateurs de Bayes. De nombreux exercices et exemples viennent compléter chaque chapitre de cet ouvrage.

Benzecri, J. P. and F. Benzecri (1986). Pratique de l'analyse de données -5- Economie, Paris : Dunod

Volle, M. (1985). Analyse des données, Paris : Economica

Présentation des méthodes de statistique descriptive (ou statistique exploratoire) et description des deux grands volets de l'analyse des données : analyse factorielle et classification automatique.

Benzecri, J. P. and F. Benzecri (1984). Pratique de l'analyse des données 1 : analyse des correspondances et classification, Paris : Dunod

Fenelon, J. P. (1981). Qu'est-ce que l'analyse des données ?, Paris : Lefonen

Lerman, I. C. (1981). Classification et analyse ordinaire des données, Paris : Dunod

Foucart, T. (1985). Analyse factorielle : programmation sur micro-ordinateur, Paris : Masson

Jérôme Pagès (2014). Multiple Factor Analysis by Example Using R

## Data mining

Tuffery, S. (2012). Data mining et statistique décisionnelle : l'intelligence des données, Paris : Editions Technip

Le data mining et la statistique sont de plus en plus répandus dans les entreprises et les organisations soucieuses d'extraire l'information pertinente de leurs bases de données, qu'elles peuvent utiliser pour expliquer et prévoir les phénomènes qui les concernent (risques, consommation, fidélisation...). Cette quatrième édition, actualisée et augmentée de 120 pages, fait le point sur le data mining, ses fondements théoriques, ses méthodes, ses outils et ses applications, qui vont du scoring jusqu'au web mining et au text mining. Nombre de ses outils appartiennent à l'analyse des données et la statistique "classique" (analyse factorielle, classification automatique, analyse discriminante, régression logistique, modèles linéaires généralisés, régression pénalisées...) mais certains sont plus spécifiques au data mining, comme les arbres de décision, les réseaux de neurones, les SVM, l'agrégation de modèles et la détection des règles d'associations. Ces outils sont disponibles dans des logiciels de plus en plus puissants et conviviaux,

aptes à exécuter de nombreux algorithmes sur de grands volumes de données. Un chapitre de l'ouvrage aide le lecteur à se diriger dans cette offre logicielle et détaille les fonctionnalités des trois principaux logiciels : R, SAS, IBM et SPSS. Ces logiciels sont aussi utilisés pour illustrer par des exemples de nombreuses explications théoriques : une partie de 50 pages est consacrée à une étude de cas complète de credit scoring, qui va de l'exploration des données jusqu'à l'élaboration de la grille de score. Les aspects méthodologiques vont de la conduite des projets jusqu'aux facteurs de réussite et aux pièges à éviter, en passant par l'évaluation et la comparaison des modèles, leur intégration dans les processus opérationnels, sans oublier les contraintes juridiques dès que l'on traite des données à caractère personnel.

## La statistique avec SAS

Decourt, O. and H. Gerard (2011). SAS l'essentiel, Paris : Dunod

Ce livre a été conçu pour maîtriser les usages les plus courants de SAS (Statistical Analysis System) que sont l'extraction de données, les requêtes, les jointures, les statistiques descriptives ou le paramétrage. Plus de 90% des informations contenues dans ce livre sont valables pour toutes les versions de SAS car elles se rapportent à des besoins que l'éditeur du logiciel a pris en compte depuis de nombreuses années. L'ouvrage comporte une introduction au code (en langage SAS, en requêtes SQL ou en langage macro) pour ceux qui veulent s'initier à ce type de programmation.

Decourt, O. and P. Letren (2008). Reporting avec SAS : mettre en forme et diffuser vos résultats avec SAS 9 et SAS 9 B1, Paris : Dunod

Cet ouvrage s'adresse aux étudiants et aux professionnels ainsi qu'à tous ceux qui ont à mettre en forme et à diffuser des tableaux, des rapports et des graphiques créés avec SAS. Le logiciel SAS se place sans conteste parmi les leaders des marchés de l'informatique décisionnelle, de la Business Intelligence et de l'analyse statistique. Mais il a également la réputation de produire des sorties assez peu esthétiques, et de demander beaucoup d'efforts pour arriver à éditer un graphique diffusable.

SAS expliqué à ma fille et autres documents

<http://od-datamining.com/documentation-en-ligne/>

Kontchou, Kouomegni H. and O. Decourt (2007). Maîtriser SAS Base et SAS Macro SAS 9.2 et versions antérieures, Paris : Dunod

SAS est sans conteste l'un des logiciels d'informatique décisionnelle les plus connus, les plus complets et les plus répandus. Le but de cet ouvrage est de présenter de manière concrète et pratique tout ce qui vous permettra de démarrer efficacement avec SAS et d'apprendre à tirer rapidement le meilleur parti de cet outil puissant. Le contenu de ce livre est adapté à toutes les versions de SAS et à tous les environnements informatiques (Windows, Unix, mainframes?). Les fonctionnalités de SAS 9 et de sa dernière version 9.2 sont décrites avec soin et repérées par rapport à celles des versions antérieures. Les nombreux exemples de programmes présentés sont tous accompagnés de leurs résultats. Les programmes, les tables SAS et les fichiers utilisés sont disponibles sur Internet. Cet ouvrage s'adresse aux professionnels qui utilisent SAS, qu'ils soient informaticiens, chargés d'études statistiques ou analystes. Il sera également utile aux étudiants qui ont à assimiler les fondamentaux de cet outil.

Remond-Tiedrez, E., et al. (1997). Le langage SAS, Paris : Eurostat

(1998). An introduction to SAS/GIS Software : Course Notes, Cary : SAS Institute



Destandau, S. and Le Guen M. (1998). Analyse exploratoire des données avec SAS/INSIGHT, Paris : INSEE

L'analyse exploratoire des données, initiée par John W. Tukey de l'Université de Princeton dans les années 70, repose sur une palette d'outils et de techniques graphiques permettant une nouvelle approche des données. SAS/INSIGHT est un module interactif de SAS qui intègre certains de ces outils et techniques, en se passant d'un langage de commande. Ce document est un ouvrage en français pour l'apprentissage du module SAS/INSIGHT.

Sautory, O. (1995). La statistique descriptive avec le système SAS, Paris : INSEE

Cet ouvrage s'adresse à tous ceux qui font de la statistique descriptive avec le logiciel SAS. Son objectif est double : présenter les diverses procédures de statistique descriptive disponibles dans la version 6.08 du logiciel SAS ; donner à l'utilisateur du logiciel la documentation théorique nécessaire à une bonne maîtrise de ces procédures. La première partie de l'ouvrage présente les éléments de statistique dont la connaissance est requise pour pouvoir comprendre les sorties et analyser les résultats donnés par les procédures de SAS. La deuxième partie contient une présentation détaillée de la syntaxe de chacune des procédures suivantes : CHART, CORR, FREQ, MEANS, SUMMARY, PLOT, PRINT, RANK, STANDARD, UNIVARIATE. De nombreux exemples d'utilisation de ces procédures figurent dans chacune de ces deux parties.

Introduction to SAS. Resources to help you learn and use SAS. UCLA: Statistical Consulting Group.

<http://www.ats.ucla.edu/stat/sas/>

SAS Reference.

[http://www.sasreference.fr/veronique\\_bourcier](http://www.sasreference.fr/veronique_bourcier)

KNOWLEDGE BASE. SAS Institute

<http://support.sas.com/resources/index.html>

Wang, J., et al. (2012). Multilevel models. Applications using SAS, Berlin : De Gruyter

Interest in multilevel statistical models for social science and public health studies has been aroused dramatically since the mids-1980. New multilevel modeling techniques are giving researchers tools for analyzing data that have a hierarchical or clustered structure. Multilevel models are now applied to a wide range of studies in sociology, population studies, education studies, psychology, economics, epidemiology and public health. This book covers a broad range of topics about multilevel modeling. The goal of the authors is to help those who are interested in analysis of multilevel data to understand the basic concepts, theoretical frameworks and application methods of multilevel modelling. The book is focusing on the methods and application of various multilevel models, using the internationally widely used statistical software, SAS (d'après la 4ème de couv.)

Littell, R. C., et al. (2006). SAS for mixed models, Cary : SAS Institute

The indispensable, up-to-date guide to mixed models using SAS?. Discover the latest capabilities available for a variety of applications featuring the MIXED, GLIMMIX, and NLMIXED procedures in this valuable edition of the comprehensive mixed models guide for data analysis, completely revised and updated for SAS?9. The theory underlying the models, the forms of the models for various applications, and a wealth of examples from different fields of study are integrated in the discussions of these models: random effect only and random coefficients models split-plot, multilocation, and repeated measures models hierarchical models with nested random effects analysis of covariance models spatial correlation models generalized linear mixed models nonlinear mixed models Professionals and students with a background in

two-way ANOVA and regression and a basic knowledge of linear models and matrix algebra will benefit from the topics covered. Includes a free CD-ROM with example SAS code.

Allison, P. D. (2000). Logistic regression using the SAS' system : theory and application, Cary : SAS Institute

Stokes, M. E., et al. (1997). Categorical data analysis using the SAS system, Cary : SAS Institute

Cet ouvrage est un manuel de référence à l'usage des statisticiens et chercheurs. Il comprend une partie théorique accompagnée d'exercices d'application.

Walker, G. A. (1997). Common statistical methods for clinical research with SAS examples, Gregy sur Yerres : SAS Institute Cary : SAS Institute

Document pédagogique sur les tests statistiques et leur application sous le logiciel SAS. Une introduction présente les méthodes statistiques de base.

## La statistique avec Stata

(2009). Stata Statistical Software : release 11 (19 volumes), College Station : Stata Corp

Cahuzac, E. and C. Bontemps (2009). Stata par la pratique : statistiques, graphiques et éléments de programmation, Texas : Stata press

Stata par la pratique par Eric Cahuzac et Christophe Bontemps propose une introduction complète à l'usage de Stata en Français. S'appuyant sur des exemples clairs écrits dans un langage simple, cet ouvrage guide l'utilisateur au travers des différentes fonctionnalités de Stata 10. L'ensemble des outils nécessaires à un travail sur données est abordé : exploration des données, statistiques descriptives, modélisation, inférence, tests, graphiques, ainsi que les sorties pour publication. En outre, l'ouvrage inclut également une introduction à la programmation et propose des extraits de code utiles pour résoudre les problèmes fréquemment rencontrés par les utilisateurs. Il contient le matériel essentiel pour transformer le débutant en expert, la clarté de l'ouvrage rendant ce processus particulièrement rapide. L'ouvrage propose un apprentissage de Stata par des approches variées. Pour certains sujets (statistiques descriptives par exemple), les commandes appropriées sont exposées et leur utilisation expliquée de manière simple. Si différents choix sont possibles, les avantages et inconvénients de chaque commande sont explicités afin de guider l'utilisateur dans ses choix. Pour les sujets plus complexes, l'approche est de type exploratoire : le lecteur est accompagné dans l'utilisation des différentes commandes pour analyser ou visualiser les données sur la base d'exemples choisis. Une autre originalité intéressante de ce livre réside dans le fait qu'il ne se limite pas aux commandes standard de Stata, mais présente aussi de nombreuses commandes additionnelles issues de la communauté des utilisateurs de Stata. Les exemples proposés sont principalement tirés de l'économie et des sciences sociales, mais sont illustratifs pour tout lecteur intéressé par une application statistique, quelle que soit sa spécialité. Cet ouvrage couvre un champ très large et s'adresse aux débutants comme aux utilisateurs confirmés, qu'ils soient étudiants ou chercheurs.

Stata 14 documentation. <http://www.stata.com/features/documentation/>

Stata Forum. <http://www.statalist.org/>

Stata FAQ

<http://www.stata.com/support/faqs/>

Introduction to Stata. Resources to help you learn and use Stata. UCLA: Statistical Consulting Group.  
<http://www.ats.ucla.edu/stat/stata/>

The Stata Blog  
<http://blog.stata.com/>

Cameron, A. C. and P. K. Trivedi (2009). Microeconometrics using Stata, Texas : Stata Press

Microeconometrics Using Stata, by A. Colin Cameron and Pravin K. Trivedi, is an outstanding introduction to microeconometrics and how to do microeconomic research using Stata. Aimed at students and researchers, this book covers topics left out of microeconometrics textbooks and omitted from basic introductions to Stata. Cameron and Trivedi provide the most complete and up-to-date survey of microeconomic methods available in Stata. Early in the book, Cameron and Trivedi introduce simulation methods and then use them to illustrate features of the estimators and tests described in the rest of the book. While simulation methods are important tools for econometricians, they are not covered in standard textbooks. By introducing simulation methods, the authors arm students and researchers with techniques they can use in future work. Cameron and Trivedi address each topic with an in-depth Stata example, and they reference their 2005 textbook, *Microeconometrics: Methods and Applications*, where appropriate. The authors also show how to use Stata's programming features to implement methods for which Stata does not have a specific command. Although the book is not specifically about Stata programming, it does show how to solve many programming problems. These techniques are essential in applied microeconometrics because there will always be new, specialized methods beyond what has already been incorporated into a software package. Cameron and Trivedi's choice of topics perfectly reflects the current practice of modern microeconometrics. After introducing the reader to Stata, the authors introduce linear regression, simulation, and generalized least-squares methods. The section on cross-sectional techniques is thorough, with up-to-date treatments of instrumental-variables methods for linear models and of quantile-regression methods. The next section of the book covers estimators for the parameters of linear panel-data models. The authors' choice of topics is unique: after addressing the standard random-effects and fixed-effects methods, the authors also describe mixed linear models' a method used in many areas outside of econometrics. Cameron and Trivedi not only address methods for nonlinear regression models but also show how to code new nonlinear estimators in Stata. In addition to detailing nonlinear methods, which are omitted from most econometrics textbooks, this section shows researchers and students how to easily implement new nonlinear estimators. The authors next describe inference using analytical and bootstrap approximations to the distribution of test statistics. This section highlights Stata's power to easily obtain bootstrap approximations, and it also introduces the basic elements of statistical inference. Cameron and Trivedi then include an extensive section about methods for different nonlinear models. They begin by detailing methods for binary dependent variables. This section is followed by sections about multinomial models, tobit and selection models, count-data models, and nonlinear panel-data models. Two appendices about Stata programming complete the book.

Cleves, M., et al. (2008). An introduction to survival analysis using Stata, Texas : Stata Press

This book is the ideal tutorial for professional data analysts who want to learn survival analysis for the first time or who are well versed in survival analysis but not as dexterous in using Stata to analyze survival data. This text also serves as a valuable reference to those who already have experience using Stata's survival analysis routines. The second edition has been updated for Stata 10, containing a new chapter on power and sample-size calculations for survival studies and sections that describe how to fit regression models (stcox and streg) in the presence of complex survey data. Other enhancements include discussions about nonparametric estimation of mean/median survival, survival graphs with embedded at-risk tables, better hazard graphs through the use of boundary kernels, and concordance measures for assessing the predictive accuracy of the Cox model, as well as an expanded discussion of model building strategies including the use of fractional polynomials. Survival analysis is a field of its own requiring specialized data management and analysis procedures. Toward this end, Stata provides the st family of commands for

organizing and summarizing survival data. The authors of this text are also the authors of Stata's `st` commands. This book provides statistical theory, step-by-step procedures for analyzing survival data, an in-depth usage guide for Stata's most widely used `st` commands, and a collection of tips for using Stata to analyze survival data and present the results. This book develops from first principles the statistical concepts unique to survival data and assumes only a knowledge of basic probability and statistics and a working knowledge of Stata. The first three chapters of the text cover basic theoretical concepts: hazard functions, cumulative hazard functions, and their interpretations; survivor functions; hazard models; and a comparison of nonparametric, semiparametric, and parametric methodologies. Chapter 4 deals with censoring and truncation. The next three chapters cover the formatting, manipulation, `stsetting`, and error checking involved in preparing survival data for analysis using Stata's `st` analysis commands. Chapter 8 covers nonparametric methods, including the Kaplan-Meier and Nelson-Aalen estimators, and the various nonparametric tests for the equality of survival experience. Chapters 9-11 discuss Cox regression and include various examples of fitting a Cox model, obtaining predictions, interpreting results, building models, and model diagnostics. The next four chapters cover parametric models, which are fit using Stata's `streg` command. These chapters include detailed derivations of all six parametric models currently supported in Stata and methods for determining which model is appropriate, as well as information on obtaining predictions, stratification, and advanced topics such as frailty models. The final chapter is devoted to power and sample-size calculations for survival studies.

Hesketh, S. and A. Skrondal (2008). Multilevel and longitudinal modeling using Stata, Texas : Stata Press

Multilevel and Longitudinal Modeling Using Stata, by Sophia Rabe-Hesketh and Anders Skrondal, looks specifically at Stata's treatment of generalized linear mixed models, also known as multilevel or hierarchical models. These models are "mixed" because they allow fixed and random effects, and they are "generalized" because they are appropriate for continuous Gaussian responses as well as binary, count, and other types of limited dependent variables. Beginning with the comparatively simple random-intercept linear model without covariates, Rabe-Hesketh and Skrondal develop the mixed model from principles, thereby familiarizing the reader with terminology, summarizing and relating the widely used estimating strategies, and providing historical perspective. Once the authors have established the mixed-model foundation, they smoothly generalize to random-intercept models with covariates and then to random-coefficient models. The middle chapters of the book apply the concepts for Gaussian models to models for binary responses (e.g., logit and probit), ordinal responses (e.g., ordered logit and ordered probit), and count responses (e.g., Poisson). The authors then consider models with multiple levels of random variation and models with crossed (nonnested) random effects. In its examples and end-of-chapter exercises, the book contains real datasets and data from the medical, social, and behavioral sciences literature. The book has several applications of generalized mixed models performed in Stata. Rabe-Hesketh and Skrondal developed `gllamm`, a Stata program that can fit many latent-variable models, of which the generalized linear mixed model is a special case. With the release of version 9, Stata introduced the `xtmixed` command for fitting linear (Gaussian) mixed models. Stata users can use `gllamm` and `xtmixed` in conjunction with the rest of the `xt` suite of commands to perform comparative mixed-model analyses for various response families. The type of models fit by these commands sometimes overlap; when this happens, the authors highlight the differences in syntax, data organization, and output for the two (or more) commands that can be used to fit the same model. The authors also point out the relative strengths and weaknesses of each command when used to fit the same model, based on considerations such as computational speed, accuracy, and available predictions and postestimation statistics. The book delineates the relationship between `gllamm` and `xtmixed`, clearly showing how they complement one another.

Blossfeld, H. P., et al. (2007). Event history analysis with Stata, New York : Lawrence Erlbaum Associates

This book presents survival analysis from a social science perspective. Introducing the mathematics and statistics of survival analysis, along with substantive discussions of social science data-specific issues, the authors give examples throughout using Stata (version 9) and data from the German Life History Study.

The text covers both basic and advanced topics, from an introduction to life tables to fitting parametric models with unobserved heterogeneity. The authors aptly illustrate the entire research path required in applying event history analysis, from the initial problems of recording event-oriented data, to data organization, to software applications, to interpreting results. Chapters 1 and 2 introduce event history data, discussed substantively, and the data structures used to contain them. Chapter 3 introduces nonparametric descriptive methods including life tables, product-limit estimation of the survivor function, and comparison of survivor functions. Chapters 4-8 focus on estimation using parametric survival functions. This section discusses not the usual exponential, Weibull, etc., models but rather issues such as period-specific effects, qualitative and quantitative covariates, time-dependent covariates, and multiepisode data. Chapter 9 discusses the Cox proportional hazards model, whereas chapter 10 covers problems with parametric model specification, including unobserved heterogeneity. The book has a parametric model focus, which for some readers will be a strength and for others, a weakness. For the latter group, the weakness is minimal because the coverage of the Cox model is adequate given the foregoing discussion. Event History Analysis with Stata is aimed at the professional social scientist but could also serve as a graduate-level text. A web site providing supporting materials for the book, including the dataset files and do-files, is available at <http://web.uni-bamberg.de/sowi/soziologie-i/eha/stata>.

Gould, W., et al. (2006). Maximum likelihood estimation with Stata, Texas : Stata Press

Maximum Likelihood Estimation with Stata is written for researchers in all disciplines who need to fit models using maximum likelihood estimation. This edition offers a wealth of material about the ml command, updated to include new features introduced in Stata 9.

Scott, Long J. and J. Freese (2006). Regression models for categorical dependent variables using Stata, Texas : Stata press

This book shows how to use Stata to fit and interpret regression models for categorical data. Nearly 50% longer than the previous edition, the second edition covers new topics for fitting and interpreting models included in Stata 9, such as multinomial probit models, the stereotype logistic model, and zero-truncated count models. Many of the interpretation techniques have been updated to include interval and point estimates. Although regression models for categorical dependent variables are common, few texts explain how to interpret such models; Regression Models for Categorical Dependent Variables Using Stata, Second Edition fills this void. To accompany the book, Long and Freese provide a suite of commands for hypothesis testing and model diagnostics. The second edition begins with an excellent introduction to Stata and follows with general treatments of estimation, testing, fit, and interpretation in this class of models. Long and Freese detail binary, ordinal, nominal, and count outcomes in separate chapters. The final chapter explains how to fit and interpret models with special characteristics, such as interaction, nonlinear terms, and ordinal and nominal independent variables. One appendix explains the syntax of the author-written commands, and a second appendix details the book's datasets. Long and Freese use many concrete examples in their second edition. All the examples, datasets, and author-written commands are available on the authors' website, so readers can easily replicate the examples when using Stata. This book is ideal for students or applied researchers who want to learn how to fit and interpret models for categorical data (4e de couverture).

## La statistique avec R

The R Project for Statistical Computing. Documentation.  
<https://www.r-project.org/>

The Comprehensive R Archive Network. Documentation.  
<https://cran.r-project.org/>

Irdes - Pôle documentation - Marie-Odile Safon – Relecture : Anissa Afrite  
[www.irdes.fr/documentation/syntheses-et-dossiers-bibliographiques.html](http://www.irdes.fr/documentation/syntheses-et-dossiers-bibliographiques.html)  
[www.irdes.fr/documentation/syntheses/statistiques-et-econometrie.pdf](http://www.irdes.fr/documentation/syntheses/statistiques-et-econometrie.pdf)  
[www.irdes.fr/documentation/syntheses/statistiques-et-econometrie.epub](http://www.irdes.fr/documentation/syntheses/statistiques-et-econometrie.epub)

W.N. Venables & D. M. Smith, [An Introduction to R](http://cran.r-project.org/doc/manuals/R-intro.pdf)  
<http://cran.r-project.org/doc/manuals/R-intro.pdf>  
<http://sites.google.com/site/r4statistics/books/free-version>

Husson, F., et al. (2009). Analyse de données avec R, Rennes : Presses Universitaires de Rennes

Avec ce manuel, le lecteur dispose d'un équipement complet (bases théoriques, exemples, logiciels) pour analyser des données multidimensionnelles. Pour chaque méthode, un exemple détaillé concrétise les éléments théoriques et chaque résultat est accompagné de la commande R qui permet de l'obtenir (logiciel FactoMineR).

Jérôme Pagès (2014). Multiple Factor Analysis by Example Using R