

Bibliographie succincte

1. Ouvrages introductifs

ERICKSON, B.H., NOSANCHUK, T.A.: *Understanding Data: An introduction to exploratory and confirmatory data analysis for students in the Social Sciences*. Milton Keynes, Open University Press, 1977, 2e édition 1992

Un ouvrage introductif à la statistique exploratoire et confirmatoire, d'une qualité pédagogique exceptionnelle (pourquoi de tels ouvrages n'existent pas en français!?)

HARTWIG, Frederick, DEARING, Brian E., *Exploratory Data Analysis*, Sage University Papers (Quantitative Applications in the Social Sciences, No 16), Beverly Hills 1979

Une bonne première introduction à l'exploration, accessible; décrit les principales techniques de base (une centaine de pages).

LEINHARDT, S. et WASSERMANN, S.S.: *Exploratory Data Analysis: An introduction to selected methods*; in: K.F. SCHUESSLER (ed.) *Sociological Methodology*. San Francisco 1979, Jossey-Bass

Une introduction générale aux outils de base de l'exploration (style "review article").

MARSH, Catherine, *Exploring Data. An Introduction to Data Analysis for Social Scientists*, Cambridge, Polity Press, 1988

Un excellent ouvrage introductif qui traite les outils de base de l'exploration. Accessible et didactique. Beaucoup d'exemples tirés des sciences sociales.

VELLEMAN, Paul F. & Hoaglin, David C., *ABC of EDA: Applications, Basics and Computing of Exploratory Data Analysis*, Boston, Mass. 1981 (Duxbury Press) *Principles of Programming*, New York 1980 (Wiley & Sons)

Excellent ouvrage introductif, très clair, très précis. Contient des programmes en Fortran et en Basic. Ces programmes sont également intégrés dans le logiciel Minitab.

2. Textes classiques

[UREDA] HOAGLIN, David C., MOSTELLER, F., TUKEY, J.W., (éditeurs) *Understanding Robust and Exploratory Data Analysis*, New York 1983 (Wiley)

L'original de Tukey (EDA) a souvent été critiqué dans les milieux de la statistique classique pour son vocabulaire qualifié d'ésotérique. Cet ouvrage explicite les liens entre la statistique plus "classique" et les techniques exploratoires et robustes, tout en les mettant dans leur contexte.

[EDT] HOAGLIN, David C., MOSTELLER, F., TUKEY, J.W., (éditeurs) *Exploring Data Tables, Trends and Shapes*, New York, Wiley, 1985

La suite de l'UREDA, collection d'articles.

HOAGLIN, David C., MOSTELLER, F., TUKEY, J.W., *Fundamentals of Exploratory Analysis of Variance*, New York 1992 (Wiley)

Le titre ne le dit qu'en partie, c'est un départ vers de nouveaux horizons.

[DAR] MOSTELLER, F., TUKEY, J.W., *Data Analysis and Regression: A Second Course in Statistics*, Reading, MA., 1977 (Addison-Wesley)

Ouvrage impressionnant, passant en revue dans une perspective exploratoire les principaux outils et l'histoire de la statistique, livre critique et très riche. Le sous-titre "deuxième cours en statistique" invite très clairement à une autre (deuxième) façon d'aborder la statistique, une réorientation statistique... après la statistique classique. Ce livre est comme EDA de Tukey - un classique.

TUKEY, John W.: The Future of Data Analysis, *Ann. Math. Statist* 33, 1962

Un article visionnaire qui n'a pas pris une ride....

- [EDA] *Exploratory Data Analysis (EDA)*, Reading, MA, 1977 (Addison- Wesley) [EDA]

Le classique. Circulait pendant longtemps sous forme de 3 volumes de photocopies. L'ouvrage est une collection de techniques très variées, des propositions d'analyse. Un livre à lire et à relire; certaines parties sont très accessibles, d'autres se découvrent après la 5e, 10e... lecture.

- We need both exploratory and confirmatory, in: *The American Statistician*, 34, 23-25
- *The Practice of Data Analysis: Essays in Honor of John W. Tukey*, Princeton UP, Princeton, 1998

Une collection d'articles pour fêter les 80 ans de Tukey... en général excellent et quelques perles (les article de Velleman et Huber); contient également une biographie détaillée de Tukey.

The Collected Works of John W. Tukey, Pacific Grove, Wadsworth & Brooks/Cole, depuis le volume 8 repris par les éditions Chapman & Hall, New York et Londres.

Vol. 1: Time Series 1949-1964 (ed. D. R. Brillinger), 1984

Vol. 2: Time Series 1965-1984 (ed. D. R. Brillinger), 1984

Vol. 3: Philosophy and Principles of Data Analysis 1949-1964 (ed. L.V. Jones), 1986

Vol. 4: Philosophy and Principles of Data Analysis 1965-1986 (ed. L.V. Jones), 1986

Vol. 5: Graphics 1965-1985 (ed. W.S. Cleveland), 1988

Vol. 6: More Mathematical 1938-1984 (ed C.L. Mallows), 1990

Vol. 7: Factorial and ANOVA 1949-1962 (ed. D.R. Cox), 1992

Vol. 8: Multiple Comparisons 1949-1983, Chapman and Hall 1994

Chaque volume contient une série d'articles, communications aux Congrès scientifiques, papiers internes qui autrement seraient inaccessibles ou très difficiles à trouver. D'autres volumes sont en voie de publication.

3. Graphique et visualisation

BASFORD, K.E.; TUKEY, J.W., *Graphical Analysis of Multiresponse Data. Illustrated with a Plant Breeding Trial*, Chapman&Hall/CRC, Boca Raton, 1999

BERTIN, Jacques, *Sémiologie graphique*, Mouton/Gauthier-Villars, Paris/La Haye, 1967; EHESS, Les Réimpressions, 1999

Un classique bien moins connu dans le monde francophone qu'ailleurs.

CARD, S.K, MACKINLAY, J.D., SHNEIDERMAN, B. (eds), *Readings in Information Visualization. Using Vision to Think*, Morgan Kaufmann, Academic Press, 1999

Bon aperçu de l'état de l'art dans du domaine de visualisation (point de vue général, non-statistique).

[GMDA] CHAMBERS, J.M., CLEVELAND, W.S., KLEINER, B., TUKEY, P.A.: *Graphical Methods for Data Analysis*, Duxbury Press, Boston 1983

Le titre dit tout; les auteurs gravitent autour de Tukey (John) et du langage S (langage de programmation graphique et statistique). Un livre clé qui introduit tous les thèmes importants de la statistique moderne.

CLEVELAND, W.S., *The Elements of Graphing Data*, Monterey, Wadsworth, 1985, 1994 (2e édition revue et augmentée)

- *Visualizing Data*, Murray Hill, AT&T Bell Laboratories/Hobart Press, 1993

Un livre essentiel sur la visualisation statistique par un des auteurs du logiciel S

CLEVELAND, W.S.; MCGILL, M.E. (ed.), *Dynamic Graphics for Statistics*, New York, Chapman & Hall, 1988

COOK, D. R.; WEISBERG, S.: *An Introduction to Regression Graphics*, J. Wiley & Sons, New York, 1994;

- *Applied Regression Including Computing and Graphics*, John Wiley & Sons 1999

La régression comme vous ne la connaissez pas (encore): l'étude graphique des relations. Utilise LispStat avec une extension appelée ARC (1999), anciennement R-Code ou R-Code2 (1994)

SPENCE, I.; LEWANDOWSKY, *Graphical Perception*, in: Fox/Long (eds), *Modern Methods of Data Analysis*, Newbury Park: Sage, 1990

Bonne bibliographie sur la perception graphique.

SPENCE, Robert, *Information Visualization*, ACM-Press/Addison-Wesley, 2001

Tour d'horizon du domaine de la visualisation (statistique et au-delà).

TUFTE, E., *The Visual Display of Quantitative Information*, Graphics Press, 1986, 2000 (2e édition)[version française disponible fin 2001]

Un "tour de force" (Tukey dixit), un ouvrage plein d'idées sur les graphiques, leurs qualités et défauts. De plus l'ouvrage est beau à voir et n'est pas du tout technique. Toute personne intéressée par les graphiques doit connaître ce livre. Premier ouvrage d'une série:

- *Envisioning information*, Graphics Press, 1990

- *Visual Explanations: Images and Quantities*, Graphics Press, 1996.

The Cognitive Style of PowerPoint, Graphics Press, 2003

WAINER, H., *How to Display Data Badly*, *American Statistician*, May 1984

Un article provocateur contenant une multitude d'exemples.

- *Visual Revelations*, Copernicus/Springer, 1997

Un livre plein d'exemples: passionnant. Le même auteur a tenu pendant des années une rubrique au même titre dans la revue "Chance" (ASA).

- *A Trout in the Milk and Other Tales of Visual Discovery*, Princeton U. Press, 2004

WAINER, H.; VELLEMAN, P.F., *Statistical Graphics: Mapping the Pathways of Science*, *Ann. Rev. Psychol.*, 2001, 32

WILKINSON, Leland, *The Grammar of Graphics*, Springer 1999

La grammaire graphique, titre choisi délibérément pour faire écho à Bertin, dont l'auteur se dit profondément influencé...

6. L'esprit exploratoire

FEYNMAN, Richard P., *Vous voulez rire, Monsieur Feynman*, Editions Odile Jacob, Paris 2000, *traduction de "Surely you're Joking Mr. Feynman! Adventures of a Curious Character"*

A lire pour comprendre comment un des physiciens les plus importants du 20e siècle (prix Nobel) aborde la science, avec humour, rigueur et avec un réel plaisir de chercher, fouiner, explorer...

5. Thèmes choisis

BARNETT, Vic (ed.), *Interpreting Multivariate Data*, Chichester 1981 (Wiley & Sons)

- BERK, Richard A., A Primer on Robust Regression, **in:** Fox/Long (eds), op. cit.
- BLASIUS, J.; GREENACRE, M.L. (eds), Visualisation of Categorical Data, Academic Press 1998
- BRILLINGER, D.R. (et al.), The Practice of Data Analysis. Essays in Honor of John W. Tukey, Princeton University Press, 1997
- DIACONIS, P.; EFRON, B., *Computer Intensive Methods in Statistics* Scientific American 5,83 (Pour la Science 7,83)
- Un bon point de départ pour comprendre les bases de la statistique moderne; introduction au bootstrap, jackknife et autres techniques de ré-échantillonnage*
- EFRON, B., *An Introduction to the bootstrap*, Chapman and Hall 1993
- FOX, J., *Describing Univariate Distributions*, **in:** Fox/Long (eds), *Modern Methods of Data Analysis*, Newbury Park: Sage, 1990
- GOODALL, C., *A Survey of Smoothing Techniques*, **in:** Fox/Long (eds), op. cit.
- FAIRLEY, W.; MOSTELLER, F., *Statistics and Public Policy*, Reading: Addison Wesley, 1977
- HAMPEL, F.R. (et al.), Robust Statistics. The Approach Based on Influence Functions, New York, Wiley, 1986
- HASTIE, T.J.; TIBSHIRANI, R.J., FRIEDMAN, J., *The Elements of Statistical Learning*, Springer, New York, 2001
- HASTIE, T.J.; TIBSHIRANI, R.J., *Generalized Additive Models*, Chapman and Hall, London, 1990
- HORBER, E., Analyse exploratoire des données et sciences sociales. *Vers une approche méthodologique pragmatique*, Université de Genève, 1990
- JOHNSTONE, I.M.; VELLEMAN, P.F., *The Resistant Line and Other Regression Methods*, JASA, 80, 1041-1059, 1985
- LADIRAY, Dominique; ROTH, Nicole, *Lissage robuste de séries chronologiques. Une étude expérimentale*, Annales d'Economie et de Statistique, No 5, 1987
- MACDONALD, K.I., *Exploratory Data Analysis: A Process and a Problem*. **in:** McKay, D. et al., *Data Analysis and the Social Sciences*, London: Frances Pinter, 1983
- MILLER, R.G., The Jackknife: A Review, Biometrika, vol. 61, 1976 pp 1-15
- RONCCHETTI, E. et al., STEP I: Statistique et Probabilités: Une Introduction, Presses polytechniques, Lausanne, 1991
- Le chapitre 2 est une introduction à l'analyse exploratoire des données (STEP I est un didacticiel qui introduit aux principes de la statistique et des probabilités).*
- ROUSSEEUW, P.J., RUTS, IDA, TUKEY. J.W., The Bagplot: A Bivariate Boxplot, The American Statistician, 1999, Vol 53,4
- STINE, R., *An Introduction to Bootstrap Methods: Examples and Ideas*, **in:** Fox/Long (eds), op. cit.
- STOTO, Michael A. and EMERSON John D.: *Power Transformations for Data Analysis*, **in:** Leinhardt, S. (ed.) *Sociological Methodology 1983/1983*, San Francisco: Jossey-Bass 1983
- TUKEY, J.W., "Robust Techniques for the User", **in:** LAUNER, R.L., & WILKINSON, G.N. (eds) *Robustness in Statistics*, New York 1979 (Academic Press).
- VELLEMAN, Paul F.: Applied nonlinear smoothing, **in:** Sociological Methodology 1982, **in:** J. Leinhardt (ed.) San Francisco, Jossey Bass, 1982, pp. 141-177

WILKINSON, L. and Task Force on Statistical Inference, "Statistical Methods in Psychology Journals: Guidelines and Explanations". *American Psychologist*, 1999, vol. 54, No. 8: 594-604

6. Logiciels

Informations supplémentaires à <http://www.unige.ch/ses/sococ/eda/explor>

BECKER, R.A.; CHAMBERS, John M.; WILKS, A.R.,. *The New S Language. A Programming Environment for Data Analysis and Graphics*, Wadsworth & Brooks/Cole, Pacific Grove, 1988

DESTANDAU, Sophie; LE GUEN Monique, *Analyse exploratoire des Données avec SAS/INSIGHT*, INSEE 1998

HORBER, E.; *EDA User's Manual*, Université de Genève, 1991

JMP In, *Statistical Visualization Software for the Macintosh*, Duxbury Press, Belmont Ca, 1993
(Contient la version "étudiant" du logiciel)

STINE, Robert; Fox, John (eds) *Statistical Computing Environments for Social Research*, Sage, Thousand Oaks, 1997o

Ceci n'est pas un ouvrage qui traite spécifiquement des logiciels exploratoires, mais contient des articles sur plusieurs logiciels exploratoires, notamment LISPStat et S-Plus, ainsi que trois articles sur les extensions de LISP-Stat (Vista, R-code et Axis).

TIERNEY, Luke, *LISP-STAT: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics*, NY, Wiley, 1990

VELLEMAN, Paul F., *Learning Data Analysis with DATA DESK*, New York: W.H. Freeman 1989
(contient la version "étudiant" du logiciel)

7. Autres langues....

FREIXA I BLANXART, M. (et al.) *Análisis exploratorio de Datos: Nuevas Técnicas estadísticas*, Barcelona PPU 1992

POLASEK, Wolfgang, *EDA Explorative Datenanalyse*, Springer, Berlin, 1988/1994 (2e édition).