

Lecture Notes in Statistics

109

Edited by P. Bickel, P. Diggle, S. Fienberg, K. Krickeberg,
I. Olkin, N. Wermuth, S. Zeger

Springer

New York

Berlin

Heidelberg

Barcelona

Budapest

Hong Kong

London

Milan

Paris

Santa Clara

Singapore

Tokyo

Helmut Rieder
(Editor)

Robust Statistics, Data Analysis, and Computer Intensive Methods

In Honor of Peter Huber's 60th Birthday



Springer

Helmut Rieder
Universität Bayreuth
Lehrstuhl Mathematik VII
D-95440 Bayreuth
Universitätstrasse 30
Gebäude NM II
Germany

Library of Congress Cataloging-in-Publication Data Available
Printed on acid-free paper.

© 1996 Springer-Verlag New York, Inc.

Softcover reprint of the hardcover 1st edition 1996

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Camera ready copy provided by the author.

9 8 7 6 5 4 3 2 1

ISBN-13: 978-0-387-94660-3 e-ISBN-13: 978-1-4612-2380-1

DOI: 10.1007/ 978-1-4612-2380-1

Table of Contents

PETER J. HUBER	vii
Preface	xiii
<i>Helmut Rieder</i>	
Bootstrap Variable-Selection and Confidence Sets	1
<i>Rudolf Beran</i>	
Robust Estimation in the Logistic Regression Model	17
<i>Ana M. Bianco and Victor J. Yohai</i>	
The m out of n Bootstrap and Goodness of Fit Tests with Double Censored Data	35
<i>Peter J. Bickel and Jian-Jian Ren</i>	
What Criterion for a Power Algorithm?	49
<i>Andreas Buja</i>	
Nonparametric Estimation of Global Functionals of Conditional Quantiles	63
<i>Probal Chaudhuri, Kjell Doksum, and Alexander Samarov</i>	
High Breakdown Point Estimators in Logistic Regression	79
<i>Andreas Christmann</i>	
The Guinea Pig of Multiple Regression	91
<i>Yadolah Dodge</i>	
When is a p -Value a Good Measure of Evidence?	119
<i>Michael B. Dollinger, Elena Kulinskaya, and Robert G. Staudte</i>	
Geometric Properties of Principal Curves in the Plane	135
<i>Tom Duchamp and Werner Stuetzle</i>	
On Robust Estimation of Variograms in Geostatistics	153
<i>Rudolf Dutter</i>	
Advances in Nonparametric Function Estimation	173
<i>Theo Gasser</i>	
On the Philosophical Foundations of Statistics: Bridges to Huber's work, and recent results	185
<i>Frank Hampel</i>	

Data Based Prototyping	197
<i>Thomas M. Huber and Matthias Nagel</i>	
Robust Regression with a Categorical Covariable	215
<i>Mia Hubert and Peter J. Rousseeuw</i>	
Robustness in Discriminant Analysis	225
<i>Yurij S. Kharin</i>	
M -Estimation and Spatial Quantiles	235
<i>Vladimir Koltchinskii</i>	
Robust Statistical Methods in Interlaboratory Analytical Studies	251
<i>Peter Lischer</i>	
Estimating Distributions with a Fixed Number of Modes	267
<i>Martin B. Mächler</i>	
Implementing M -Estimators of the Gamma distribution	277
<i>Alfio Marazzi and C. Ruffieux</i>	
Constrained M -Estimation for Regression	299
<i>Beatriz Mendes and David E. Tyler</i>	
Inference for the Direction of the Larger of Two Eigenvectors: The Case of Circular Elongation	321
<i>Stephan Morgenthaler and John W. Tukey</i>	
High Breakdown Point Designs	353
<i>Christine H. Müller</i>	
On Bayesian Robustness: An Asymptotic Approach	361
<i>Daniel Peña and Ruben H. Zamar</i>	
Computational Aspects of Trimmed Single-Link Clustering	375
<i>Evangelos Tabakis</i>	
Interval Probability on Finite Sample Spaces	391
<i>Kurt Weichselberger</i>	
On Consistency of Recursive Multivariate M -Estimators in Linear Models	411
<i>Yuehua Wu</i>	
List of Participants	425

Peter J. Huber

born in Wohlen/Aargau, Switzerland

25 March 1934

Education

Ph. D. ETH Zürich (Mathematics)	1961
Diploma ETH Zürich (Mathematics)	1958

Teaching and Research Positions

Universität Bayreuth	
Lehrstuhl für Mathematische Statistik	1992–present
Massachusetts Institute of Technology	
Full Professor (Applied Mathematics)	1988–1992
Harvard University	
Full Professor (Statistics)	1978–1988
Eidg. Technische Hochschule Zürich	
Full Professor (Mathematical Statistics)	1964–1978
Harvard University	
Visiting Professor	1977–1978
Yale University	
Visiting Professor	1971–1972
Princeton University	
Visiting Professor	1970–1971
Cornell University	
Visiting Professor	1966–1967
Visiting Professor	1963–1964
University of California, Berkeley	
MSRI, Visiting Professor	1982
Miller Fellow	1962–1963
Research Fellowship (Schweiz. Nationalfonds)	1961–1962

Memberships

International Statistical Institute
Institute of Mathematical Statistics (Fellow)
American Mathematical Society
Association for Computing Machinery
Bernoulli Society
American Association for the Advancement of Science (Fellow)
National Computer Graphics Association
American Academy of Arts and Sciences (Fellow)

Professional Activities

Board Member, Institute for Mathematics and its Applications	1984–1987
Council Member, Institute of Mathematical Statistics	1983–1986
Member, Panel on Computers in Statistics Research (ONR “Eddy Panel”)	1986
Member, Panel on Future Directions in Computational Mathematics (NSF, “Rheinboldt Panel”)	1985
Associate Editor—Zeitschrift für Wahrscheinlichkeitstheorie	1972–1987
Associate Editor—Annals of Statistics	
Chairman of several conferences (IMS/AMS Summer School, Oberwolfach, DMV–Seminar, . . .)	

Honors

Dr. h. c. , Université de Neuchâtel	1994
Humboldt Award	1988
Wald Lecturer, IMS Annual Meeting	1972
NSF Senior Foreign Scientist Fellowship, Princeton	1970–1971
Silver Medal of the Eidgenössische Technische Hochschule	1961

List of Publications

- [1] Zu einem Mathematischen Keilschrifttext (VAT 8512). *Isis* **46** (1955) 104–106.
- [2] Bemerkungen über Mathematische Keilschrifttexte. *Enseignement mathématique* **3** (1957) 19–27.
- [3] Zur täglichen Bewegung des Jupiter nach Babylonischen Texten. *Z. Assyriol. N. F.* **18** (1957) 265–303.
- [4] Über den Nullpunkt der Babylonischen Ekliptik. *Centaurus* **5** (1958) 192–208.
- [5] Homotopical cohomology and Čech cohomology. *Math. Ann.* **144** (1961) 73–76.
- [6] Homotopy theory in general categories. *Math. Ann.* **144** (1961) 361–385.
- [7] Standard constructions in abelian categories. *Math. Ann.* **146** (1962) 321–325.
- [8] A remark on a paper of Trawinski and David. *Ann. Math. Stat.* **34** (1963) 92–94.
- [9] (with H. Bühlmann). Pairwise comparison and ranking in tournaments. *Ann. Math. Stat.* **34** (1963) 501–510.
- [10] Pairwise comparison and ranking: Optimum properties of the row sum procedure. *Ann. Math. Stat.* **34** (1963) 511–520.
- [11] Robust estimation of a location parameter. *Ann. Math. Stat.* **35** (1964) 73–101.
- [12] A robust version of the probability ratio test. *Ann. Math. Stat.* **36** (1965) 1753–1758.
- [13] Neuere Fortschritte in der Theorie der Bayesschen Statistik. *Mitteil. d. Vereinig. schweiz. Versicherungsmath.* **65** (1965) 163–171.
- [14] Strict efficiency excludes superefficiency (abstract). *Ann. Math. Stat.* **37** (1966) 1425.
- [15] The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. 5th Berkeley Symp. Math. Stat. Prob.*, Vol. I (1967) 221–233.
- [16] Robust confidence limits. *Z. Wahrsch. verw. Gebiete* **10** (1968) 269–278.
- [17] *Théorie de l'Inférence Statistique Robuste*. Les Presses de l' Université de Montréal (1969).
- [18] Robust estimation. *Mathematical Centre Tracts* **27** (1968) 1–25, Amsterdam.
- [19] Studentizing robust estimates. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.), (1970) 453–463, Bloomington.
- [20] (with B. Kleiner, H. Fluhler, G. Dumermuth). Spectrum analysis of the electroencephalogram. *Computer Programs in Biomedicine*. Amsterdam, Vol. I (1970) 183–197.
- [21] (with B. Kleiner, T. Gasser, G. Dumermuth). Statistical methods for investigating phase relations in stationary stochastic processes. *IEEE Transactions on Audio and Electroacoustics* Vol. AU-19 (1971) 78–86.
- [22] Robust statistics: A review (The A. Wald lecture). *Ann. Math. Stat.* **43** (1972) 1041–1067.
- [23] (with Andrews, Bickel, Hampel, Rogers, Tukey). *Robust Estimates of Location: Survey and Advances*. Princeton University Press, Princeton, New Jersey (1972).
- [24] (with V. Strassen). Minimax tests and the Neyman–Pearson lemma for capacities. *Ann. Statist.* **1** (1973) 251–263, **2** (1974) 223–224.

- [25] Robust regression: Asymptotics, conjectures and Monte Carlo (Wald memorial lecture). *Ann. Statist.* **1** (1973) 799–821.
- [26] The use of choquet capacities in statistics (1973). *Bull. I.S.I.* Vol. XLV (1973) 181–188.
- [27] Fisher information and spline interpolation. *Ann. Statist.* **2** (1974) 1029–1033.
- [28] (with R. Dutter). Numerical solution of robust regression problems. COMPSTAT 1974, *Proc. in Computational Statistics* (G. Bruckman, ed.), Physica-Verlag, Vienna (1974) 165–172.
- [29] Some mathematical problems arising in robust statistics. *Proc. Int. Congress of Mathematicians*, Vancouver (1974) 821–824.
- [30] *Einführung in die Wahrscheinlichkeit und Statistik* (nach einer Vorlesung). Verlag der Fachvereine an der ETH Zürich (1975).
- [31] Robustness and designs. In *A Survey of Statistical Design and Linear Models* (J. N. Srivastava, ed.), North-Holland Publishing Company (1975) 287–301.
- [32] Application vs. abstraction: The selling out of mathematical statistics? In *Proc. Conf. Directions for Mathematical Statistics* (S. G. Ghurye, ed.), Applied Probability Trust (1975) 84–89.
- [33] Kapazitäten statt Wahrscheinlichkeiten? Gedanken zur Grundlegung der Statistik. *Jber. Deutsch. Math.-Verein.* **78** H.2 (1976) 81–92.
- [34] Robust methods of estimation of regression coefficients. *Math. Operationsforsch. Statist. Ser. Statistics* **8** No. 1 (1977) 41–53.
- [35] Early cuneiform evidence for the planet Venus. In *Scientists Confront Velikovsky* (D. Goldsmith, ed.), Cornell University Press (1977) 117–144.
- [36] Robust covariances. In *Statistical Decision Theory and Related Topics II* (S. S. Gupta and D. S. Moore, eds.), Academic Press (1977) 165–191.
- [37] *Robust Statistical Procedures*. Society for Industrial and Applied Mathematics, Philadelphia (1977).
- [38] (with A. Aaboe). A text concerning subdivision of the synodic motion of Venus from Babylon: BM 37151. In *Ancient Near Eastern Studies in Memory of J. J. Finkelstein*, Connecticut Acad. Arts & Sciences, Mem. **19** (1977) 1–4.
- [39] (with Gasser, Largo, Prader and Stützel). Analysis of the adolescent growth spurt using smoothing spline functions. *Ann. Human Biology* **5** (1978) 421–434.
- [40] Robust smoothing. In *Robustness in Statistics* (R. L. Launer and G. N. Wilkinson, eds.), Academic Press (1979) 33–47.
- [41] Robustness, computing and data analysis. In *Interactive Statistics* (D. McNeil, ed.), North-Holland (1979) 85–90.
- [42] (with Gasser, Largo, Molinari, Prader and Stützel). Shape-invariant modelling. *Ann. Human Biology* **7** (1980) 507–528.
- [43] *Robust Statistics*. Wiley, New York (1981), 308 pp.
- [44] (with R. Dutter). Numerical methods for the nonlinear robust regression problem. *J. Statist. Comput. Simul.* (1981) 79–113.
- [45] (with D. Donoho and H. M. Thoma). The use of kinematics displays to represent high dimensional data. In *Computer Science and Statistics: Proc. 13th Symp. Interface* (W. F. Eddy, ed.), Springer-Verlag (1981) 274–278.

- [46] *Astronomical Dating of Babylon II and UR III*. Undena Press, Malibu, CA (1982).
- [47] (with D. Donoho). The notion of breakdown point. In *Festschrift in Honor of Erich Lehmann* (Bickel, Doksum and Hodges, eds.), Wadsworth (1982) 157–184.
- [48] (with D. Donoho, E. Ramos and H. M. Thoma). Kinematic display of multivariate data. *Proc. 3rd Ann. Conf. NCGA* (1982) 393–400.
- [49] Discussion of Burguete et al., “On unification of the asymptotic theory of nonlinear economic models”. *Econometric Reviews* **1** (1982) 191–192.
- [50] Current issues in robust statistics. In *Proc. Symp. Recent Advances in Statistics*. Academic Press, New York (1983) 183–196.
- [51] Statistical graphics: History and overview. *Proc. 4th Ann. Conf. NCGA* (1983) 667–676.
- [52] Minimax aspects of bounded influence regression (with discussions). *J. Amer. Statist. Assoc.* **78** (1983) 66–80.
- [53] Kinematic displays. In *Encyclopedia of Statistical Sciences* (Kotz and Johnson, eds.), **4**, John Wiley (1983) 381–384.
- [54] Finite sample breakdown of M- and P-estimators. *Ann. Statist.* **12** (1984) 119–126.
- [55] Data analysis: In search of an identity. *Proc. Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer*, Vol I (Lucien M. LeCam and Richard A. Olshen, eds.), Wadsworth (1985) 65–78.
- [56] Projection pursuit. *Ann. Statist.* **13** (1985) 435–475. Invited Discussion Paper.
- [57] Environments for supporting statistical strategy. In *Artificial Intelligence* (W. A. Gale, ed.), Addison Wesley, Reading, MA (1986) 285–294.
- [58] Data analysis implications for command language design. In *Foundations for Human-Computer Communication* (R. Hopper and I. A. Newman, eds.), Elsevier, North-Holland (1986) 488–504.
- [59] Discussion of R. A. Thisted, “Computing environments for data analysis”. *Statistical Science* **1** (1986) 273.
- [60] Dating by lunar eclipse omina. In *From Ancient Omens to Statistical Mechanics* (J. L. Berggren and B. R. Goldstein, eds.), Copenhagen (1987) 3–13.
- [61] Experiences with 3-d scatter plots. *J. Amer. Statist. Assoc.* **82** (1987) 448–453.
- [62] Astronomical evidence for the long and against the middle and short chronologies. In *High, Middle or Low? Acts of an Intl. Coll. on Absolute Chronology, Gothenburg, 20–22. August 1987* (P. Åström, ed.), Studies in Mediterranean Archaeology and Literature, P. Åströms Förlag, Gothenburg, Vol. 56, 5–17.
- [63] The place of the L_1 -norm in robust estimation. *Comput. Statist. and Data Analysis* **5** (1987) 255–262.
- [64] Discussion of Becker et al., “Dynamic graphics for data analysis”. *Statistical Science* **2** (1987) 385–386.
- [65] Discussion of J. Hodges, “Uncertainty, policy analysis and statistics”. *Statistical Science* **2** (1987) 279–281.
- [66] Spurious structure? Or: Improved bounds for the Kolmogorov index. Technical Report CICS-P-77 (1988).
- [67] (with W. Vach). The design of a “Lab assistant” for data analysis. Technical Report (1988).

- [68] (with E. H. Huber-Buser). ISP: Why a Command Language? In *Fortschritte der Statistik-Software I* (F. Faulbaum und H.-M. Uehlinger, eds.), Gustav Fischer Verlag, Stuttgart (1988) 349–360.
- [69] *Projection Pursuit and Exploratory Data Analysis*. Technical Report, M.I.T. Dept. of Mathematics (1990).
- [70] Between robustness and diagnostics. In *Directions in Robust Statistics and Diagnostics* (W. Stahel and S. Weisberg, eds.), IMA Volumes in Mathematics and Its Applications (1991) **33** 121–130.
- [71] Goals versus algorithms. In *Data Analysis and Statistical Inference* (S. Schach and G. Trenkler, eds.), (1992) 567–579.
- [72] Issues in computational data analysis. In *Comput. Statist.* (Y. Dodge and J. Whitaker, eds.), (1992) Vol. 2, 3–13.
- [73] Projection pursuit and robustness. In *New Directions in Statistical Data Analysis and Robustness*. Birkhäuser Verlag, (1993) 139–146.
- [74] Huge data sets. In *Proc. Comput. Statist.* (R. Dutter and W. Grossmann, eds.), (1994) 3–13.
- [75] Languages for statistics and data analysis. In *Comput. Statist.* (P. Dirschedl and R. Ostermann, eds.), (1994), 53–80.
- [76] Zufall oder Unordnung? In *Festschrift für Kurt Weichselberger* (H. Rinne, B. Rürger and H. Strecker, eds.), Physica-Verlag, Vienna (1995) 45–59.
- [77] Robustness: Where are we now? (1995) *Student*, to appear.
- [78] Speculations on the Path of Statistics. Tukey–Festschrift, to appear.
- [79] Massive Data Sets Workshop: The Morning After. To appear.

