

1. Record Nr.	UNINA9910818514303321
Autore	Maronna Ricardo A.
Titolo	Robust statistics : theory and methods (with R) // Ricardo A. Maronna, Universidad Nacional de La Plata, Argentina, R. Douglas Martin, University of Washington, USAe, Victor J. Yohai, University of Buenos Aires, and CONICRT, Argentina, Matias Salibian-Barrera, University of British Columbia, Canada
Pubbl/distr/stampa	Hoboken, New Jersey : , : Wiley, , 2019 [Piscataqay, New Jersey] : , : IEEE Xplore, , [2019]
ISBN	1-119-21467-X 1-119-21466-1 1-119-21465-3
Edizione	[Second edition.]
Descrizione fisica	1 online resource (431 pages)
Collana	Wiley series in probability and statistics THEi Wiley ebooks
Disciplina	519.5
Soggetti	Robust statistics R (Computer program language)
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Note: sections marked with an asterisk can be skipped on first reading -- Preface xv -- Preface to the First Edition xxi -- About the Companion Website xxix -- 1 Introduction 1 -- 1.1 Classical and robust approaches to statistics 1 -- 1.2 Mean and standard deviation 2 -- 1.3 The "three sigma rule" 6 -- 1.4 Linear regression 8 -- 1.4.1 Straight-line regression 8 -- 1.4.2 Multiple linear regression 9 -- 1.5 Correlation coefficients 12 -- 1.6 Other parametric models 13 -- 1.7 Problems 16 -- 2 Location and Scale 17 -- 2.1 The location model 17 -- 2.2 Formalizing departures from normality 19 -- 2.3 M-estimators of location 22 -- 2.3.1 Generalizing maximum likelihood 22 -- 2.3.2 The distribution of M-estimators 25 -- 2.3.3 An intuitive view of M-estimators 28 -- 2.3.4 Redescending M-estimators 29 -- 2.4 Trimmed and Winsorized means 31 -- 2.5 M-estimators of scale 33 -- 2.6 Dispersion estimators 35 -- 2.7 M-estimators of location with unknown dispersion 37 -- 2.7.1 Previous estimation of

dispersion 38 -- 2.7.2 Simultaneous M-estimators of location and dispersion 38 -- 2.8 Numerical computing of M-estimators 40 -- 2.8.1 Location with previously-computed dispersion estimation 40 -- 2.8.2 Scale estimators 41 -- 2.8.3 Simultaneous estimation of location and dispersion 42 -- 2.9 Robust confidence intervals and tests 42 -- 2.9.1 Confidence intervals 42 -- 2.9.2 Tests 44 -- 2.10 Appendix: proofs and complements 45 -- 2.10.1 Mixtures 45 -- 2.10.2 Asymptotic normality of M-estimators 46 -- 2.10.3 Slutsky's lemma 47 -- 2.10.4 Quantiles 47 -- 2.10.5 Alternative algorithms for M-estimators 47 -- 2.11 Recommendations and software 48 -- 2.12 Problems 49 -- 3 Measuring Robustness 51 -- 3.1 The influence function 55 -- 3.1.1 *The convergence of the SC to the IF 57 -- 3.2 The breakdown point 58 -- 3.2.1 Location M-estimators 59 -- 3.2.2 Scale and dispersion estimators 59 -- 3.2.3 Location with previously-computed dispersion estimator 60 -- 3.2.4 Simultaneous estimation 61. 3.2.5 Finite-sample breakdown point 61 -- 3.3 Maximum asymptotic bias 62 -- 3.4 Balancing robustness and efficiency 64 -- 3.5 *'Optimal' robustness 66 -- 3.5.1 Bias- and variance-optimality of location estimators 66 -- 3.5.2 Bias optimality of scale and dispersion estimators 66 -- 3.5.3 The infinitesimal approach 67 -- 3.5.4 The Hampel approach 68 -- 3.5.5 Balancing bias and variance: the general problem 70 -- 3.6 Multidimensional parameters 70 -- 3.7 *Estimators as functionals 72 -- 3.8 Appendix: Proofs of results 76 -- 3.8.1 IF of general M-estimators 76 -- 3.8.2 Maximum BP of location estimators 76 -- 3.8.3 BP of location M-estimators 77 -- 3.8.4 Maximum bias of location M-estimators 79 -- 3.8.5 The minimax bias property of the median 80 -- 3.8.6 Minimizing the GES 80 -- 3.8.7 Hampel optimality 82 -- 3.9 Problems 85 -- 4 Linear Regression 1 87 -- 4.1 Introduction 87 -- 4.2 Review of the least squares method 91 -- 4.3 Classical methods for outlier detection 94 -- 4.4 Regression M-estimators 97 -- 4.4.1 M-estimators with known scale 99 -- 4.4.2 M-estimators with preliminary scale 100 -- 4.4.3 Simultaneous estimation of regression and scale 102 -- 4.5 Numerical computing of monotone M-estimators 103 -- 4.5.1 The L1 estimator 103 -- 4.5.2 M-estimators with smooth ψ -function 104 -- 4.6 BP of monotone regression estimators 104 -- 4.7 Robust tests for linear hypothesis 106 -- 4.7.1 Review of the classical theory 106 -- 4.7.2 Robust tests using M-estimators 108 -- 4.8 *Regression quantiles 109 -- 4.9 Appendix: Proofs and complements 110 -- 4.9.1 Why equivariance? 110 -- 4.9.2 Consistency of estimated slopes under asymmetric errors 110 -- 4.9.3 Maximum FBP of equivariant estimators 111 -- 4.9.4 The FBP of monotone M-estimators 112 -- 4.10 Recommendations and software 113 -- 4.11 Problems 113 -- 5 Linear Regression 2 115 -- 5.1 Introduction 115 -- 5.2 The linear model with random predictors 118 -- 5.3 M-estimators with a bounded ψ -function 119. 5.3.1 Properties of M-estimators with a bounded ψ -function 120 -- 5.4 Estimators based on a robust residual scale 124 -- 5.4.1 S-estimators 124 -- 5.4.2 L-estimators of scale and the LTS estimator 126 -- 5.4.3 σ -estimators 127 -- 5.5 MM-estimators 128 -- 5.6 Robust inference and variable selection for M-estimators 133 -- 5.6.1 Bootstrap robust confidence intervals and tests 134 -- 5.6.2 Variable selection 135 -- 5.7 Algorithms 138 -- 5.7.1 Finding local minima 140 -- 5.7.2 Starting values: the subsampling algorithm 141 -- 5.7.3 A strategy for faster subsampling-based algorithms 143 -- 5.7.4 Starting values: the Peñahai estimator 144 -- 5.7.5 Starting values with numeric and categorical predictors 146 -- 5.7.6 Comparing initial estimators 149 -- 5.8 Balancing asymptotic bias and

efficiency 150 -- 5.8.1 "Optimal"; redescending M-estimators 153 -- 5.9 Improving the efficiency of robust regression estimators 155 -- 5.9.1 Improving efficiency with one-step reweighting 155 -- 5.9.2 A fully asymptotically efficient one-step procedure 156 -- 5.9.3 Improving finite-sample efficiency and robustness 158 -- 5.9.4 Choosing a regression estimator 164 -- 5.10 Robust regularized regression 164 -- 5.10.1 Ridge regression 165 -- 5.10.2 Lasso regression 168 -- 5.10.3 Other regularized estimators 171 -- 5.11 *Other estimators 172 -- 5.11.1 Generalized M-estimators 172 -- 5.11.2 Projection estimators 174 -- 5.11.3 Constrained M-estimators 175 -- 5.11.4 Maximum depth estimators 175 -- 5.12 Other topics 176 -- 5.12.1 The exact fit property 176 -- 5.12.2 Heteroskedastic errors 177 -- 5.12.3 A robust multiple correlation coefficient 180 -- 5.13 *Appendix: proofs and complements 182 -- 5.13.1 The BP of monotone M-estimators with random X 182 -- 5.13.2 Heavy-tailed x 183 -- 5.13.3 Proof of the exact fit property 183 -- 5.13.4 The BP of S-estimators 184 -- 5.13.5 Asymptotic bias of M-estimators 186 -- 5.13.6 Hampel optimality for GM-estimators 187 -- 5.13.7 Justification of RFPE; 188.

5.14 Recommendations and software 191 -- 5.15 Problems 191 -- 6 Multivariate Analysis 195 -- 6.1 Introduction 195 -- 6.2 Breakdown and efficiency of multivariate estimators 200 -- 6.2.1 Breakdown point 200 -- 6.2.2 The multivariate exact fit property 201 -- 6.2.3 Efficiency 201 -- 6.3 M-estimators 202 -- 6.3.1 Collinearity 205 -- 6.3.2 Size and shape 205 -- 6.3.3 Breakdown point 206 -- 6.4 Estimators based on a robust scale 207 -- 6.4.1 The minimum volume ellipsoid estimator 208 -- 6.4.2 S-estimators 208 -- 6.4.3 The MCD estimator 210 -- 6.4.4 S-estimators for high dimension 210 -- 6.4.5 "S-estimators 214 -- 6.4.6 One-step reweighting 215 -- 6.5 MM-estimators 215 -- 6.6 The Stahel-Donoho estimator 217 -- 6.7 Asymptotic bias 219 -- 6.8 Numerical computing of multivariate estimators 220 -- 6.8.1 Monotone M-estimators 220 -- 6.8.2 Local solutions for S-estimators 221 -- 6.8.3 Subsampling for estimators based on a robust scale 221 -- 6.8.4 The MVE 223 -- 6.8.5 Computation of S-estimators 223 -- 6.8.6 The MCD 223 -- 6.8.7 The Stahel-Donoho estimator 224 -- 6.9 Faster robust scatter matrix estimators 224 -- 6.9.1 Using pairwise robust covariances 224 -- 6.9.2 The Peña-Prieto procedure 228 -- 6.10 Choosing a location/scatter estimator 229 -- 6.10.1 Efficiency 230 -- 6.10.2 Behavior under contamination 231 -- 6.10.3 Computing times 232 -- 6.10.4 Tuning constants 233 -- 6.10.5 Conclusions 233 -- 6.11 Robust principal components 234 -- 6.11.1 Spherical principal components 236 -- 6.11.2 Robust PCA based on a robust scale 237 -- 6.12 Estimation of multivariate scatter and location with missing data 240 -- 6.12.1 Notation 240 -- 6.12.2 GS estimators for missing data 241 -- 6.13 Robust estimators under the cellwise contamination model 242 -- 6.14 Regularized robust estimators of the inverse of the covariance matrix 245 -- 6.15 Mixed linear models 246 -- 6.15.1 Robust estimation for MLM 248 -- 6.15.2 Breakdown point of MLM estimators 248 -- 6.15.3 S-estimators for MLMs 250. 6.15.4 Composite "S-estimators 250 -- 6.16 *Other estimators of location and scatter 254 -- 6.16.1 Projection estimators 254 -- 6.16.2 Constrained M-estimators 255 -- 6.16.3 Multivariate depth 256 -- 6.17 Appendix: proofs and complements 256 -- 6.17.1 Why affine equivariance? 256 -- 6.17.2 Consistency of equivariant estimators 256 -- 6.17.3 The estimating equations of the MLE 257 -- 6.17.4 Asymptotic BP of monotone M-estimators 258 -- 6.17.5 The estimating equations for S-estimators 260 -- 6.17.6 Behavior of S-

estimators for high p 261 -- 6.17.7 Calculating the asymptotic covariance matrix of location M-estimators 262 -- 6.17.8 The exact fit property 263 -- 6.17.9 Elliptical distributions 264 -- 6.17.10 Consistency of Gnanadesikan-Kettenring correlations 265 -- 6.17.11 Spherical principal components 266 -- 6.17.12 Fixed point estimating equations and computing algorithm for the GS estimator 267 -- 6.18 Recommendations and software 268 -- 6.19 Problems 269 -- 7

Generalized Linear Models 271 -- 7.1 Binary response regression 271 -- 7.2 Robust estimators for the logistic model 275 -- 7.2.1 Weighted MLEs 275 -- 7.2.2 Redescending M-estimators 276 -- 7.3 Generalized linear models 281 -- 7.3.1 Conditionally unbiased bounded influence estimators 283 -- 7.4 Transformed M-estimators 284 -- 7.4.1 Definition of transformed M-estimators 284 -- 7.4.2 Some examples of variance-stabilizing transformations 286 -- 7.4.3 Other estimators for GLMs 286 -- 7.5 Recommendations and software 289 -- 7.6 Problems 290 -- 8 Time Series 293 -- 8.1 Time series outliers and their impact 294 -- 8.1.1 Simple examples of outliers influence 296 -- 8.1.2 Probability models for time series outliers 298 -- 8.1.3 Bias impact of AOs 301 -- 8.2 Classical estimators for AR models 302 -- 8.2.1 The Durbin-Levinson algorithm 305 -- 8.2.2 Asymptotic distribution of classical estimators 307 -- 8.3 Classical estimators for ARMA models 308 -- 8.4 M-estimators of ARMA models 310 -- 8.4.1 M-estimators and their asymptotic distribution 310. 8.4.2 The behavior of M-estimators in AR processes with additive outliers 311 -- 8.4.3 The behavior of LS and M-estimators for ARMA processes with infinite innovation variance 312 -- 8.5 Generalized M-estimators 313 -- 8.6 Robust AR estimation using robust filters 315 -- 8.6.1 Naive minimum robust scale autoregression estimators 315 -- 8.6.2 The robust filter algorithm 316 -- 8.6.3 Minimum robust scale estimators based on robust filtering 318 -- 8.6.4 A robust Durbin-Levinson algorithm 319 -- 8.6.5 Choice of scale for the robust Durbin-Levinson procedure 320 -- 8.6.6 Robust identification of AR order 320 -- 8.7 Robust model identification 321 -- 8.8 Robust ARMA model estimation using robust filters 324 -- 8.8.1 M-estimators of ARMA models 324 -- 8.8.2 Robust filters for ARMA models 326 -- 8.8.3 Robustly filtered M-estimators 328 -- 8.9 ARIMA and SARIMA models 329 -- 8.10 Detecting time series outliers and level shifts 333 -- 8.10.1 Classical detection of time series outliers and level shifts 334 -- 8.10.2 Robust detection of outliers and level shifts for ARIMA models 336 -- 8.10.3 REGARIMA models: estimation and outlier detection 338 -- 8.11 Robustness measures for time series 340 -- 8.11.1 Influence function 340 -- 8.11.2 Maximum bias 342 -- 8.11.3 Breakdown point 343 -- 8.11.4 Maximum bias curves for the AR (1) model 343 -- 8.12 Other approaches for ARMA models 345 -- 8.12.1 Estimators based on robust autocovariances 345 -- 8.12.2 Estimators based on memory- m prediction residuals 346 -- 8.13 High-efficiency robust location estimators 347 -- 8.14 Robust spectral density estimation 348 -- 8.14.1 Definition of the spectral density 348 -- 8.14.2 AR spectral density 349 -- 8.14.3 Classic spectral density estimation methods 349 -- 8.14.4 Prewhitening 350 -- 8.14.5 Influence of outliers on spectral density estimators 351 -- 8.14.6 Robust spectral density estimation 353 -- 8.14.7 Robust time-average spectral density estimator 354 -- 8.15 Appendix A: Heuristic derivation of the asymptotic distribution of M-estimators for ARMA models 356. 8.16 Appendix B: Robust filter covariance recursions 359 -- 8.17 Appendix C: ARMA model state-space representation 360 -- 8.18 Recommendations and software 361 -- 8.19 Problems 361 -- 9 Numerical Algorithms 363 -- 9.1 Regression M-estimators 363 -- 9.2

Regression S-estimators 366 -- 9.3 The LTS-estimator 366 -- 9.4
Scale M-estimators 367 -- 9.4.1 Convergence of the fixed-point
algorithm 367 -- 9.4.2 Algorithms for the non-concave case 368 --
9.5 Multivariate M-estimators 369 -- 9.6 Multivariate S-estimators 370
-- 9.6.1 S-estimators with monotone weights 370 -- 9.6.2 The MCD
371 -- 9.6.3 S-estimators with non-monotone weights 371 -- 9.6.4
*Proof of (9.27) 372 -- 10 Asymptotic Theory of M-estimators 373 --
10.1 Existence and uniqueness of solutions 374 -- 10.1.1
Redescending location estimators 375 -- 10.2 Consistency 376 -- 10.3
Asymptotic normality 377 -- 10.4 Convergence of the SC to the IF 379
-- 10.5 M-estimators of several parameters 381 -- 10.6 Location M-
estimators with preliminary scale 384 -- 10.7 Trimmed means 386 --
10.8 Optimality of the MLE 386 -- 10.9 Regression M-estimators:
existence and uniqueness 388 -- 10.10 Regression M-estimators:
asymptotic normality 389 -- 10.10.1 Fixed X 389 -- 10.10.2
Asymptotic normality: random X 394 -- 10.11 Regression M
estimators: Fisher-consistency 394 -- 10.11.1 Redescending
estimators 394 -- 10.11.2 Monotone estimators 396 -- 10.12
Nonexistence of moments of the sample median 398 -- 10.13
Problems 399 -- 11 Description of Datasets 401 -- References 407 --
Index 423.

Sommario/riassunto

A new edition of this popular text on robust statistics, thoroughly updated to include new and improved methods and focus on implementation of methodology using the increasingly popular open-source software R. Classical statistics fail to cope well with outliers associated with deviations from standard distributions. Robust statistical methods take into account these deviations when estimating the parameters of parametric models, thus increasing the reliability of fitted models and associated inference. This new, second edition of *Robust Statistics: Theory and Methods "with R"* presents a broad coverage of the theory of robust statistics that is integrated with computing methods and applications. Updated to include important new research results of the last decade and focus on the use of the popular software package R, it features in-depth coverage of the key methodology, including regression, multivariate analysis, and time series modeling. The book is illustrated throughout by a range of examples and applications that are supported by a companion website featuring data sets and R code that allow the reader to reproduce the examples given in the book. Unlike other books on the market, *Robust Statistics: Theory and Methods "with R"* offers the most comprehensive, definitive, and up-to-date treatment of the subject. It features chapters on estimating location and scale; measuring robustness; linear regression with fixed and with random predictors; multivariate analysis; generalized linear models; time series; numerical algorithms; and asymptotic theory of M-estimates. . Explains both the use and theoretical justification of robust methods. Guides readers in selecting and using the most appropriate robust methods for their problems. Features computational algorithms for the core methods Robust statistics research results from the past decade included in this 2nd edition are: fast deterministic robust regression, finite-sample robustness, robust regularized regression, robust location and scatter estimation with missing data, robust estimation with independent outliers in variables, and robust mixed linear models. *Robust Statistics* aims to stimulate the use of robust methods as a powerful tool to increase the reliability and accuracy of statistical modelling and data analysis. It is an ideal resource for researchers, practitioners, and graduate students in statistics, engineering, computer science, and physical and social sciences.
