LINEAR REGRESSION MODELS AND ADEQUACY PARAMETERS FOR SCATTERED DATA WITH OUTLIERS

A. V. PANTELEIMONOV, D. O. ANOKHIN, A. B. ZAKHAROV, I. V. KHRISTENKO, A. I. KOROBOV, V. V. IVANOV

References:

1. Harris, C. R.; Millman, K. J.; van der Walt S. J. *et al*. *Array programming with NumPy. Nature*. **2020**, *585*, 357–362.
2. Kholin, Yu.V. Quantitative physicochemical analysis of complex formation in solutions and on the surface of chemically modified silicas: meaningful models, mathematical methods and their applications; Folio, Kharkiv, 2000, 288 p. [in Rus].
3. Huber, P. *Robust Statistics*, J. Wiley and Sons, New York, 1981, 308 p.
4. Clarke B. R. *Linear Models. The Theory and Application of Analysis of Variance*. J. Wiley & Sons, 2008, 241 p.
5. Burnett, V.; Lewis, T. *Outliers in statistical data*. J. Wiley and Sons, New York, 1978, 365 p.
6. Hawkins, D. M. *Identification of Outliers*. Chapman & Hal, 1980, 188 p.
7. Pope, A. J. *The statistics of residuals and the detection of outliers*. NOAA Technical Report NOS 65 NGS1. 1976. (https://www.ngs.noaa.gov/PUBS_LIB/TRNOS65NGS1.pdf)
8. Fox J. *Regression Diagnostics*. Sage publications, London, 1991, 92 p.
9. *Stone, M.; Cross-Validatory Choice and Assessment of Statistical Predictions. Journal of the Royal Statistical Society, Series B.* **1974***, 36 (2), 111–147*.
10. Gray, J.B.; Woodall, W.H. The Maximum Size of standardized and Internally Studentized Residuals in Regression Analysis. *The American Statistician*, **1994**, *48 (2)*, 111-113.
11. Mahalanobis, P. C. On the Generalised Distance in Statistic. *Proceedings of the National Institute of Sciences of India*. **1936**, *2 (1)*, 49-55.
12. Maesschalck, R. De; Jouan-Rimbaud D.; Massart, D.L. The Mahalanobis distance. *Chemometrics and Intelligent Laboratory System*. **2000**, *50*, 1–18.
13. Li, X.; Deng, S.; Li, L.; Jiang Y. Outlier Detection Based on Robust Mahalanobis Distance and Its Application. *Open Journal of Statistics*. **2019**, *9*, 15-26.
14. Cook, R. D. Detection of influential observations in linear regression. *Technometrics*. **1977**, *19(1)*, 15-18.
15. Cook, R. D. *Regression Graphics*. J. Wiley and Sons, New York, 1998, 349 p.
16. Johnson, N. L.; Kotz, S.; Balakrishnan, N. *Continuous univariate distributions*. Wiley-Interscience, 1995, 732 p.
17. Rousseeuw, P. J.; Leroy, A. M. *Robust Regression and Outlier Detection*. J. Wiley and Sons, New York, 1987, 329 p.
18. *Rawlings,* J. O.; *Pantula,* S.G.; *Dickey,* D.A. *Applied Regression Analysis. A Research Tool*. 2nd Edition, *Springer*, 1998, 678 p.
19. Belsley, D. A.; Kuh, E.; Welsch, R. E. *Regression Diagnostics. Identifying Influential Data and Sources of Collinearity*. J. Wiley and Sons, New York, 1980, 292 p.
20. Huber, P. J. *Robust Statistic Procedures*. Society industrial and Applied Mathematics, Philadelphia, 1996, 67 p.
21. Huber, P.J. Robust Statistics. Maximum likelihood type estimates (M-estimates), P. 43-55. J. Wiley and Sons, 1981, New York.
22. Wang, L. ; Zheng, C. ; Zhou, W.; *et. al.* A new principle for tuning-free Huber regression. *Statistica Sinica*, **2021**, *31 (4)*, 2153-2177.
23. Yu C., Yao W. Robust linear regression: A review and comparison. *Communications in Statistics - Simulation and Computation*. **2017**, *46 (8)*, 6261-6282.
24. Yohai, V. J. High breakdown-point and high efficiency robust estimates for regression. *Annals of Statistics*. **1987**, *15(20)*, 642-656.
25. Rousseeuw, P. J.; Driessen, K. V. Computing LTS Regression for Large Data Sets. *Data Mining and Knowledge Discovery*. **2006***, 12*, 29–45.
26. Heyde, C. C.; Seneta, E. (Ed.) *Statisticians of the centuries*. Springer-Verlag, New York. 2001, 500 p.
27. Edgeworth, F. Y. On Observations Relating to Several Quantities. *Hermathena*. **1887**, *6*, 279-285.
28. Branham, R. L. Jr. Alternatives to least squares. *The Astronomical Journal*. **1982**, *87(6)*, 928-937.
29. Stigler, S. M. Newcomb S., Percy D. and the History of Robust Estimation 1885–1920. *Journal of the American Statistical Association*. **1973**, *68*, 872-879.
30. Li, Y.; Arce, G. R. A Maximum Likelihood Approach to Least Absolute Deviation Regression. *EURASIP Journal on Applied Signal Processing*. **2004**, *12*, 1762–1769.

31. Bloomfield, P.; Steiger, W. L. Least Absolute Deviations: Theory, Applications and Algorithms. *Progress in probability and statistics*. Birkhauser, **1983**, 349 p,
32. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. Roy. Statist. Soc.* **1996**, *58 (1)*, 267–288.
33. Gordon, M.; Zhu, J.; Wang, L. Regularized Least Absolute Deviations Regression and an Efficient Algorithm for Parameter Tuning. *Sixth International Conference on Data Mining (ICDM'06)*, Hong Kong, **2006**, 690-700.
34. Dodge, Y. (Eds.) Statistical Data Analysis Based on the L1-Norm and Related Methods. Springer Basel AG, **2002**, 454 p.
35. Farebrother, R. *L_1-Norm and L_∞-Norm Estimation: An Introduction to the Least Absolute Residuals, the Minimax Absolute Residual and Related Fitting Procedures.* Springer Berlin Heidelberg, **2013**, 58 p.
36. Berdnyk, M. I.; Zakharov, A. B.; Ivanov, V. V. Application of L_1-regularization approach in QSAR problem. Linear regression and artificial neural networks. *Methods and Objects of Chemical Analysis*. **2019**, *14 (2)*, 79-90.
37. Ivanov, V.V.; Berdnik, M. I.; Adamowicz, L. L_1 – regularisation of the Coupled Cluster solutions. *Molecular Physics*. **2017**, *115(21-22)*, 2892-2902.
38. Ivanov, V.V. L_1-regularized solutions of coupled cluster theory equations. Test system F_2. *Kharkov University Bulletin. Chemical Series*. **2017**, *28(51)*, 30-34.
39. Fernandes, R.; Leblanc, S. G. Parametric (modified least squares) and non-parametric (Theil-Sen) linear regressions for predicting biophysical parameters in the presence of measurement errors. *Remote Sensing of Environment*. **2005**, *95*, 303–316.
40. Ahn, S. J. *Least Squares Orthogonal Distance Fitting of Curves and Surfaces in Space.* Springer-Verlag: Berlin, Heidelberg, **2004**, 127 p.
41. Onizhuk, M. O.; Ivanov, V. V.; Panteleimonov, A. V. *et al.* Alternative Methods for Constructing of Linear Regressions. *Methods and Objects of Chemical Analysis*. **2017**, *12(3)*, 105–111.
42. Berdnyk, M. I.; Onizhuk, M. O.; Ivanov, V. V. Methods for building linear regression equations in the "structure-property" problems. *Kharkov University Bulletin. Chemical Series*. **2018**, *30 (53)*, 6-17.
43. Bugaevsky, A. A.; Kholin, Yu. V. Computer-aided determination of the composition and stability of complex compounds in solutions with complicated equilibria. *Anal. Chim. Acta*. **1991**, *249 (2)*, 353–365.
44. Myerniy, S.A.; Konyaev, D. S.; Kholin, Yu.V. Robust parameter estimation in problems of quantitative physical and chemical analysis. *Kharkov University Bulletin. Chemical Series*. **1998**, *420 (2)*, 6–17. [in Rus].
45. Myerniy, S.; Varshal, G.; Kholin, Yu. Determination of affinity distributions: numerical algorithm and its application for estimating energetic heterogeneity of complexing silicas and humic substances. *Adsorption Science & Technology*. **2000**, *18 (3)*, 267–294.
46. Khristenko, I. V; Panteleimonov, A. V.; Iliashenko R. Yu.; *et al* Heterogeneous polarity and surface acidity of silica-organic materials with fixed 1-n-propyl-3-methylimidazolium chloride as probed by solvatochromic and fluorescent dyes. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*. **2018**, *538,* 280-286.