Linear regression

The **linear regression,** which is a special case of the general concept of regression analysis , is a statistical method used to explain an observed dependent variable by one or more independent variables . The word "linear" results from the fact that the dependent variable is a linear combination of the regression coefficients (but not necessarily the independent variable). The term "regression" or " regression" to the center was mainly characterized by the statistician Francis Galton .

Table of Contents

Simple Linear Regression

→ *Main article : Simple linear regression*

The simple linear regression model is based on only two metric variables: an influencing variable X And a target size Y, The simple linear regression uses two parameters to place a straight line through a point cloud so that the linear relationship between$X$ and YAs well as possible. The equation of the linear single regression is given by

$$Y_i = \beta_0 + \beta_1 x_i + U_i, \quad i = 1, \ldots, n,$$

Multiple Linear Regression

→ *Main article : Multiple linear regression*

The multiple linear regression represents a generalization of the simple linear regression, assuming K regressors to explain the dependent variable. Thus, in addition to the variation over the observations, a variation over the regressors is assumed, resulting in a linear system of equations that can be summarized as follows in matrix notation:

$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$$

Normal linear model

→ *Main article : Classic linear model*

If the assumption of the normal distribution of the error terms is also made with respect to the previous multiple linear model, then one also speaks of a classical linear model. The assumption of the normal distribution of the error terms is required to perform statistical inference, It is required to be able to calculate confidence intervals and the like.

$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\varepsilon} \text{ With } \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_T),$$

Paneldate Regression

→ *Main article : Linear panel data models*

The general linear panel data model allows the axis section and the slope parameters on the one hand over the individuals$i$ (In cross section) and on the other hand over time tVary (non-time invariant). The general linear panel data model is:

$$y_{it} = \alpha_{it} + \mathbf{x}_{it}^\top \beta_{it} + \varepsilon_{it}, \;\; i = 1, \ldots, N; \;\; t = 1, \ldots, T$$

With the variance covariance matrix:

$$\operatorname{Cov}(\boldsymbol{\varepsilon}) = \operatorname{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top) = \mathbf{\Sigma} \otimes \mathbf{I}_T = \mathbf{\Phi}$$

In this case y_{it} A scalar dependent dependent variable, $\mathbf{x}_{it}^\top$ is a $(K \times 1)$Vector of independent variables, ε_{it}Is a scalar error. Since this model is too general and can not be estimated when there are more parameters than observations,α_{it} and β_{it} With i and tAnd limiting assumptions regarding the behavior of the error term. These additional restrictions and the models based on them are topics of the linear paneldate models and the Paneldatenanalyse .

Generalized linear models

→ *Main articles : Generalized linear models*

Linear models can be extended in such a way that no fixed data matrix is examined, but this is randomly adhered to. In this case, the investigation methods do not change substantially, but become much more complex and therefore more complex.

General Linear Models

→ *Main article : General linear model*

The general linear model considers the situation where the dependent variable YIs not a scalar but a vector. In this case also conditioned linearity becomes$\operatorname{E}(\mathbf{y}|\mathbf{X}) = \mathbf{BX}$ As assumed in the classical linear model, but with a matrix BRepresenting the vector βOf the classical linear model. Multivariate pendants to the usual method of least squares and to the method of generalized least squares were developed. *General linear models* are also called "multivariate linear models". However, these are not to be confused with multiple linear models. The general linear model is given by

$$\mathbf{Y} = \mathbf{XB} + \mathbf{U},$$

Orthogonal Regression

→ *Main article : Orthogonal Regression*

The orthogonal regression (more precisely: orthogonal linear regression) is used to calculate a balance line for a finite set of metrically scaled data pairs (x_i, y_i) According to the method of least squares.

Regularization of Regression

In order to ensure a desired behavior of the regression and thus to avoid an over-adjustment to the training data set , there is the possibility of providing the regression term with *penalty* tables, which occur as secondary conditions .

Among the best known regularizations are: [1] [2]

- The L_1-regulatoryisation (also called LASSO regularization): by $\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}}(\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|_1)$

  Are preferably individual elements of the vector $\hat{\boldsymbol{\beta}}$minimized. However, the other elements of the vector can (in terms of magnitude) assume large values. This favors the formation of thin-stranded matrices , which allows for more efficient algorithms.
- The L_2-Regularization (also called Ridge regularization): Through $\hat{\boldsymbol{\beta}} = \arg\min_{\beta}(\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|^2)$ Is the whole vector $\hat{\boldsymbol{\beta}}$Uniformly minimized, but the matrices are *fuller.*
- The elastic net: Here, by the expression $\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}}(\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda_2\|\boldsymbol{\beta}\|^2 + \lambda_1\|\boldsymbol{\beta}\|_1)$ Both the L_1

  As well as theL_2-regulation.

Regression analysis applications

Special applications of the regression analysis also refer to the analysis of discrete and restricted dependent variables. Here, a distinction can be made between the type of dependent variables and the type of the restriction of the value range. The regression models, which can be applied at this point, are listed below. More details can be found in Frone (1997) [3] and Long (1997). [4]

Models for different types-dependent variables (generalized linear models):

- Binary: Logistic regression and probit regression
- Ordinal: Ordinal logistic regression and ordinal probit regression
- Absolute: Poisson regression , negative binomial regression
- Nominal: Multinomial logistic regression

Models for different types of restricted values:

- Censored: Tobit model
- Truncated: truncated regression
- Sample- *selected:* sample-selegated regression

Application in Econometrics

For quantitative economic analyzes within the framework of the regression analysis, such as econometrics , are particularly suitable:

- Growth functions , such as the law of organic growth or interest rate calculation ,
- Decay functions , such as the hyperbolic distribution function or the Korachian price function ,
- Schwanenhalsfunktionen , such as the logistic function used in the context of the logistic regression , the Johnson function or the potency exponential function ,
- Degressive saturation functions , such as the Gompertz function or the traverse function .

Single Signature

1. Andrew Y. Ng: *Feature Selection, L_1 vs.. L_2 regularization, and rotational invariance*. *(http://www.machinelearning.org/proceedings/icml2004/papers/354.pdf)*In: *International Conference on Machine Learning (ICML)*. (PDF).
2. Hui Zou, Trevor Hastie: *Regularization and Variable Selection via the Elastic Net*. *(http://web.stanford.edu/~hastie/TALKS/enet_talk.pdf)*(PDF).
3. MR Frone: *Regression models for discrete and limited dependent variables*. Research Methods Forum No. 2, 1997, *online*. (https://web.archive.org/web/20070107012608/http://division.aomonline.org/rm/1997_forum_regression_models.html)(Memento of 7 January 2007 in the *Internet Archive*).
4. JS Long: *Regression models for categorical and limited dependent variables*. Sage, Thousand Oaks, CA 1997.

See also

- measurement error
- Generalized method of least squares
- regression analysis

References

- Norman R. Draper, Harry Smith: *Applied Regression Analysis*. 3. Edition. Wiley, New York 1998, ISBN 0-471-17082-8 .
- Ludwig Fahrmeir , Thomas Kneib, Stefan Lang: *Regression: Models, Methods and Applications*. Springer Verlag, Berlin, Heidelberg / New York 2007, ISBN 978-3-540-33932-8 .
- Gerhard Opfer: *Numerical Mathematics for Beginners*. 2nd Edition. Vieweg Verlag, 1994, ISBN 3-528-17265-7 .
- Volker Oppitz, Volker Nollau: *Paperback Profitability* Analysis . Carl Hanser Verlag, 2003, ISBN 3-446-22463-7 .
- Volker Oppitz: *Gabler Lexicon Profitability Analysis*. Gabler Verlag, 1995, ISBN 3-409-19951-9 .
- Peter Schönfeld: *Methods of Econometrics* . Berlin / Frankfurt 1969.
- Dieter Urban, Jochen Mayerl: *Regression Analysis: Theory, Technology and Application*. 2nd, revised. Edition. VS Verlag, Wiesbaden, 2006, ISBN 3-531-33739-4 .
- E. Zeidler (Eds.): *Taschenbuch der Mathematik*. (Known as *Bronstein and Semendaev.*) Stuttgart / Leipzig / Wiesbaden 2003, ISBN 3-519-20012-0 .
- K. Backhaus, B. Erichson, W. Plinke, R. Weiber: *Multivariate Analysis Methods*. *An application-oriented introduction*. *(http://www.multivariate.de/)*12., fully overh. Edition. Berlin & a. 2008, ISBN 978-3-540-85044-1 .
- W. Zucchini, A. Schlegel, O. Nenadíc, S. Sperlich: *Statistics for Bachelor and Master students*. Springer Verlag, Berlin, Heidelberg, ISBN 978-3-540-88986-1 .
- A. Colin Cameron, Pravin K. Trivedi: *Microeconometrics*. *Methods and Applications*. Cambridge University Press, 2005, ISBN 0-521-84805-9 .
- G. Judge, R. Carterhill: *Introduction to the Theory and Practice of Econometrics*. Wiley, New York 1988, ISBN 0-471-62414-4 .

Web Links

Wikibooks: Introduction to regression analysis - learning and teaching materials
Commons: Linear Regression (https://commons.wikimedia.org/wiki/Category:Linear_regression?uselang=de) - Collection of pictures, videos and audio files

- Evaluate bivariate data collections (regression line, correlation coefficient) (http://ne.lo-net2.de/selbstlernmaterial/m/wk/lr/lrindex.html)
- Examples of circular regression and exponential regression (http://www.block-net.de/messtechnik/p-messtechnik.html)
- Detailed video on the linear regression including demonstration of a case study (http://www.phimea.de/fachbeitraege/video-tutorial-lineare-regression)