Linear regression

In statistics, **linear regression** is an approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X. The case of one explanatory variable is called *simple linear regression*. For more than one explanatory variable, the process is called *multiple linear regression*.[1] (This term should be distinguished from *multivariate linear regression*, where multiple correlated dependent variables are predicted, rather than a single scalar variable.)[2]

In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called *linear models*.[3] Most commonly, the conditional mean of y given the value of X is assumed to be an affine function of X; less commonly, the median or some other quantile of the conditional distribution of y given X is expressed as a linear function of X. Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of y given X, rather than on the joint probability distribution of y and X, which is the domain of multivariate analysis.

Linear regression was the first type of regression analysis to be studied rigorously, and to be used extensively in practical applications.[4] This is because models which depend linearly on their unknown parameters are easier to fit than models which are non-linearly related to their parameters and because the statistical properties of the resulting estimators are easier to determine.

Linear regression has many practical uses. Most applications fall into one of the following two broad categories:

- If the goal is prediction, or forecasting, or error reduction, linear regression can be used to fit a predictive model to an observed data set of y and X values. After developing such a model, if an additional value of X is then given without its accompanying value of y, the fitted model can be used to make a prediction of the value of y.

- Given a variable y and a number of variables X_1, ..., Xp that may be related to y, linear regression analysis can be applied to quantify the strength of the relationship between y and the Xj, to assess which Xj may have no relationship with y at all, and to identify which subsets of the Xj contain redundant information about y.

Linear regression models are often fitted using the least squares approach, but they may also be fitted in other ways, such as by minimizing the "lack of fit" in some other norm (as with least absolute deviations regression), or by minimizing a penalized version of the least squares loss function as in ridge regression (L2-norm penalty) and lasso (L1-norm penalty). Conversely, the least squares approach can be used to fit models that are not linear models. Thus, although the terms "least squares" and "linear model" are closely linked, they are not synonymous.

1 Introduction
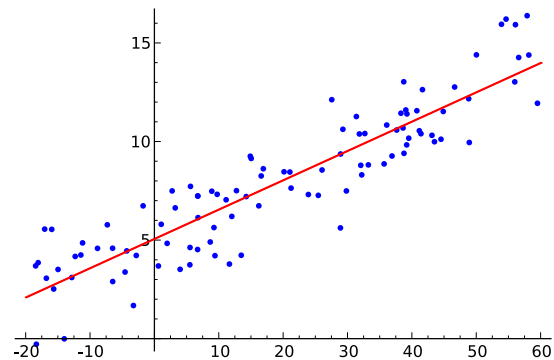
Example of simple linear regression, which has one independent variable
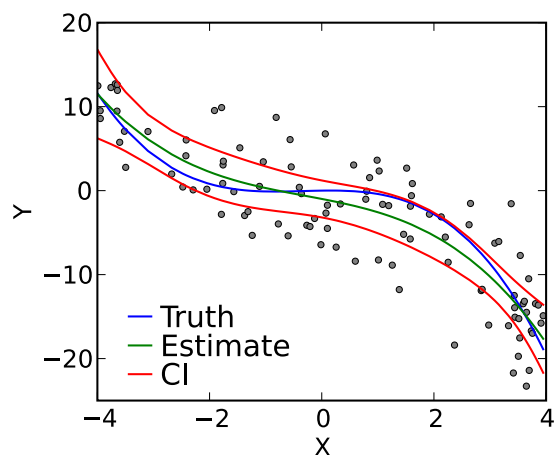
Example of a cubic polynomial regression, which is a type of linear regression.

Given a data set $\{y_i,\, x_{i1}, \ldots, x_{ip}\}_{i=1}^{n}$ of n statistical units, a linear regression model assumes that the relationship between the dependent variable yi and the p-vector

of regressors xi is linear. This relationship is modeled through a *disturbance term* or *error variable* εi — an unobserved random variable that adds noise to the linear relationship between the dependent variable and regressors. Thus the model takes the form

$$y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^{\mathsf{T}} \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where $^{\mathsf{T}}$ denotes the transpose, so that $xi^{\mathsf{T}}\beta$ is the inner product between vectors xi and β.

Often these n equations are stacked together and written in vector form as

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

where

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^{\mathsf{T}} \\ \mathbf{x}_2^{\mathsf{T}} \\ \vdots \\ \mathbf{x}_n^{\mathsf{T}} \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix},$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Some remarks on terminology and general use:

- y_i is called the *regressand*, *endogenous variable*, *response variable*, *measured variable*, *criterion variable*, or *dependent variable* (see dependent and independent variables.) The decision as to which variable in a data set is modeled as the dependent variable and which are modeled as the independent variables may be based on a presumption that the value of one of the variables is caused by, or directly influenced by the other variables. Alternatively, there may be an operational reason to model one of the variables in terms of the others, in which case there need be no presumption of causality.

- $x_{i1}, x_{i2}, \ldots, x_{ip}$ are called *regressors*, *exogenous variables*, *explanatory variables*, *covariates*, *input variables*, *predictor variables*, or *independent variables* (see dependent and independent variables, but not to be confused with independent random variables). The matrix $\mathbf{X}$ is sometimes called the design matrix.

- Usually a constant is included as one of the regressors. For example, we can take $xi_1 = 1$ for $i = 1, \ldots, n$. The corresponding element of $\boldsymbol{\beta}$ is called the *intercept*. Many statistical inference procedures for linear models require an intercept to be present, so it is often included even if theoretical considerations suggest that its value should be zero.

- Sometimes one of the regressors can be a non-linear function of another regressor or of the data, as in polynomial regression and segmented regression. The model remains linear as long as it is linear in the parameter vector $\boldsymbol{\beta}$.

- The regressors xij may be viewed either as random variables, which we simply observe, or they can be considered as predetermined fixed values which we can choose. Both interpretations may be appropriate in different cases, and they generally lead to the same estimation procedures; however different approaches to asymptotic analysis are used in these two situations.

- $\boldsymbol{\beta}$ is a p-dimensional *parameter vector*. Its elements are also called *effects*, or *regression coefficients*. Statistical estimation and inference in linear regression focuses on $\boldsymbol{\beta}$. The elements of this parameter vector are interpreted as the partial derivatives of the dependent variable with respect to the various independent variables.

- ε_i is called the *error term*, *disturbance term*, or *noise*. This variable captures all other factors which influence the dependent variable yi other than the regressors xi. The relationship between the error term and the regressors, for example whether they are correlated, is a crucial step in formulating a linear regression model, as it will determine the method to use for estimation.

Example. Consider a situation where a small ball is being tossed up in the air and then we measure its heights of ascent hi at various moments in time ti. Physics tells us that, ignoring the drag, the relationship can be modeled as

$$h_i = \beta_1 t_i + \beta_2 t_i^2 + \varepsilon_i,$$

where β_1 determines the initial velocity of the ball, β_2 is proportional to the standard gravity, and εi is due to measurement errors. Linear regression can be used to estimate the values of β_1 and β_2 from the measured data. This model is non-linear in the time variable, but it is linear in the parameters β_1 and β_2; if we take regressors $xi = (xi_1, xi_2) = (ti, ti^2)$, the model takes on the standard form

$$h_i = \mathbf{x}_i^{\mathsf{T}} \boldsymbol{\beta} + \varepsilon_i.$$

1.1   Assumptions

Standard linear regression models with standard estimation techniques make a number of assumptions about the predictor variables, the response variables and their relationship. Numerous extensions have been developed that allow each of these assumptions to be relaxed (i.e. reduced to a weaker form), and in some cases eliminated entirely. Some methods are general enough that they can relax multiple assumptions at once, and in other cases this can be achieved by combining different extensions. Generally these extensions make the estimation procedure more complex and time-consuming, and may also require more data in order to produce an equally precise model.

The following are the major assumptions made by standard linear regression models with standard estimation techniques (e.g. ordinary least squares):

- **Weak exogeneity**. This essentially means that the predictor variables x can be treated as fixed values, rather than random variables. This means, for example, that the predictor variables are assumed to be error-free—that is, not contaminated with measurement errors. Although this assumption is not realistic in many settings, dropping it leads to significantly more difficult errors-in-variables models.

- **Linearity**. This means that the mean of the response variable is a linear combination of the parameters (regression coefficients) and the predictor variables. Note that this assumption is much less restrictive than it may at first seem. Because the predictor variables are treated as fixed values (see above), linearity is really only a restriction on the parameters. The predictor variables themselves can be arbitrarily transformed, and in fact multiple copies of the same underlying predictor variable can be added, each one transformed differently. This trick is used, for example, in polynomial regression, which uses linear regression to fit the response variable as an arbitrary polynomial function (up to a given rank) of a predictor variable. This makes linear regression an extremely powerful inference method. In fact, models such as polynomial regression are often "too powerful", in that they tend to overfit the data. As a result, some kind of regularization must typically be used to prevent unreasonable solutions coming out of the estimation process. Common examples are ridge regression and lasso regression. Bayesian linear regression can also be used, which by its nature is more or less immune to the problem of overfitting. (In fact, ridge regression and lasso regression can both be viewed as special cases of Bayesian linear regression, with particular types of prior distributions placed on the regression coefficients.)

- **Constant variance** (a.k.a. **homoscedasticity**). This means that different response variables have the same variance in their errors, regardless of the values of the predictor variables. In practice this assumption is invalid (i.e. the errors are heteroscedastic) if the response variables can vary over a wide scale. In order to determine for heterogeneous error variance, or when a pattern of residuals violates model assumptions of homoscedasticity (error is equally variable around the 'best-fitting line' for all points of x), it is prudent to look for a "fanning effect" between residual error and predicted values. This is to say there will be a systematic change in the absolute or squared residuals when plotted against the predicting outcome. Error will not be evenly distributed across the regression line. Heteroscedasticity will result in the averaging over of distinguishable variances around the points to get a single variance that is inaccurately representing all the variances of the line. In effect, residuals appear clustered and spread apart on their predicted plots for larger and smaller values for points along the linear regression line, and the mean squared error for the model will be wrong. Typically, for example, a response variable whose mean is large will have a greater variance than one whose mean is small. For example, a given person whose income is predicted to be \$100,000 may easily have an actual income of \$80,000 or \$120,000 (a standard deviation of around \$20,000), while another person with a predicted income of \$10,000 is unlikely to have the same \$20,000 standard deviation, which would imply their actual income would vary anywhere between -\$10,000 and \$30,000. (In fact, as this shows, in many cases—often the same cases where the assumption of normally distributed errors fails—the variance or standard deviation should be predicted to be proportional to the mean, rather than constant.) Simple linear regression estimation methods give less precise parameter estimates and misleading inferential quantities such as standard errors when substantial heteroscedasticity is present. However, various estimation techniques (e.g. weighted least squares and heteroscedasticity-consistent standard errors) can handle heteroscedasticity in a quite general way. Bayesian linear regression techniques can also be used when the variance is assumed to be a function of the mean. It is also possible in some cases to fix the problem by applying a transformation to the response variable (e.g. fit the logarithm of the response variable using a linear regression model, which implies that the response variable has a log-normal distribution rather than a normal distribution).

- **Independence** of errors. This assumes that the errors of the response variables are uncorrelated with each other. (Actual statistical independence is a stronger condition than mere lack of correlation and is often not needed, although it can be exploited if it is known to hold.) Some methods (e.g. generalized

least squares) are capable of handling correlated errors, although they typically require significantly more data unless some sort of regularization is used to bias the model towards assuming uncorrelated errors. Bayesian linear regression is a general way of handling this issue.

- **Lack of multicollinearity** in the predictors. For standard least squares estimation methods, the design matrix X must have full column rank p; otherwise, we have a condition known as multicollinearity in the predictor variables. This can be triggered by having two or more perfectly correlated predictor variables (e.g. if the same predictor variable is mistakenly given twice, either without transforming one of the copies or by transforming one of the copies linearly). It can also happen if there is too little data available compared to the number of parameters to be estimated (e.g. fewer data points than regression coefficients). In the case of multicollinearity, the parameter vector β will be non-identifiable—it has no unique solution. At most we will be able to identify some of the parameters, i.e. narrow down its value to some linear subspace of $\mathbf{R}^p$. See partial least squares regression. Methods for fitting linear models with multicollinearity have been developed;[5][6][7][8] some require additional assumptions such as "effect sparsity"—that a large fraction of the effects are exactly zero.
Note that the more computationally expensive iterated algorithms for parameter estimation, such as those used in generalized linear models, do not suffer from this problem—and in fact it's quite normal when handling categorically valued predictors to introduce a separate indicator variable predictor for each possible category, which inevitably introduces multicollinearity.

Beyond these assumptions, several other statistical properties of the data strongly influence the performance of different estimation methods:

- The statistical relationship between the error terms and the regressors plays an important role in determining whether an estimation procedure has desirable sampling properties such as being unbiased and consistent.

- The arrangement, or probability distribution of the predictor variables x has a major influence on the precision of estimates of β. Sampling and design of experiments are highly developed subfields of statistics that provide guidance for collecting data in such a way to achieve a precise estimate of β.

1.2　Interpretation

A fitted linear regression model can be used to identify the relationship between a single predictor variable xj and the
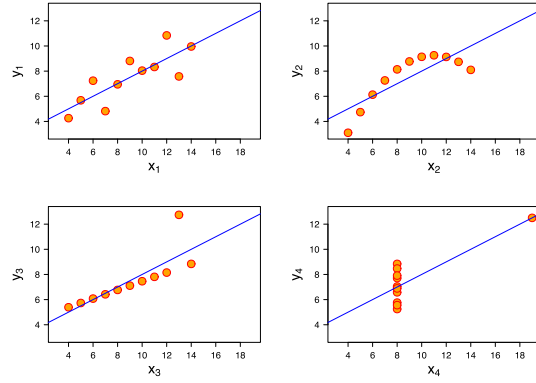
The sets in the Anscombe's quartet have the same linear regression line but are themselves very different.

response variable y when all the other predictor variables in the model are "held fixed". Specifically, the interpretation of βj is the expected change in y for a one-unit change in xj when the other covariates are held fixed—that is, the expected value of the partial derivative of y with respect to xj. This is sometimes called the *unique effect* of xj on y. In contrast, the *marginal effect* of xj on y can be assessed using a correlation coefficient or simple linear regression model relating xj to y; this effect is the total derivative of y with respect to xj.

Care must be taken when interpreting regression results, as some of the regressors may not allow for marginal changes (such as dummy variables, or the intercept term), while others cannot be held fixed (recall the example from the introduction: it would be impossible to "hold ti fixed" and at the same time change the value of ti^2).

It is possible that the unique effect can be nearly zero even when the marginal effect is large. This may imply that some other covariate captures all the information in xj, so that once that variable is in the model, there is no contribution of xj to the variation in y. Conversely, the unique effect of xj can be large while its marginal effect is nearly zero. This would happen if the other covariates explained a great deal of the variation of y, but they mainly explain variation in a way that is complementary to what is captured by xj. In this case, including the other variables in the model reduces the part of the variability of y that is unrelated to xj, thereby strengthening the apparent relationship with xj.

The meaning of the expression "held fixed" may depend on how the values of the predictor variables arise. If the experimenter directly sets the values of the predictor variables according to a study design, the comparisons of interest may literally correspond to comparisons among units whose predictor variables have been "held fixed" by the experimenter. Alternatively, the expression "held fixed" can refer to a selection that takes place in the context of data analysis. In this case, we "hold a variable fixed" by restricting our attention to the subsets of the data that happen to have a common value for the given predic-

tor variable. This is the only interpretation of "held fixed" that can be used in an observational study.

The notion of a "unique effect" is appealing when studying a complex system where multiple interrelated components influence the response variable. In some cases, it can literally be interpreted as the causal effect of an intervention that is linked to the value of a predictor variable. However, it has been argued that in many cases multiple regression analysis fails to clarify the relationships between the predictor variables and the response variable when the predictors are correlated with each other and are not assigned following a study design.[9] A commonality analysis may be helpful in disentangling the shared and unique impacts of correlated independent variables.[10]

2 Extensions

Numerous extensions of linear regression have been developed, which allow some or all of the assumptions underlying the basic model to be relaxed.

2.1 Simple and multiple regression

The very simplest case of a single scalar predictor variable x and a single scalar response variable y is known as *simple linear regression*. The extension to multiple and/or vector-valued predictor variables (denoted with a capital X) is known as *multiple linear regression*, also known as *multivariable linear regression*. Nearly all real-world regression models involve multiple predictors, and basic descriptions of linear regression are often phrased in terms of the multiple regression model. Note, however, that in these cases the response variable y is still a scalar. Another term *multivariate linear regression* refers to cases where y is a vector, i.e., the same as *general linear regression*. The difference between *multivariate* linear regression and *multivariable* linear regression should be emphasized as it causes much confusion and misunderstanding in the literature.

2.2 General linear models

The general linear model considers the situation when the response variable Y is not a scalar but a vector. Conditional linearity of $E(y|x) = Bx$ is still assumed, with a matrix B replacing the vector β of the classical linear regression model. Multivariate analogues of Ordinary Least-Squares (OLS) and Generalized Least-Squares (GLS) have been developed. The term "general linear models" is equivalent to "multivariate linear models". It should be noted the difference of "multivariate linear models" and "multivariable linear models," where the former is the same as "general linear models" and the latter is the same as "multiple linear models."

2.3 Heteroscedastic models

Various models have been created that allow for heteroscedasticity, i.e. the errors for different response variables may have different variances. For example, weighted least squares is a method for estimating linear regression models when the response variables may have different error variances, possibly with correlated errors. (See also Weighted linear least squares, and generalized least squares.) Heteroscedasticity-consistent standard errors is an improved method for use with uncorrelated but potentially heteroscedastic errors.

2.4 Generalized linear models

Generalized linear models (GLMs) are a framework for modeling a response variable y that is bounded or discrete. This is used, for example:

- when modeling positive quantities (e.g. prices or populations) that vary over a large scale—which are better described using a skewed distribution such as the log-normal distribution or Poisson distribution (although GLMs are not used for log-normal data, instead the response variable is simply transformed using the logarithm function);

- when modeling categorical data, such as the choice of a given candidate in an election (which is better described using a Bernoulli distribution/binomial distribution for binary choices, or a categorical distribution/multinomial distribution for multi-way choices), where there are a fixed number of choices that cannot be meaningfully ordered;

- when modeling ordinal data, e.g. ratings on a scale from 0 to 5, where the different outcomes can be ordered but where the quantity itself may not have any absolute meaning (e.g. a rating of 4 may not be "twice as good" in any objective sense as a rating of 2, but simply indicates that it is better than 2 or 3 but not as good as 5).

Generalized linear models allow for an arbitrary *link function g* that relates the mean of the response variable to the predictors, i.e. $E(y) = g(\beta'x)$. The link function is often related to the distribution of the response, and in particular it typically has the effect of transforming between the $(-\infty, \infty)$ range of the linear predictor and the range of the response variable.

Some common examples of GLMs are:

- Poisson regression for count data.

- Logistic regression and probit regression for binary data.

- Multinomial logistic regression and multinomial probit regression for categorical data.

- Ordered probit regression for ordinal data.

Single index models allow some degree of nonlinearity in the relationship between *x* and *y*, while preserving the central role of the linear predictor $\beta'x$ as in the classical linear regression model. Under certain conditions, simply applying OLS to data from a single-index model will consistently estimate β up to a proportionality constant.[11]

2.5 Hierarchical linear models

Hierarchical linear models (or *multilevel regression*) organizes the data into a hierarchy of regressions, for example where *A* is regressed on *B*, and *B* is regressed on *C*. It is often used where the variables of interest have a natural hierarchical structure such as in educational statistics, where students are nested in classrooms, classrooms are nested in schools, and schools are nested in some administrative grouping, such as a school district. The response variable might be a measure of student achievement such as a test score, and different covariates would be collected at the classroom, school, and school district levels.

2.6 Errors-in-variables

Errors-in-variables models (or "measurement error models") extend the traditional linear regression model to allow the predictor variables *X* to be observed with error. This error causes standard estimators of β to become biased. Generally, the form of bias is an attenuation, meaning that the effects are biased toward zero.

## 2.7 Others

- In Dempster–Shafer theory, or a linear belief function in particular, a linear regression model may be represented as a partially swept matrix, which can be combined with similar matrices representing observations and other assumed normal distributions and state equations. The combination of swept or unswept matrices provides an alternative method for estimating linear regression models.

3 Estimation methods

A large number of procedures have been developed for parameter estimation and inference in linear regression. These methods differ in computational simplicity of algorithms, presence of a closed-form solution, robustness with respect to heavy-tailed distributions, and theoretical assumptions needed to validate desirable statistical properties such as consistency and asymptotic efficiency.

Some of the more common estimation techniques for linear regression are summarized below.
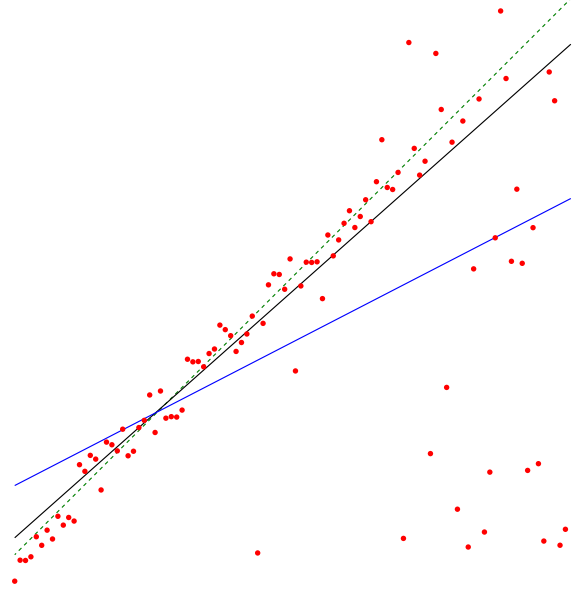
Comparison of the Theil–Sen estimator (black) and simple linear regression (blue) for a set of points with outliers.

3.1 Least-squares estimation and related techniques

- **Ordinary least squares** (OLS) is the simplest and thus most common estimator. It is conceptually simple and computationally straightforward. OLS estimates are commonly used to analyze both experimental and observational data.

  The OLS method minimizes the sum of squared residuals, and leads to a closed-form expression for the estimated value of the unknown parameter β:

  $$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} = \left(\sum \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}}\right)^{-1}\left(\sum \mathbf{x}_i y_i\right).$$

  The estimator is unbiased and consistent if the errors have finite variance and are uncorrelated with the regressors[12]

  $$\mathrm{E}\big[\,\mathbf{x}_i \varepsilon_i\,\big] = 0.$$

  It is also efficient under the assumption that the errors have finite variance and are homoscedastic, meaning that $\mathrm{E}[\varepsilon_i^2|\mathbf{x}_i]$ does not depend on i. The condition that the errors are uncorrelated with the regressors will generally be satisfied in an experiment, but in the case of observational data, it is difficult to exclude the possibility of an omitted covariate z that is related to both the observed covariates and the response variable. The existence of such a covariate will generally lead to a correlation between the regressors and the response variable, and hence to an inconsistent estimator of $\boldsymbol{\beta}$. The condition of homoscedasticity can fail with either experimental

or observational data. If the goal is either inference or predictive modeling, the performance of OLS estimates can be poor if multicollinearity is present, unless the sample size is large.

In simple linear regression, where there is only one regressor (with a constant), the OLS coefficient estimates have a simple form that is closely related to the correlation coefficient between the covariate and the response.

- **Generalized least squares** (GLS) is an extension of the OLS method, that allows efficient estimation of β when either heteroscedasticity, or correlations, or both are present among the error terms of the model, as long as the form of heteroscedasticity and correlation is known independently of the data. To handle heteroscedasticity when the error terms are uncorrelated with each other, GLS minimizes a weighted analogue to the sum of squared residuals from OLS regression, where the weight for the ith case is inversely proportional to var(εi). This special case of GLS is called "weighted least squares". The GLS solution to estimation problem is

$$\hat{\beta} = (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{y},$$

where $\boldsymbol{\Omega}$ is the covariance matrix of the errors. GLS can be viewed as applying a linear transformation to the data so that the assumptions of OLS are met for the transformed data. For GLS to be applied, the covariance structure of the errors must be known up to a multiplicative constant.

- **Percentage least squares** focuses on reducing percentage errors, which is useful in the field of forecasting or time series analysis. It is also useful in situations where the dependent variable has a wide range without constant variance, as here the larger residuals at the upper end of the range would dominate if OLS were used. When the percentage or relative error is normally distributed, least squares percentage regression provides maximum likelihood estimates. Percentage regression is linked to a multiplicative error model, whereas OLS is linked to models containing an additive error term.[13]

- **Iteratively reweighted least squares** (IRLS) is used when heteroscedasticity, or correlations, or both are present among the error terms of the model, but where little is known about the covariance structure of the errors independently of the data.[14] In the first iteration, OLS, or GLS with a provisional covariance structure is carried out, and the residuals are obtained from the fit. Based on the residuals, an improved estimate of the covariance structure of the errors can usually be obtained. A subsequent GLS iteration is then performed using this estimate of the error structure to define the weights. The process can be iterated to convergence, but in many cases, only one iteration is sufficient to achieve an efficient estimate of β.[15][16]

- **Instrumental variables** regression (IV) can be performed when the regressors are correlated with the errors. In this case, we need the existence of some auxiliary *instrumental variables* $\mathbf{z}i$ such that E[$\mathbf{z}i\varepsilon i$] = 0. If $\mathbf{Z}$ is the matrix of instruments, then the estimator can be given in closed form as

$$\hat{\beta} = (\mathbf{X}^{\mathrm{T}}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{y}.$$

- **Optimal instruments** regression is an extension of classical IV regression to the situation where E[$\varepsilon i|\mathbf{z}i$] = 0.

- **Total least squares** (TLS)[17] is an approach to least squares estimation of the linear regression model that treats the covariates and response variable in a more geometrically symmetric manner than OLS. It is one approach to handling the "errors in variables" problem, and is also sometimes used even when the covariates are assumed to be error-free.

3.2 Maximum-likelihood estimation and related techniques

- **Maximum likelihood estimation** can be performed when the distribution of the error terms is known to belong to a certain parametric family $f\theta$ of probability distributions.[18] When $f\theta$ is a normal distribution with zero mean and variance θ, the resulting estimate is identical to the OLS estimate. GLS estimates are maximum likelihood estimates when ε follows a multivariate normal distribution with a known covariance matrix.

- **Ridge regression**,[19][20][21] and other forms of penalized estimation such as **Lasso regression**,[5] deliberately introduce bias into the estimation of β in order to reduce the variability of the estimate. The resulting estimators generally have lower mean squared error than the OLS estimates, particularly when multicollinearity is present or when overfitting is a problem. They are generally used when the goal is to predict the value of the response variable y for values of the predictors x that have not yet been observed. These methods are not as commonly used when the goal is inference, since it is difficult to account for the bias.

- **Least absolute deviation** (LAD) regression is a robust estimation technique in that it is less sensitive to the presence of outliers than OLS (but is less efficient than OLS when no outliers are present). It is equivalent to maximum likelihood estimation under a Laplace distribution model for ε.[22]

- **Adaptive estimation**. If we assume that error terms are independent from the regressors $\varepsilon_i \perp \mathbf{x}_i$, the

optimal estimator is the 2-step MLE, where the first step is used to non-parametrically estimate the distribution of the error term.[23]

3.3   Other estimation techniques

- **Bayesian linear regression** applies the framework of Bayesian statistics to linear regression. (See also Bayesian multivariate linear regression.) In particular, the regression coefficients β are assumed to be random variables with a specified prior distribution. The prior distribution can bias the solutions for the regression coefficients, in a way similar to (but more general than) ridge regression or lasso regression. In addition, the Bayesian estimation process produces not a single point estimate for the "best" values of the regression coefficients but an entire posterior distribution, completely describing the uncertainty surrounding the quantity. This can be used to estimate the "best" coefficients using the mean, mode, median, any quantile (see quantile regression), or any other function of the posterior distribution.

- **Quantile regression** focuses on the conditional quantiles of y given X rather than the conditional mean of y given X. Linear quantile regression models a particular conditional quantile, for example the conditional median, as a linear function $\beta^T x$ of the predictors.

- **Mixed models** are widely used to analyze linear regression relationships involving dependent data when the dependencies have a known structure. Common applications of mixed models include analysis of data involving repeated measurements, such as longitudinal data, or data obtained from cluster sampling. They are generally fit as parametric models, using maximum likelihood or Bayesian estimation. In the case where the errors are modeled as normal random variables, there is a close connection between mixed models and generalized least squares.[24] Fixed effects estimation is an alternative approach to analyzing this type of data.

- **Principal component regression** (PCR)[7][8] is used when the number of predictor variables is large, or when strong correlations exist among the predictor variables. This two-stage procedure first reduces the predictor variables using principal component analysis then uses the reduced variables in an OLS regression fit. While it often works well in practice, there is no general theoretical reason that the most informative linear function of the predictor variables should lie among the dominant principal components of the multivariate distribution of the predictor variables. The partial least squares regression is the extension of the PCR method which does not suffer from the mentioned deficiency.

- **Least-angle regression**[6] is an estimation procedure for linear regression models that was developed to handle high-dimensional covariate vectors, potentially with more covariates than observations.

- The **Theil–Sen estimator** is a simple robust estimation technique that chooses the slope of the fit line to be the median of the slopes of the lines through pairs of sample points. It has similar statistical efficiency properties to simple linear regression but is much less sensitive to outliers.[25]

- Other robust estimation techniques, including the **α-trimmed mean** approach, and **L-, M-, S-, and R-estimators** have been introduced.

3.4   Further discussion

In statistics and numerical analysis, the problem of **numerical methods for linear least squares** is an important one because linear regression models are one of the most important types of model, both as formal statistical models and for exploration of data sets. The majority of statistical computer packages contain facilities for regression analysis that make use of linear least squares computations. Hence it is appropriate that considerable effort has been devoted to the task of ensuring that these computations are undertaken efficiently and with due regard to numerical precision.

Individual statistical analyses are seldom undertaken in isolation, but rather are part of a sequence of investigatory steps. Some of the topics involved in considering numerical methods for linear least squares relate to this point. Thus important topics can be

- Computations where a number of similar, and often nested, models are considered for the same data set. That is, where models with the same dependent variable but different sets of independent variables are to be considered, for essentially the same set of data points.

- Computations for analyses that occur in a sequence, as the number of data points increases.

- Special considerations for very extensive data sets.

Fitting of linear models by least squares often, but not always, arises in the context of statistical analysis. It can therefore be important that considerations of computational efficiency for such problems extend to all of the auxiliary quantities required for such analyses, and are not restricted to the formal solution of the linear least squares problem.

Matrix calculations, like any others, are affected by rounding errors. An early summary of these effects, regarding the choice of computational methods for matrix inversion, was provided by Wilkinson.[26]

3.5 Using linear algebra

It follows that one can find a "best" approximation of another function by minimizing the area between two functions, a continuous function f on $[a, b]$ and a function $g \in W$ where W is a subspace of $C[a, b]$:

$$\text{Area} = \int_a^b |f(x) - g(x)| \ dx$$

all within the subspace W . Due to the frequent difficulty of evaluating integrands involving absolute value, one can instead define

$$\int_a^b [f(x) - g(x)]^2 \, dx$$

as an adequate criterion for obtaining the least squares approximation, function g , of f with respect to the inner product space W .

As such, $\|f - g\|^2$ or, equivalently, $\|f - g\|$, can thus be written in vector form:

$$\int_a^b [f(x) - g(x)]^2 \, dx = \langle f - g, f - g \rangle = \|f - g\|^2$$

In other words, the least squares approximation of f is the function $g \in$ subspace W closest to f in terms of the inner product $\langle f, g \rangle$. Furthermore, this can be applied with a theorem:

> Let f be continuous on $[a, b]$, and let W be a finite-dimensional subspace of $C[a, b]$. The least squares approximating function of f with respect to W is given by
>
> $$g = \langle f, \vec{w}_1 \rangle \, \vec{w}_1 + \langle f, \vec{w}_2 \rangle \, \vec{w}_2 + \cdots + \langle f, \vec{w}_n \rangle \, \vec{w}_n$$
>
> where $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ is an orthonormal basis for W .

4 Applications of linear regression

Linear regression is widely used in biological, behavioral and social sciences to describe possible relationships between variables. It ranks as one of the most important tools used in these disciplines.

4.1 Trend line

Main article: Trend estimation

A **trend line** represents a trend, the long-term movement in time series data after other components have been accounted for. It tells whether a particular data set (say GDP, oil prices or stock prices) have increased or decreased over the period of time. A trend line could simply be drawn by eye through a set of data points, but more properly their position and slope is calculated using statistical techniques like linear regression. Trend lines typically are straight lines, although some variations use higher degree polynomials depending on the degree of curvature desired in the line.

Trend lines are sometimes used in business analytics to show changes in data over time. This has the advantage of being simple. Trend lines are often used to argue that a particular action or event (such as training, or an advertising campaign) caused observed changes at a point in time. This is a simple technique, and does not require a control group, experimental design, or a sophisticated analysis technique. However, it suffers from a lack of scientific validity in cases where other potential changes can affect the data.

4.2 Epidemiology

Early evidence relating tobacco smoking to mortality and morbidity came from observational studies employing regression analysis. In order to reduce spurious correlations when analyzing observational data, researchers usually include several variables in their regression models in addition to the variable of primary interest. For example, suppose we have a regression model in which cigarette smoking is the independent variable of interest, and the dependent variable is lifespan measured in years. Researchers might include socio-economic status as an additional independent variable, to ensure that any observed effect of smoking on lifespan is not due to some effect of education or income. However, it is never possible to include all possible confounding variables in an empirical analysis. For example, a hypothetical gene might increase mortality and also cause people to smoke more. For this reason, randomized controlled trials are often able to generate more compelling evidence of causal relationships than can be obtained using regression analyses of observational data. When controlled experiments are not feasible, variants of regression analysis such as instrumental variables regression may be used to attempt to estimate causal relationships from observational data.

4.3 Finance

The capital asset pricing model uses linear regression as well as the concept of beta for analyzing and quantifying the systematic risk of an investment. This comes directly from the beta coefficient of the linear regression model that relates the return on the investment to the return on all risky assets.

4.4   Economics

Main article: Econometrics

Linear regression is the predominant empirical tool in economics. For example, it is used to predict consumption spending,[27] fixed investment spending, inventory investment, purchases of a country's exports,[28] spending on imports,[28] the demand to hold liquid assets,[29] labor demand,[30] and labor supply.[30]

4.5   Environmental science

Linear regression finds application in a wide range of environmental science applications. In Canada, the Environmental Effects Monitoring Program uses statistical analyses on fish and benthic surveys to measure the effects of pulp mill or metal mine effluent on the aquatic ecosystem.[31]

5   See also

- Analysis of variance
- Censored regression model
- Cross-sectional regression
- Curve fitting
- Empirical Bayes methods
- Errors and residuals
- Lack-of-fit sum of squares
- Line fitting
- Linear classifier
- Linear equation
- Logistic regression
- M-estimator
- MLPACK contains a C++ implementation of linear regression
- Multivariate adaptive regression splines
- Nonlinear regression
- Nonparametric regression
- Normal equations
- Projection pursuit regression
- Segmented linear regression
- Stepwise regression
- Support vector machine
- Truncated regression model

6   Notes

[1] David A. Freedman (2009). *Statistical Models: Theory and Practice*. Cambridge University Press. p. 26. A simple regression equation has on the right hand side an intercept and an explanatory variable with a slope coefficient. A multiple regression equation has two or more explanatory variables on the right hand side, each with its own slope coefficient

[2] Rencher, Alvin C.; Christensen, William F. (2012), "Chapter 10, Multivariate regression – Section 10.1, Introduction", *Methods of Multivariate Analysis*, Wiley Series in Probability and Statistics, **709** (3rd ed.), John Wiley & Sons, p. 19, ISBN 9781118391679.

[3] Hilary L. Seal (1967). "The historical development of the Gauss linear model". *Biometrika*. **54** (1/2): 1–24. doi:10.1093/biomet/54.1-2.1.

[4] Yan, Xin (2009), *Linear Regression Analysis: Theory and Computing*, World Scientific, pp. 1–2, ISBN 9789812834119, Regression analysis ... is probably one of the oldest topics in mathematical statistics dating back to about two hundred years ago. The earliest form of the linear regression was the least squares method, which was published by Legendre in 1805, and by Gauss in 1809 ... Legendre and Gauss both applied the method to the problem of determining, from astronomical observations, the orbits of bodies about the sun.

[5] Tibshirani, Robert (1996). "Regression Shrinkage and Selection via the Lasso". *Journal of the Royal Statistical Society, Series B*. **58** (1): 267–288. JSTOR 2346178.

[6] Efron, Bradley; Hastie, Trevor; Johnstone,Iain; Tibshirani,Robert (2004). "Least Angle Regression". *The Annals of Statistics*. **32** (2): 407–451. doi:10.1214/009053604000000067. JSTOR 3448465.

[7] Hawkins, Douglas M. (1973). "On the Investigation of Alternative Regressions by Principal Component Analysis". *Journal of the Royal Statistical Society, Series C*. **22** (3): 275–286. JSTOR 2346776.

[8] Jolliffe, Ian T. (1982). "A Note on the Use of Principal Components in Regression". *Journal of the Royal Statistical Society, Series C*. **31** (3): 300–303. JSTOR 2348005.

[9] Berk, Richard A. *Regression Analysis: A Constructive Critique*. Sage. doi:10.1177/0734016807304871.

[10] Warne, R. T. (2011). Beyond multiple regression: Using commonality analysis to better understand R2 results. *Gifted Child Quarterly, 55*, 313-318. doi:10.1177/0016986211422217

[11] Brillinger, David R. (1977). "The Identification of a Particular Nonlinear Time Series System". *Biometrika*. **64** (3): 509–515. doi:10.1093/biomet/64.3.509. JSTOR 2345326.

[12] Lai, T.L.; Robbins, H.; Wei, C.Z. (1978). "Strong consistency of least squares estimates in multiple regression". *PNAS*. **75** (7): 3034–3036. Bibcode:1978PNAS...75.3034L. doi:10.1073/pnas.75.7.3034. JSTOR 68164.

[13] Tofallis, C (2009). "Least Squares Percentage Regression". *Journal of Modern Applied Statistical Methods.* **7**: 526–534. doi:10.2139/ssrn.1406472.

[14] del Pino, Guido (1989). "The Unifying Role of Iterative Generalized Least Squares in Statistical Algorithms". *Statistical Science.* **4** (4): 394–403. doi:10.1214/ss/1177012408. JSTOR 2245853.

[15] Carroll, Raymond J. (1982). "Adapting for Heteroscedasticity in Linear Models". *The Annals of Statistics.* **10** (4): 1224–1233. doi:10.1214/aos/1176345987. JSTOR 2240725.

[16] Cohen, Michael; Dalal, Siddhartha R.; Tukey,John W. (1993). "Robust, Smoothly Heterogeneous Variance Regression". *Journal of the Royal Statistical Society, Series C.* **42** (2): 339–353. JSTOR 2986237.

[17] Nievergelt, Yves (1994). "Total Least Squares: State-of-the-Art Regression in Numerical Analysis". *SIAM Review.* **36** (2): 258–264. doi:10.1137/1036055. JSTOR 2132463.

[18] Lange, Kenneth L.; Little, Roderick J. A.; Taylor,Jeremy M. G. (1989). "Robust Statistical Modeling Using the t Distribution". *Journal of the American Statistical Association.* **84** (408): 881–896. doi:10.2307/2290063. JSTOR 2290063.

[19] Swindel, Benee F. (1981). "Geometry of Ridge Regression Illustrated". *The American Statistician.* **35** (1): 12–15. doi:10.2307/2683577. JSTOR 2683577.

[20] Draper, Norman R.; van Nostrand; R. Craig (1979). "Ridge Regression and James-Stein Estimation: Review and Comments". *Technometrics.* **21** (4): 451–466. doi:10.2307/1268284. JSTOR 1268284.

[21] Hoerl, Arthur E.; Kennard,Robert W.; Hoerl,Roger W. (1985). "Practical Use of Ridge Regression: A Challenge Met". *Journal of the Royal Statistical Society, Series C.* **34** (2): 114–120. JSTOR 2347363.

[22] Narula, Subhash C.; Wellington, John F. (1982). "The Minimum Sum of Absolute Errors Regression: A State of the Art Survey". *International Statistical Review.* **50** (3): 317–326. doi:10.2307/1402501. JSTOR 1402501.

[23] Stone, C. J. (1975). "Adaptive maximum likelihood estimators of a location parameter". *The Annals of Statistics.* **3** (2): 267–284. doi:10.1214/aos/1176343056. JSTOR 2958945.

[24] Goldstein, H. (1986). "Multilevel Mixed Linear Model Analysis Using Iterative Generalized Least Squares". *Biometrika.* **73** (1): 43–56. doi:10.1093/biomet/73.1.43. JSTOR 2336270.

[25] Theil, H. (1950). "A rank-invariant method of linear and polynomial regression analysis. I, II, III". *Nederl. Akad. Wetensch., Proc.* **53**: 386–392, 521–525, 1397–1412. MR 0036489; Sen, Pranab Kumar (1968). "Estimates of the regression coefficient based on Kendall's tau". *Journal of the American Statistical Association.* **63** (324): 1379–1389. doi:10.2307/2285891. JSTOR 2285891. MR 0258201.

[26] Wilkinson, J.H. (1963) "Chapter 3: Matrix Computations", *Rounding Errors in Algebraic Processes*, London: Her Majesty's Stationery Office (National Physical Laboratory, Notes in Applied Science, No.32)

[27] Deaton, Angus (1992). *Understanding Consumption.* Oxford University Press. ISBN 0-19-828824-7.

[28] Krugman, Paul R.; Obstfeld, M.; Melitz, Marc J. (2012). *International Economics: Theory and Policy* (9th global ed.). Harlow: Pearson. ISBN 9780273754091.

[29] Laidler, David E. W. (1993). "The Demand for Money: Theories, Evidence, and Problems" (4th ed.). New York: Harper Collins. ISBN 0065010981.

[30] Ehrenberg; Smith (2008). *Modern Labor Economics* (10th international ed.). London: Addison-Wesley. ISBN 9780321538963.

[31] EEMP webpage

7 References

- Cohen, J., Cohen P., West, S.G., & Aiken, L.S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences.* (2nd ed.) Hillsdale, NJ: Lawrence Erlbaum Associates

- Charles Darwin. *The Variation of Animals and Plants under Domestication.* (1868) *(Chapter XIII describes what was known about reversion in Galton's time. Darwin uses the term "reversion".)*

- Draper, N.R.; Smith, H. (1998). *Applied Regression Analysis* (3rd ed.). John Wiley. ISBN 0-471-17082-8.

- Francis Galton. "Regression Towards Mediocrity in Hereditary Stature," *Journal of the Anthropological Institute*, 15:246-263 (1886). *(Facsimile at:)*

- Robert S. Pindyck and Daniel L. Rubinfeld (1998, 4h ed.). *Econometric Models and Economic Forecasts*, ch. 1 (Intro, incl. appendices on Σ operators & derivation of parameter est.) & Appendix 4.3 (mult. regression in matrix form).

8 Further reading

- Barlow, Jesse L. (1993). "Chapter 9: Numerical aspects of Solving Linear Least Squares Problems". In Rao, C.R. *Computational Statistics*. Handbook of Statistics. **9**. North-Holland. ISBN 0-444-88096-8

- Björck, Åke (1996). *Numerical methods for least squares problems*. Philadelphia: SIAM. ISBN 0-89871-360-9.

- Goodall, Colin R. (1993). "Chapter 13: Computation using the QR decomposition". In Rao, C.R. *Computational Statistics*. Handbook of Statistics. **9**. North-Holland. ISBN 0-444-88096-8

- Pedhazur, Elazar J (1982). "Multiple regression in behavioral research: Explanation and prediction" (2nd ed.). New York: Holt, Rinehart and Winston. ISBN 0-03-041760-0.

- National Physical Laboratory (1961). "Chapter 1: Linear Equations and Matrices: Direct Methods". *Modern Computing Methods*. Notes on Applied Science. **16** (2nd ed.). Her Majesty's Stationery Office

- National Physical Laboratory (1961). "Chapter 2: Linear Equations and Matrices: Direct Methods on Automatic Computers". *Modern Computing Methods*. Notes on Applied Science. **16** (2nd ed.). Her Majesty's Stationery Office

9 Text and image sources, contributors, and licenses

9.1 Text

- **Linear regression** *Source:* https://en.wikipedia.org/wiki/Linear_regression?oldid=733900248 *Contributors:* The Anome, Taw, Ap, Larry Sanger, Danny, Miguel~enwiki, Rade Kutil, Edward, Patrick, Michael Hardy, GABaker, Shyamal, Kku, Dcljr, Tomi, TakuyaMurata, Den fjättrade ankan~enwiki, Kevin Baas, Rossami, Hike395, Jitse Niesen, Andrewman327, Taxman, Donarreiskoffer, Robbot, Jaredwf, Benwing, Gak, ZimZalaBim, Yelyos, Babbage, Henrygb, Jcoleff, Wile E. Heresiarch, Giftlite, BenFrantzDale, Fleminra, Alison, Duncharris, Jason Quinn, Ato, Utcursch, Pgan002, MarkSweep, Piotrus, Wurblzap~enwiki, Icairns, Urhixidur, Natrij, Discospinster, Rich Farmbrough, Pak21, Paul August, Bender235, Violetriga, Elwikipedista~enwiki, Gauge, MisterSheik, Spoon!, Perfecto, O18, Davidswelt, R. S. Shaw, Tobacman, Arcadian, NickSchweitzer, 99of9, Crust, Landroni, Storm Rider, Musiphil, Arthena, ABCD, Kotasik, Avenue, Snowolf, LFaraone, Forderud, Drummond, Oleg Alexandrov, Abanima, Tappancsa, Mindmatrix, BlaiseFEgan, Btyner, Joerg Kurt Wegner, Lacurus~enwiki, Graham87, Qwertyus, Rjwilmsi, Vegaswikian, Matt Deres, TeaDrinker, Chobot, Manscher, FrankTobia, Wavelength, RussBot, Gaius Cornelius, Bug42, Afelton, Thiseye, Cruise, Moe Epsilon, Voidxor, Dggoldst, Arch o median, Arthur Rubin, Drallim, Caballero1967, Anarch21, SolarMcPanel, SmackBot, NickyMcLean, Quazar777, Prodego, InverseHypercube, Jtneill, DanielPenfield, Evanreyes, Commander Keane bot, Ohnoitsjamie, Hraefen, Afa86, Markush, Amatulic, Feinstein, Oli Filth, John Reaves, Berland, Wolf87, Cybercobra, Semanticprecision, G716, Unco, Theblackgecko, Lambiam, Jonas August, Vjeet a, Nijdam, Beetstra, Emurph, Hu12, Pjrm, LAlawMedMBA, JoeBot, Chris53516, AlainD, Jsorens, Tawkerbot2, CmdrObot, JRavn, CBM, Anakata, Chrike, Harej bot, Thomasmeeks, Neelix, Cassmus, Bumbulski, MaxEnt, 137 0, Pedrolapinto, Mmmooonnnsssttteeerrr, Farshidforouz~enwiki, FrancoGG, Talgalili, Thijs!bot, Epbr123, Tolstoy the Cat, Jfaller, Whooooooknows, Natalie Erin, Woollymammoth, Mack2, JAnDbot, MER-C, Jeff560, Ph.eyes, Hectorlamadrid, Magioladitis, Tripbeetle, MastCell, Albmont, Baccyak4H, Ddr~enwiki, David Eppstein, Joostw, Apal~enwiki, Yonaa, R'n'B, Noyder, Charlesmartin14, Kawautar, Mbhiii, J.delanoy, Scythe of Death, Salih, TomyDuby, Jaxha, HyDeckar, Jewzip, Mr-Paul84, Copsi, Llorenzi, VolkovBot, Smarty07, Muzzamo, Mfreund~enwiki, Jsd115, Zhenqinli, Greswik, P1h3r1e3d13, Ricardo MC, Jhedengren, Petergans, Karthik Sarma, Rlendog, Zsniew, Quest for Truth, Saginer, Paolo.dL, OKBot, Water and Land, Scottyoak2, Melcombe, Gpap.gpap, Tanvir Ahmmed, ClueBot, HairyFotr, Cp111, Rhubbarb, Dromedario~enwiki, Alexbot, Ecov, Kaspar.jan, Tokorode~enwiki, Skbkekas, Stephen Milborrow, Diaa abdelmoneim, Qwfp, Bigoperm, Sunsetsky, XLinkBot, Tofallis, W82~enwiki, Tayste, Addbot, RPHv, Fgnievinski, Doronp, MrOllie, Download, Forich, Zorrobot, Ettrig, Luckas-bot, Yobot, Sked123, It's Been Emotional, AnomieBOT, Rubinbot, IRP, Tucoxn, Materialscientist, HanPritcher, Citation bot, Lixiaoxu, Sketchmoose, Flavio Guitian, Gtfjbl, Istrill, Mstangeland, Aa77zz, Imran.fanaswala, Fstonedahl, PhysicsJoe, Nickruiz, FrescoBot, X7q, Citation bot 1, AstaBOTh15, Boxplot, Pinethicket, Elockid, Kiefer.Wolfowitz, Rdecker02, Jonesey95, Stpasha, Oldrrb, Trappist the monk, Wotnow, Duoduoduo, PAC2, Wombathammer, Diannaa, RjwilmsiBot, Elitropia, John of Reading, Dewritech, Sugarfoot1001, Julienbarlan, Wikieconometrician, Bkearb, NGPriest, Bartlebythe-Scrivener, Chewings72, Esaintpierre, Manipande, Haresfur, ClueBot NG, Mathstat, Frietjes, BlueScreenD, Helpful Pixie Bot, Grandwgy, Daonng, Mark Arsten, MyWikiNik, ChrisGualtieri, Illia Connell, Dansbecker, Jerry Hintze, Dexbot, Hkoslik, Sa publishers, Ossifragus, Bha100710, PeterLFlomPhD, Prajwel, Drvikas74, Melonkelon, Asif usa, Pandadai, Bryanrutherford0, Tertius51, Logan.dunbar, Monkbot, Jpeterson1346, Bob nau, JC713, Moorshed, Velvel2, Whatfoxsays, 18trevor3695, Split97, HelpUsStopSpam, T9520, Marianna251 and Anonymous: 383

9.2 Images

- **File:Anscombe'{}s_quartet_3.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/e/ec/Anscombe%27s_quartet_3.svg *License:* CC BY-SA 3.0 *Contributors:*

- Anscombe.svg *Original artist:* Anscombe.svg: Schutz

- **File:Commons-logo.svg** *Source:* https://upload.wikimedia.org/wikipedia/en/4/4a/Commons-logo.svg *License:* CC-BY-SA-3.0 *Contributors:* ? *Original artist:* ?

- **File:Fisher_iris_versicolor_sepalwidth.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/4/40/Fisher_iris_versicolor_sepalwidth.svg *License:* CC BY-SA 3.0 *Contributors:* en:Image:Fisher iris versicolor sepalwidth.png *Original artist:* en:User:Qwfp (original); Pbroks13 (talk) (redraw)

- **File:Folder_Hexagonal_Icon.svg** *Source:* https://upload.wikimedia.org/wikipedia/en/4/48/Folder_Hexagonal_Icon.svg *License:* Cc-by-sa-3.0 *Contributors:* ? *Original artist:* ?

- **File:Linear_regression.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/3/3a/Linear_regression.svg *License:* Public domain *Contributors:* Own work *Original artist:* Sewaqu

- **File:People_icon.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/3/37/People_icon.svg *License:* CC0 *Contributors:* OpenClipart *Original artist:* OpenClipart

- **File:Polyreg_scheffe.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/8/8b/Polyreg_scheffe.svg *License:* CC BY 3.0 *Contributors:* Own work *Original artist:* Skbkekas

- **File:Portal-puzzle.svg** *Source:* https://upload.wikimedia.org/wikipedia/en/f/fd/Portal-puzzle.svg *License:* Public domain *Contributors:* ? *Original artist:* ?

- **File:Thiel-Sen_estimator.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/e/e9/Thiel-Sen_estimator.svg *License:* CC0 *Contributors:* Own work *Original artist:* David Eppstein

- **File:Wiki_letter_w_cropped.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/1/1c/Wiki_letter_w_cropped.svg *License:* CC-BY-SA-3.0 *Contributors:* This file was derived from Wiki letter w.svg: *Original artist:* Derivative work by Thumperward

- **File:Wikibooks-logo-en-noslogan.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/d/df/Wikibooks-logo-en-noslogan.svg *License:* CC BY-SA 3.0 *Contributors:* Own work *Original artist:* User:Bastique, User:Ramac et al.

9.3   Content license