Linear regression

From Wikipedia, the free encyclopedia

|  |
| --- |
| Part of a series on [Statistics](https://en.wikipedia.org/wiki/Statistics) |
| [**Regression analysis**](https://en.wikipedia.org/wiki/Regression_analysis) |
| [Linear regression.svg](https://en.wikipedia.org/wiki/File:Linear_regression.svg) |
| **Models** |
| * + **Linear regression**   + [Simple regression](https://en.wikipedia.org/wiki/Simple_linear_regression)   + [Ordinary least squares](https://en.wikipedia.org/wiki/Ordinary_least_squares)   + [Polynomial regression](https://en.wikipedia.org/wiki/Polynomial_regression)   + [General linear model](https://en.wikipedia.org/wiki/General_linear_model) |
| * + [Generalized linear model](https://en.wikipedia.org/wiki/Generalized_linear_model)   + [Discrete choice](https://en.wikipedia.org/wiki/Discrete_choice)   + [Logistic regression](https://en.wikipedia.org/wiki/Logistic_regression)   + [Multinomial logit](https://en.wikipedia.org/wiki/Multinomial_logit)   + [Mixed logit](https://en.wikipedia.org/wiki/Mixed_logit)   + [Probit](https://en.wikipedia.org/wiki/Probit_model)   + [Multinomial probit](https://en.wikipedia.org/wiki/Multinomial_probit)   + [Ordered logit](https://en.wikipedia.org/wiki/Ordered_logit)   + [Ordered probit](https://en.wikipedia.org/wiki/Ordered_probit)   + [Poisson](https://en.wikipedia.org/wiki/Poisson_regression) |
| * + [Multilevel model](https://en.wikipedia.org/wiki/Multilevel_model)   + [Fixed effects](https://en.wikipedia.org/wiki/Fixed_effects_model)   + [Random effects](https://en.wikipedia.org/wiki/Random_effects_model)   + [Mixed model](https://en.wikipedia.org/wiki/Mixed_model) |
| * + [Nonlinear regression](https://en.wikipedia.org/wiki/Nonlinear_regression)   + [Nonparametric](https://en.wikipedia.org/wiki/Nonparametric_regression)   + [Semiparametric](https://en.wikipedia.org/wiki/Semiparametric_regression)   + [Robust](https://en.wikipedia.org/wiki/Robust_regression)   + [Quantile](https://en.wikipedia.org/wiki/Quantile_regression)   + [Isotonic](https://en.wikipedia.org/wiki/Isotonic_regression)   + [Principal components](https://en.wikipedia.org/wiki/Principal_component_regression)   + [Least angle](https://en.wikipedia.org/wiki/Least-angle_regression)   + [Local](https://en.wikipedia.org/wiki/Local_regression)   + [Segmented](https://en.wikipedia.org/wiki/Segmented_regression) |
| * + [Errors-in-variables](https://en.wikipedia.org/wiki/Errors-in-variables_models) |
| **Estimation** |
| * + [Least squares](https://en.wikipedia.org/wiki/Least_squares)   + [Ordinary least squares](https://en.wikipedia.org/wiki/Ordinary_least_squares)   + [Linear (math)](https://en.wikipedia.org/wiki/Linear_least_squares_(mathematics))   + [Partial](https://en.wikipedia.org/wiki/Partial_least_squares_regression)   + [Total](https://en.wikipedia.org/wiki/Total_least_squares)   + [Generalized](https://en.wikipedia.org/wiki/Generalized_least_squares)   + [Weighted](https://en.wikipedia.org/wiki/Weighted_least_squares)   + [Non-linear](https://en.wikipedia.org/wiki/Non-linear_least_squares)   + [Non-negative](https://en.wikipedia.org/wiki/Non-negative_least_squares)   + [Iteratively reweighted](https://en.wikipedia.org/wiki/Iteratively_reweighted_least_squares)   + [Ridge regression](https://en.wikipedia.org/wiki/Tikhonov_regularization) |
| * + [Least absolute deviations](https://en.wikipedia.org/wiki/Least_absolute_deviations)   + [Bayesian](https://en.wikipedia.org/wiki/Bayesian_linear_regression)   + [Bayesian multivariate](https://en.wikipedia.org/wiki/Bayesian_multivariate_linear_regression) |
| **Background** |
| * + [Regression model validation](https://en.wikipedia.org/wiki/Regression_model_validation)   + [Mean and predicted response](https://en.wikipedia.org/wiki/Mean_and_predicted_response)   + [Errors and residuals](https://en.wikipedia.org/wiki/Errors_and_residuals_in_statistics)   + [Goodness of fit](https://en.wikipedia.org/wiki/Goodness_of_fit)   + [Studentized residual](https://en.wikipedia.org/wiki/Studentized_residual)   + [Gauss–Markov theorem](https://en.wikipedia.org/wiki/Gauss%E2%80%93Markov_theorem) |
| * + [Fisher iris versicolor sepalwidth.svg](https://en.wikipedia.org/wiki/File:Fisher_iris_versicolor_sepalwidth.svg)   [**Statistics portal**](https://en.wikipedia.org/wiki/Portal:Statistics) |
| * + [v](https://en.wikipedia.org/wiki/Template:Regression_bar)   + [t](https://en.wikipedia.org/wiki/Template_talk:Regression_bar)   + [e](https://en.wikipedia.org/w/index.php?title=Template:Regression_bar&action=edit) |

In [statistics](https://en.wikipedia.org/wiki/Statistics), **linear regression** is a [linear](https://en.wikipedia.org/wiki/Linear) approach for modelling the relationship between a scalar [dependent variable](https://en.wikipedia.org/wiki/Dependent_variable) *y* and one or more [explanatory variables](https://en.wikipedia.org/wiki/Explanatory_variable) (or independent variables) denoted *X*. The case of one explanatory variable is called [*simple linear regression*](https://en.wikipedia.org/wiki/Simple_linear_regression). For more than one explanatory variable, the process is called *multiple linear regression*.[[1]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-Freedman09-1) (This term is distinct from [*multivariate linear regression*](https://en.wikipedia.org/wiki/Multivariate_linear_regression), where multiple correlated dependent variables are predicted, rather than a single scalar variable.)[[2]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-2)

In linear regression, the relationships are modeled using [linear predictor functions](https://en.wikipedia.org/wiki/Linear_predictor_function) whose unknown model [parameters](https://en.wikipedia.org/wiki/Parameters) are [estimated](https://en.wikipedia.org/wiki/Estimation_theory) from the [data](https://en.wikipedia.org/wiki/Data). Such models are called [*linear models*](https://en.wikipedia.org/wiki/Linear_model).[[3]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-3) Most commonly, the [conditional mean](https://en.wikipedia.org/wiki/Conditional_expectation) of *y* given the value of *X* is assumed to be an [affine function](https://en.wikipedia.org/wiki/Affine_transformation) of *X*; less commonly, the [median](https://en.wikipedia.org/wiki/Median) or some other [quantile](https://en.wikipedia.org/wiki/Quantile) of the conditional distribution of *y* given *X* is expressed as a linear function of *X*. Like all forms of [regression analysis](https://en.wikipedia.org/wiki/Regression_analysis), linear regression focuses on the [conditional probability distribution](https://en.wikipedia.org/wiki/Conditional_probability_distribution) of *y* given *X*, rather than on the [joint probability distribution](https://en.wikipedia.org/wiki/Joint_probability_distribution) of *y* and *X*, which is the domain of [multivariate analysis](https://en.wikipedia.org/wiki/Multivariate_analysis).

Linear regression was the first type of regression analysis to be studied rigorously, and to be used extensively in practical applications.[[4]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-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*.

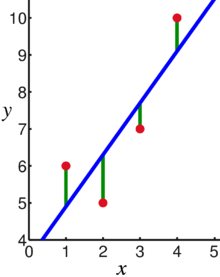
Linear regression models are often fitted using the [least squares](https://en.wikipedia.org/wiki/Least_squares) approach, but they may also be fitted in other ways, such as by minimizing the "lack of fit" in some other [norm](https://en.wikipedia.org/wiki/Norm_(mathematics)) (as with [least absolute deviations](https://en.wikipedia.org/wiki/Least_absolute_deviations) regression), or by minimizing a penalized version of the least squares [loss function](https://en.wikipedia.org/wiki/Loss_function) as in [ridge regression](https://en.wikipedia.org/wiki/Ridge_regression) (*L*2-norm penalty) and [lasso](https://en.wikipedia.org/wiki/Lasso_(statistics)) (*L*1-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.

**Contents**

  [hide]

* [1Introduction](https://en.wikipedia.org/wiki/Linear_regression#Introduction)
* [1.1Assumptions](https://en.wikipedia.org/wiki/Linear_regression#Assumptions)
* [1.2Interpretation](https://en.wikipedia.org/wiki/Linear_regression#Interpretation)
* [2Extensions](https://en.wikipedia.org/wiki/Linear_regression#Extensions)
* [2.1Simple and multiple linear regression](https://en.wikipedia.org/wiki/Linear_regression#Simple_and_multiple_linear_regression)
* [2.2General linear models](https://en.wikipedia.org/wiki/Linear_regression#General_linear_models)
* [2.3Heteroscedastic models](https://en.wikipedia.org/wiki/Linear_regression#Heteroscedastic_models)
* [2.4Generalized linear models](https://en.wikipedia.org/wiki/Linear_regression#Generalized_linear_models)
* [2.5Hierarchical linear models](https://en.wikipedia.org/wiki/Linear_regression#Hierarchical_linear_models)
* [2.6Errors-in-variables](https://en.wikipedia.org/wiki/Linear_regression#Errors-in-variables)
* [2.7Others](https://en.wikipedia.org/wiki/Linear_regression#Others)
* [3Estimation methods](https://en.wikipedia.org/wiki/Linear_regression#Estimation_methods)
* [3.1Least-squares estimation and related techniques](https://en.wikipedia.org/wiki/Linear_regression#Least-squares_estimation_and_related_techniques)
* [3.2Maximum-likelihood estimation and related techniques](https://en.wikipedia.org/wiki/Linear_regression#Maximum-likelihood_estimation_and_related_techniques)
* [3.3Other estimation techniques](https://en.wikipedia.org/wiki/Linear_regression#Other_estimation_techniques)
* [3.4Further discussion](https://en.wikipedia.org/wiki/Linear_regression#Further_discussion)
* [3.5Using linear algebra](https://en.wikipedia.org/wiki/Linear_regression#Using_linear_algebra)
* [4Applications of linear regression](https://en.wikipedia.org/wiki/Linear_regression#Applications_of_linear_regression)
* [4.1Trend line](https://en.wikipedia.org/wiki/Linear_regression#Trend_line)
* [4.2Epidemiology](https://en.wikipedia.org/wiki/Linear_regression#Epidemiology)
* [4.3Finance](https://en.wikipedia.org/wiki/Linear_regression#Finance)
* [4.4Economics](https://en.wikipedia.org/wiki/Linear_regression#Economics)
* [4.5Environmental science](https://en.wikipedia.org/wiki/Linear_regression#Environmental_science)
* [5See also](https://en.wikipedia.org/wiki/Linear_regression#See_also)
* [6Notes](https://en.wikipedia.org/wiki/Linear_regression#Notes)
* [7References](https://en.wikipedia.org/wiki/Linear_regression#References)
* [8Further reading](https://en.wikipedia.org/wiki/Linear_regression#Further_reading)
* [9External links](https://en.wikipedia.org/wiki/Linear_regression#External_links)

Introduction[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=1)]

[](https://en.wikipedia.org/wiki/File:Linear_least_squares_example2.png)

In linear regression, the observations (red) are assumed to be the result of random deviations (green) from an underlying relationship (blue) between the dependent variable (*y*) and independent variable (*x*).

Given a [data](https://en.wikipedia.org/wiki/Data) set

 of *n* [statistical units](https://en.wikipedia.org/wiki/Statistical_unit), a linear regression model assumes that the relationship between the dependent variable *yi* and the [*p*-vector](https://en.wikipedia.org/wiki/Multivector) of regressors ***x****i* is [linear](https://en.wikipedia.org/wiki/Linear_function). This relationship is modeled through a *disturbance term* or *error variable* *εi* — an unobserved [random variable](https://en.wikipedia.org/wiki/Random_variable) that adds noise to the linear relationship between the dependent variable and regressors. Thus the model takes the form

where T denotes the [transpose](https://en.wikipedia.org/wiki/Transpose), so that ***x****i*T***β*** is the [inner product](https://en.wikipedia.org/wiki/Inner_product) between [vectors](https://en.wikipedia.org/wiki/Coordinate_vector) ***x****i* and ***β***.

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

where

Some remarks on terminology and general use:

 is called the *regressand*, *endogenous variable*, *response variable*, *measured variable*, *criterion variable*, or *dependent variable* (see [dependent and independent variables](https://en.wikipedia.org/wiki/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.

 are called *regressors*, *exogenous variables*, *explanatory variables*, *covariates*, *input variables*, *predictor variables*, or *independent variables* (see [dependent and independent variables](https://en.wikipedia.org/wiki/Dependent_and_independent_variables), but not to be confused with [independent random variables](https://en.wikipedia.org/wiki/Independent_random_variables)). The matrix

 is sometimes called the [design matrix](https://en.wikipedia.org/wiki/Design_matrix).

* Usually a constant is included as one of the regressors. For example, we can take *xi*1 = 1 for *i* = 1, ..., *n*. The corresponding element of ***β*** is called the [*intercept*](https://en.wikipedia.org/wiki/Y-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](https://en.wikipedia.org/wiki/Polynomial_regression) and [segmented regression](https://en.wikipedia.org/wiki/Segmented_regression). The model remains linear as long as it is linear in the parameter vector ***β***.
* The regressors *xij* may be viewed either as [random variables](https://en.wikipedia.org/wiki/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.

 is a (*p* + 1)-dimensional *parameter vector*, where

 is the constant (offset) term. Its elements are also called *effects*, and the estimates of it are called "estimated effects" or *regression coefficients*. Statistical [estimation](https://en.wikipedia.org/wiki/Estimation_theory) and [inference](https://en.wikipedia.org/wiki/Statistical_inference) in linear regression focuses on ***β***. The elements of this parameter vector are interpreted as the [partial derivatives](https://en.wikipedia.org/wiki/Partial_derivative) of the dependent variable with respect to the various independent variables.

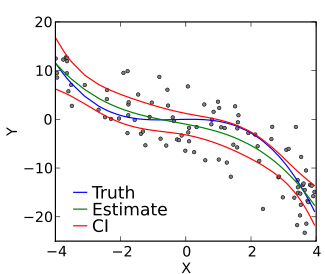
 is called the *error term*, *disturbance term*, or *noise*. This variable captures all other factors which influence the dependent variable *yi* other than the regressors ***x****i*. The relationship between the error term and the regressors, for example whether they are [correlated](https://en.wikipedia.org/wiki/Correlation), 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

where *β*1 determines the initial velocity of the ball, *β*2 is proportional to the [standard gravity](https://en.wikipedia.org/wiki/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 ***x****i* = (*xi*1, *xi*2)  = (*ti*, *ti*2), the model takes on the standard form

**Assumptions**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=2)]

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. 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.

[](https://en.wikipedia.org/wiki/File:Polyreg_scheffe.svg)

Example of a cubic polynomial regression, which is a type of linear regression.

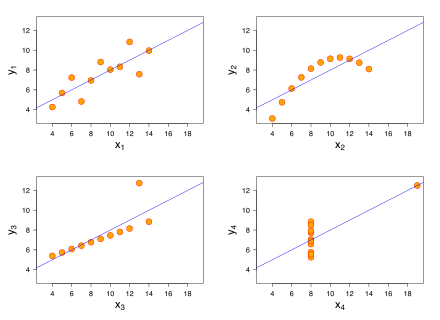
The following are the major assumptions made by standard linear regression models with standard estimation techniques (e.g. [ordinary least squares](https://en.wikipedia.org/wiki/Ordinary_least_squares)):

* **Weak exogeneity**. This essentially means that the predictor variables *x* can be treated as fixed values, rather than [random variables](https://en.wikipedia.org/wiki/Random_variable). 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](https://en.wikipedia.org/wiki/Errors-in-variables_model).
* **Linearity**. This means that the mean of the response variable is a [linear combination](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Polynomial_regression), which uses linear regression to fit the response variable as an arbitrary [polynomial](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Overfit) the data. As a result, some kind of [regularization](https://en.wikipedia.org/wiki/Regularization_(mathematics)) must typically be used to prevent unreasonable solutions coming out of the estimation process. Common examples are [ridge regression](https://en.wikipedia.org/wiki/Ridge_regression) and [lasso regression](https://en.wikipedia.org/wiki/Lasso_regression). [Bayesian linear regression](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Ridge_regression) and [lasso regression](https://en.wikipedia.org/wiki/Lasso_regression) can both be viewed as special cases of Bayesian linear regression, with particular types of [prior distributions](https://en.wikipedia.org/wiki/Prior_distribution) placed on the regression coefficients.)
* **Constant variance** (a.k.a. [**homoscedasticity**](https://en.wikipedia.org/wiki/Homoscedasticity)). This means that different values of the response variable have the same [variance](https://en.wikipedia.org/wiki/Variance) in their errors, regardless of the values of the predictor variables. In practice this assumption is invalid (i.e. the errors are [heteroscedastic](https://en.wikipedia.org/wiki/Heteroscedasticity)) if the response variable can vary over a wide scale. In order to check 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 predictive variables. Errors 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](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Weighted_least_squares) and [heteroscedasticity-consistent standard errors](https://en.wikipedia.org/wiki/Heteroscedasticity-consistent_standard_errors)) can handle heteroscedasticity in a quite general way. [Bayesian linear regression](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Logarithm) of the response variable using a linear regression model, which implies that the response variable has a [log-normal distribution](https://en.wikipedia.org/wiki/Log-normal_distribution) rather than a [normal distribution](https://en.wikipedia.org/wiki/Normal_distribution)).
* **Independence** of errors. This assumes that the errors of the response variables are uncorrelated with each other. (Actual [statistical independence](https://en.wikipedia.org/wiki/Independence_(probability_theory)) 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](https://en.wikipedia.org/wiki/Generalized_least_squares)) are capable of handling correlated errors, although they typically require significantly more data unless some sort of [regularization](https://en.wikipedia.org/wiki/Regularization_(mathematics)) is used to bias the model towards assuming uncorrelated errors. [Bayesian linear regression](https://en.wikipedia.org/wiki/Bayesian_linear_regression) is a general way of handling this issue.
* **Lack of perfect multicollinearity** in the predictors. For standard [least squares](https://en.wikipedia.org/wiki/Least_squares) estimation methods, the design matrix *X* must have full [column rank](https://en.wikipedia.org/wiki/Column_rank) *p*; otherwise, we have a condition known as perfect [multicollinearity](https://en.wikipedia.org/wiki/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 perfect multicollinearity, the parameter vector *β* will be [non-identifiable](https://en.wikipedia.org/wiki/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 **R***p*. See [partial least squares regression](https://en.wikipedia.org/wiki/Partial_least_squares_regression). Methods for fitting linear models with multicollinearity have been developed;[[5]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-tibs_lasso-5)[[6]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-efron_lars-6)[[7]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-hawkins_pcr-7)[[8]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-joliffe_pcr-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](https://en.wikipedia.org/wiki/Generalized_linear_model), do not suffer from this problem.

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](https://en.wikipedia.org/wiki/Probability_distribution) of the predictor variables *x* has a major influence on the precision of estimates of *β*. [Sampling](https://en.wikipedia.org/wiki/Sampling_(statistics)) and [design of experiments](https://en.wikipedia.org/wiki/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*.

**Interpretation**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=3)]

[](https://en.wikipedia.org/wiki/File:Anscombe%27s_quartet_3.svg)

The data sets in the [Anscombe's quartet](https://en.wikipedia.org/wiki/Anscombe%27s_quartet) are designed to have the same linear regression line (as well as identical means, standard deviations, and correlations) but are graphically very different. This illustrates the pitfalls of relying solely on a fitted model to understand the relationship between variables.

A fitted linear regression model can be used to identify the relationship between a single predictor variable *xj* and the response variable *y* when all the other predictor variables in the model are "held fixed". Specifically, the interpretation of *βj* is the [expected](https://en.wikipedia.org/wiki/Expected_value) 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](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Pearson_correlation) or [simple linear regression](https://en.wikipedia.org/wiki/Simple_linear_regression) model relating only *xj* to *y*; this effect is the [total derivative](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Dummy_variable_(statistics)), 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*.

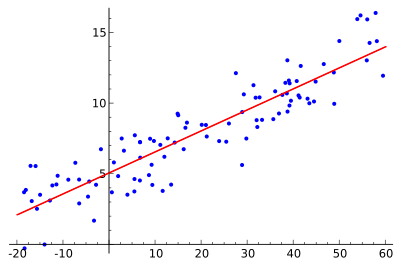
The meaning of the expression "held fixed" may depend on how the values of the predictor variables arise. If the experimenter directly sets the values of the predictor variables according to a study design, the comparisons of interest may literally correspond to comparisons among units whose predictor 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 predictor 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]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-9) A commonality analysis may be helpful in disentangling the shared and unique impacts of correlated independent variables.[[10]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-10)

Extensions[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=4)]

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

**Simple and multiple linear regression**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=5)]

[](https://en.wikipedia.org/wiki/File:Linear_regression.svg)

Example of [simple linear regression](https://en.wikipedia.org/wiki/Simple_linear_regression), which has one independent variable

The very simplest case of a single [scalar](https://en.wikipedia.org/wiki/Scalar_(mathematics)) predictor variable *x* and a single scalar response variable *y* is known as [*simple linear regression*](https://en.wikipedia.org/wiki/Simple_linear_regression). The extension to multiple and/or [vector](https://en.wikipedia.org/wiki/Euclidean_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*.

**General linear models**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=6)]

The [general linear model](https://en.wikipedia.org/wiki/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. "General linear models" are also called "multivariate linear models". These are not the same as multivariable linear models (also called "multiple linear models").

**Heteroscedastic models**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=7)]

Various models have been created that allow for [heteroscedasticity](https://en.wikipedia.org/wiki/Heteroscedasticity), i.e. the errors for different response variables may have different [variances](https://en.wikipedia.org/wiki/Variance). For example, [weighted least squares](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Linear_least_squares_(mathematics)#Weighted_linear_least_squares), and [Generalized least squares](https://en.wikipedia.org/wiki/Generalized_least_squares).) [Heteroscedasticity-consistent standard errors](https://en.wikipedia.org/wiki/Heteroscedasticity-consistent_standard_errors) is an improved method for use with uncorrelated but potentially heteroscedastic errors.

**Generalized linear models**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=8)]

[Generalized linear models](https://en.wikipedia.org/wiki/Generalized_linear_model) (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](https://en.wikipedia.org/wiki/Skewed_distribution) such as the [log-normal distribution](https://en.wikipedia.org/wiki/Log-normal_distribution) or [Poisson distribution](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Categorical_data), such as the choice of a given candidate in an election (which is better described using a [Bernoulli distribution](https://en.wikipedia.org/wiki/Bernoulli_distribution)/[binomial distribution](https://en.wikipedia.org/wiki/Binomial_distribution) for binary choices, or a [categorical distribution](https://en.wikipedia.org/wiki/Categorical_distribution)/[multinomial distribution](https://en.wikipedia.org/wiki/Multinomial_distribution) for multi-way choices), where there are a fixed number of choices that cannot be meaningfully ordered;
* when modeling [ordinal data](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Mean) of the response variable to the predictors, i.e. *E*(*y*) = *g*(*β*′*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

 range of the linear predictor and the range of the response variable.

Some common examples of GLMs are:

* [Poisson regression](https://en.wikipedia.org/wiki/Poisson_regression) for count data.
* [Logistic regression](https://en.wikipedia.org/wiki/Logistic_regression) and [probit regression](https://en.wikipedia.org/wiki/Probit_regression) for binary data.
* [Multinomial logistic regression](https://en.wikipedia.org/wiki/Multinomial_logistic_regression) and [multinomial probit](https://en.wikipedia.org/wiki/Multinomial_probit) regression for categorical data.
* [Ordered probit](https://en.wikipedia.org/wiki/Ordered_probit) regression for ordinal data.

Single index models[[*clarification needed*](https://en.wikipedia.org/wiki/Wikipedia:Please_clarify)] allow some degree of nonlinearity in the relationship between *x* and *y*, while preserving the central role of the linear predictor *β*′*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]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-11)

**Hierarchical linear models**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=9)]

[Hierarchical linear models](https://en.wikipedia.org/wiki/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.

**Errors-in-variables**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=10)]

[Errors-in-variables models](https://en.wikipedia.org/wiki/Errors-in-variables_model) (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.

**Others**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=11)]

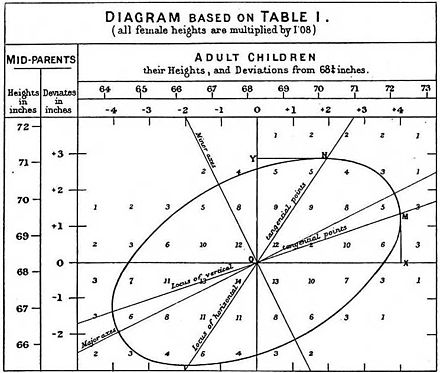
* In [Dempster–Shafer theory](https://en.wikipedia.org/wiki/Dempster%E2%80%93Shafer_theory), or a [linear belief function](https://en.wikipedia.org/wiki/Linear_belief_function) in particular, a linear regression model may be represented as a partially swept matrix, which can be combined with similar 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.

Estimation methods[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=12)]

A large number of procedures have been developed for [parameter](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Consistent_estimator) and asymptotic [efficiency](https://en.wikipedia.org/wiki/Efficiency_(statistics)).

Some of the more common estimation techniques for linear regression are summarized below.

**Least-squares estimation and related techniques**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=13)]

[](https://en.wikipedia.org/wiki/File:Galton%27s_correlation_diagram_1875.jpg)

[Francis Galton's](https://en.wikipedia.org/wiki/Francis_Galton) 1875 illustration of the correlation between the heights of adults and their parents. The observation that adult children's heights tended to deviate less from the mean height than their parents suggested the concept of "[regression toward the mean](https://en.wikipedia.org/wiki/Regression_toward_the_mean)", giving regression its name. The "locus of horizontal tangential points" passing through the leftmost and rightmost points on the ellipse (which is a [level curve](https://en.wikipedia.org/wiki/Level_curve) of the [bivariate normal distribution](https://en.wikipedia.org/wiki/Bivariate_normal_distribution) estimated from the data), is the [OLS](https://en.wikipedia.org/wiki/Ordinary_least_squares) estimate of the regression of parents' heights on children's heights, whereas the "locus of vertical tangential points" is the OLS estimate of the regression of children's heights on parent's heights. The major axis of the ellipse is the [TLS](https://en.wikipedia.org/wiki/Total_least_squares) estimate.

* [**Ordinary least squares**](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Experiment) and [observational](https://en.wikipedia.org/wiki/Observational_study) data.  
  The OLS method minimizes the sum of squared [residuals](https://en.wikipedia.org/wiki/Errors_and_residuals_in_statistics), and leads to a closed-form expression for the estimated value of the unknown parameter *β*:

The estimator is [unbiased](https://en.wikipedia.org/wiki/Bias_of_an_estimator) and [consistent](https://en.wikipedia.org/wiki/Consistent_estimator) if the errors have finite variance and are uncorrelated with the regressors[[12]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-12)

It is also [efficient](https://en.wikipedia.org/wiki/Efficiency_(statistics)) under the assumption that the errors have finite variance and are [homoscedastic](https://en.wikipedia.org/wiki/Homoscedasticity), meaning that E[*εi*2|**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 **β**. 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](https://en.wikipedia.org/wiki/Multicollinearity) is present, unless the sample size is large.  
In [simple linear regression](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Pearson_correlation_coefficient) between the covariate and the response.

* [**Generalized least squares**](https://en.wikipedia.org/wiki/Generalized_least_squares) (GLS) is an extension of the OLS method, that allows efficient estimation of *β* when either [heteroscedasticity](https://en.wikipedia.org/wiki/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 *i*th case is inversely proportional to var(*εi*). This special case of GLS is called "weighted least squares". The GLS solution to estimation problem is

where **Ω** 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.

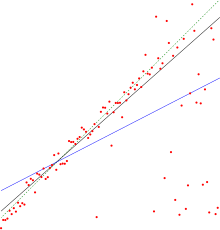
* [**Percentage least squares**](https://en.wikipedia.org/w/index.php?title=Percentage_least_squares&action=edit&redlink=1) focuses on reducing percentage errors, which is useful in the field of forecasting or time series analysis. It is also useful in situations where the dependent variable has a wide 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]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-13)
* [**Iteratively reweighted least squares**](https://en.wikipedia.org/wiki/Iteratively_reweighted_least_squares) (IRLS) is used when [heteroscedasticity](https://en.wikipedia.org/wiki/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]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-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]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-15)[[16]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-16)
* [**Instrumental variables**](https://en.wikipedia.org/wiki/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* **z***i* such that E[**z***iεi*] = 0. If **Z** is the matrix of instruments, then the estimator can be given in closed form as
* **Optimal instruments** regression is an extension of classical IV regression to the situation where E[*εi* | **z***i*] = 0.
* [**Total least squares**](https://en.wikipedia.org/wiki/Total_least_squares) (TLS)[[17]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-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.

**Maximum-likelihood estimation and related techniques**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=14)]

* [**Maximum likelihood estimation**](https://en.wikipedia.org/wiki/Maximum_likelihood_estimation) can be performed when the distribution of the error terms is known to belong to a certain parametric family *ƒθ* of [probability distributions](https://en.wikipedia.org/wiki/Probability_distribution).[[18]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-18) When *f*θ is a normal distribution with zero [mean](https://en.wikipedia.org/wiki/Expected_value) 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**](https://en.wikipedia.org/wiki/Ridge_regression),[[19]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-19)[[20]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-20)[[21]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-21) and other forms of penalized estimation such as [**Lasso regression**](https://en.wikipedia.org/wiki/Lasso_(statistics)),[[5]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-tibs_lasso-5) deliberately introduce [bias](https://en.wikipedia.org/wiki/Bias_of_an_estimator) into the estimation of *β* in order to reduce the [variability](https://en.wikipedia.org/wiki/Variance) of the estimate. The resulting estimators generally have lower [mean squared error](https://en.wikipedia.org/wiki/Mean_squared_error) than the OLS estimates, particularly when [multicollinearity](https://en.wikipedia.org/wiki/Multicollinearity) is present or when [overfitting](https://en.wikipedia.org/wiki/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**](https://en.wikipedia.org/wiki/Least_absolute_deviation) (LAD) regression is a [robust estimation](https://en.wikipedia.org/wiki/Robust_regression) technique in that it is less sensitive to the presence of outliers than OLS (but is less [efficient](https://en.wikipedia.org/wiki/Efficiency_(statistics)) than OLS when no outliers are present). It is equivalent to maximum likelihood estimation under a [Laplace distribution](https://en.wikipedia.org/wiki/Laplace_distribution) model for *ε*.[[22]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-22)
* **Adaptive estimation**. If we assume that error terms are [independent](https://en.wikipedia.org/wiki/Independence_(probability_theory)) from the regressors

, 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]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-23)

**Other estimation techniques**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=15)]

[](https://en.wikipedia.org/wiki/File:Thiel-Sen_estimator.svg)

Comparison of the [Theil–Sen estimator](https://en.wikipedia.org/wiki/Theil%E2%80%93Sen_estimator) (black) and [simple linear regression](https://en.wikipedia.org/wiki/Simple_linear_regression) (blue) for a set of points with outliers.

* [**Bayesian linear regression**](https://en.wikipedia.org/wiki/Bayesian_linear_regression) applies the framework of [Bayesian statistics](https://en.wikipedia.org/wiki/Bayesian_statistics) to linear regression. (See also [Bayesian multivariate linear regression](https://en.wikipedia.org/wiki/Bayesian_multivariate_linear_regression).) In particular, the regression coefficients β are assumed to be [random variables](https://en.wikipedia.org/wiki/Random_variable) with a specified [prior distribution](https://en.wikipedia.org/wiki/Prior_distribution). The prior distribution can bias the solutions for the regression coefficients, in a way similar to (but more general than) [ridge regression](https://en.wikipedia.org/wiki/Ridge_regression) or [lasso regression](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Quantile_regression)), or any other function of the posterior distribution.
* [**Quantile regression**](https://en.wikipedia.org/wiki/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 βT*x* of the predictors.
* [**Mixed models**](https://en.wikipedia.org/wiki/Mixed_model) 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](https://en.wikipedia.org/wiki/Parametric_statistics) models, using maximum likelihood or Bayesian estimation. In the case where the errors are modeled as [normal](https://en.wikipedia.org/wiki/Normal_distribution) random variables, there is a close connection between mixed models and generalized least squares.[[24]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-24) [Fixed effects estimation](https://en.wikipedia.org/wiki/Fixed_effects_estimation) is an alternative approach to analyzing this type of data.
* [**Principal component regression**](https://en.wikipedia.org/wiki/Principal_component_regression) (PCR)[[7]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-hawkins_pcr-7)[[8]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-joliffe_pcr-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](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Partial_least_squares_regression) is the extension of the PCR method which does not suffer from the mentioned deficiency.
* [**Least-angle regression**](https://en.wikipedia.org/wiki/Least-angle_regression)[[6]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-efron_lars-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**](https://en.wikipedia.org/wiki/Theil%E2%80%93Sen_estimator) is a simple [robust estimation](https://en.wikipedia.org/wiki/Robust_regression) 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](https://en.wikipedia.org/wiki/Outlier).[[25]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-25)
* Other robust estimation techniques, including the **α-trimmed mean** approach, and **L-, M-, S-, and R-estimators** have been introduced.

**Further discussion**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=16)]

In [statistics](https://en.wikipedia.org/wiki/Statistics) and [numerical analysis](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Statistical_model) and for exploration of data sets. The majority of [statistical computer packages](https://en.wikipedia.org/wiki/Comparison_of_statistical_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](https://en.wikipedia.org/wiki/Precision_(computer_science)).

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](https://en.wikipedia.org/wiki/Dependent_variable) but different sets of [independent variables](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Linear_least_squares_(mathematics)) problem.

Matrix calculations, like any others, are affected by [rounding errors](https://en.wikipedia.org/wiki/Rounding_error). An early summary of these effects, regarding the choice of computational methods for matrix inversion, was provided by Wilkinson.[[26]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-26)

**Using linear algebra**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=17)]

It follows that one can find a "best" approximation of another function by minimizing the area between two functions, a continuous function

 on

 and a function

 where

 is a subspace of

:

all within the subspace

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

as an adequate criterion for obtaining the least squares approximation, function

, of

 with respect to the inner product space

.

As such,

 or, equivalently,

, can thus be written in vector form:

In other words, the least squares approximation of

 is the function

 closest to

 in terms of the inner product

. Furthermore, this can be applied with a theorem:

Let

 be continuous on

, and let

 be a finite-dimensional subspace of

. The least squares approximating function of

 with respect to

 is given by

where

 is an orthonormal basis for

.

Applications of linear regression[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=18)]

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.

**Trend line**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=19)]

*Main article:*[*Trend estimation*](https://en.wikipedia.org/wiki/Trend_estimation)

A **trend line** represents a trend, the long-term movement in [time series](https://en.wikipedia.org/wiki/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.

**Epidemiology**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=20)]

Early evidence relating [tobacco smoking](https://en.wikipedia.org/wiki/Tobacco_smoking) to mortality and [morbidity](https://en.wikipedia.org/wiki/Morbidity) came from [observational studies](https://en.wikipedia.org/wiki/Observational_studies) employing regression analysis. In order to reduce [spurious correlations](https://en.wikipedia.org/wiki/Spurious_correlation) 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](https://en.wikipedia.org/wiki/Randomized_controlled_trial) 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](https://en.wikipedia.org/wiki/Instrumental_variables) regression may be used to attempt to estimate causal relationships from observational data.

**Finance**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=21)]

The [capital asset pricing model](https://en.wikipedia.org/wiki/Capital_asset_pricing_model) uses linear regression as well as the concept of [beta](https://en.wikipedia.org/wiki/Beta_(finance)) 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.

**Economics**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=22)]

*Main article:*[*Econometrics*](https://en.wikipedia.org/wiki/Econometrics)

Linear regression is the predominant empirical tool in [economics](https://en.wikipedia.org/wiki/Economics). For example, it is used to predict [consumption spending](https://en.wikipedia.org/wiki/Consumption_(economics)),[[27]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-27) [fixed investment](https://en.wikipedia.org/wiki/Fixed_investment) spending, [inventory investment](https://en.wikipedia.org/wiki/Inventory_investment), purchases of a country's [exports](https://en.wikipedia.org/wiki/Exports),[[28]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-Krugman-28) spending on [imports](https://en.wikipedia.org/wiki/Imports),[[28]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-Krugman-28) the [demand to hold liquid assets](https://en.wikipedia.org/wiki/Money_demand),[[29]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-29) [labor demand](https://en.wikipedia.org/wiki/Labour_economics),[[30]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-Ehrenberg-30) and [labor supply](https://en.wikipedia.org/wiki/Labor_supply).[[30]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-Ehrenberg-30)

**Environmental science**[[edit](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=23)]

|  |  |
| --- | --- |
| [[icon]](https://en.wikipedia.org/wiki/File:Wiki_letter_w_cropped.svg) | **This section needs expansion**. You can help by [adding to it](https://en.wikipedia.org/w/index.php?title=Linear_regression&action=edit&section=).*(January 2010)* |

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](https://en.wikipedia.org/wiki/Benthic_zone) surveys to measure the effects of pulp mill or metal mine effluent on the aquatic ecosystem.[[31]](https://en.wikipedia.org/wiki/Linear_regression#cite_note-31)