Contents

[Linear Regression ( OLS) 1](#_Toc504753292)

[Advanced types of linear regression 1](#_Toc504753293)

[o Ridge regression 1](#_Toc504753294)

[o Lasso regression 1](#_Toc504753295)

[o Partial least squares (PLS) regression 1](#_Toc504753296)

[Nonlinear regression 1](#_Toc504753297)

[Binary Logistic Regression 1](#_Toc504753298)

[Ordinal Logistic Regression 2](#_Toc504753299)

[Nominal Logistic Regression 2](#_Toc504753300)

[Poisson regression 2](#_Toc504753301)

[Alternatives to Poisson regression for count data 3](#_Toc504753302)

[Negative binomial regression 3](#_Toc504753303)

[Zero-inflated models 3](#_Toc504753304)

# Linear Regression ( OLS)

# Advanced types of linear regression

Linear models are the oldest type of regression. It was designed so that [statisticians](http://statisticsbyjim.com/glossary/statistics/) can do the calculations by hand. However, OLS has several weaknesses, including a sensitivity to both [outliers](http://statisticsbyjim.com/glossary/outliers/) and [multicollinearity](http://statisticsbyjim.com/regression/multicollinearity-in-regression-analysis/), and it is prone to [overfitting](http://statisticsbyjim.com/regression/overfitting-regression-models/). To address these problems, statisticians have developed several advanced variants:

* Ridge regression allows you to analyze data even when severe multicollinearity is present and helps prevent overfitting. This type of model reduces the large, problematic variance that multicollinearity causes by introducing a slight bias in the estimates. The procedure trades away much of the variance in exchange for a little bias, which produces more useful [coefficient](http://statisticsbyjim.com/glossary/regression-coefficient/) estimates when multicollinearity is present.
* Lasso regression (least absolute shrinkage and selection operator) performs variable selection that aims to increase prediction accuracy by identifying a simpler model. It is similar to Ridge regression but with variable selection.
* Partial least squares (PLS) regressionis useful when you have very few observations compared to the number of independent variables or when your independent variables are highly correlated. PLS decreases the independent variables down to a smaller number of uncorrelated components, similar to Principal Components Analysis. Then, the procedure performs linear regression on these components rather the original data. PLS emphasizes developing predictive models and is not used for screening variables. Unlike OLS, you can include multiple continuous dependent variables. PLS uses the [correlation](http://statisticsbyjim.com/glossary/correlation/)structure to identify smaller effects and model multivariate patterns in the dependent variables.

# Nonlinear regression

Nonlinear regression also requires a continuous dependent variable, but it provides a greater flexibility to fit curves than linear regression.

## Binary Logistic Regression

Use [binary logistic regression](http://statisticsbyjim.com/glossary/binary-logistic-regression/) to understand how changes in the independent variables are associated with changes in the probability of an event occurring. This type of model requires a binary dependent variable. A [binary variable](http://statisticsbyjim.com/glossary/binary-variables/) has only two possible values, such as pass and fail.

**Example:** Political scientists assess the odds of the incumbent U.S. President winning reelection based on stock market performance.

Read my post about a binary logistic model that [estimates the probability of House Republicans belonging to the Freedom Caucus](http://statisticsbyjim.com/regression/statistical-analysis-republican-split/).

## Ordinal Logistic Regression

[Ordinal logistic regression](http://statisticsbyjim.com/glossary/ordinal-logistic-regression/) models the relationship between a set of predictors and an [ordinal response](http://statisticsbyjim.com/glossary/ordinal-variables/) variable. An ordinal response has at least three groups which have a natural order, such as hot, medium, and cold.

**Example:** Market analysts want to determine which variables influence the decision to buy large, medium, or small popcorn at the movie theater.

## Nominal Logistic Regression

[Nominal logistic regression](http://statisticsbyjim.com/glossary/nominal-logistic-regression/) models the relationship between a set of independent variables and a nominal dependent variable. A [nominal variable](http://statisticsbyjim.com/glossary/nominal-variables/) has at least three groups which do not have a natural order, such as scratch, dent, and tear.

**Example**: A quality analyst studies the variables that affect the odds of the type of product defects: scratches, dents, and tears.

# Poisson regression

Count data frequently follow the Poisson distribution, which makes Poisson Regression a good possibility. [Poisson variables](http://statisticsbyjim.com/glossary/poisson-variables/) are a count of something over a constant amount of time, area, or another consistent length of observation. With a Poisson variable, you can calculate and assess a rate of occurrence. A classic example of a Poisson dataset is provided by Ladislaus Bortkiewicz, a Russian economist, who analyzed annual deaths caused by horse kicks in the Prussian Army from 1875-1984.

Use Poisson regression to model how changes in the independent variables are associated with changes in the counts. Poisson models are similar to logistic models because they use Maximum Likelihood Estimation and transform the dependent variable using the natural log. Poisson models can be suitable for rate data, where the rate is a count of events divided by a measure of that unit’s exposure (a consistent unit of observation). For example, homicides per month.

**Example**: An analyst uses Poisson regression to model the number of calls that a call center receives daily.

# Alternatives to Poisson regression for count data

Not all count data follow the Poisson distribution because this distribution has some stringent restrictions. Fortunately, there are alternative analyses you can perform when you have count data.

Negative binomial regression: Poisson regression assumes that the variance equals the mean. When the variance is greater than the mean, your model has overdispersion. A negative binomial model, also known as NB2, can be more appropriate when overdispersion is present.

Zero-inflated models: Your count data might have too many zeros to follow the Poisson distribution. In other words, there are more zeros than the Poisson regression predicts. Zero-inflated models assume that two separate processes work together to produce the excessive zeros. One process determines whether there are zero events or more than zero events. The other is the Poisson process that determines how many events occur, some of which some can be zero. An example makes this clearer!

Suppose park rangers count the number of fish caught by each park visitor as they exit the park. A zero-inflated model might be appropriate for this scenario because there are two processes for catching zero fish:

* Some park visitors catch zero fish because they did not go fishing.
* Other visitors went fishing, and some of these people caught zero fish.

Whew! That’s many different types of regression analysis! If you’re trying to figure out which one to choose, I hope you will use this information to point yourself in the right direction!

If you’re learning regression, check out my [Regression Tutorial](http://statisticsbyjim.com/regression/regression-tutorial-analysis-examples/)!