

Linear Regression in R

Rahul Goswami

2022-01-22

Linear Regression is a statistical method for estimating the parameters of a linear model. where the model can be written as $y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$, *in matrix form we can write the same as $y = Xb$* where X is the design matrix and b is the vector of parameters. The linear regression model is a generalization of the ordinary least squares (OLS) model. It has certain underlying assumptions and properties for OLS Estimate such as the model is linear in parameters, non-multicollinearity of the variables, etc. The ordinary least squares (OLS) Estimate of the model can be estimated by minimizing the sum of squares of the residuals. The residuals are the difference between the observed values and the predicted values. The OLS estimate for the parameters is given by the formula:

$$b = (X^T X)^{-1} X^T y$$

Example

Generating Dataset

```
nobs = 10000
x1 = runif(nobs)
x2 = runif(nobs)
x3 = runif(nobs)
```

Model

```
y = 3*x1 + 2*x2 + x3 + rnorm(nobs)
```

Now we will fit the model using OLS

```
X = matrix(c(x1,x2,x3),ncol=3)
b = solve(t(X) %*% X, t(X) %*% y)
cat("Estimated coefficients:",b)
```

```
## Estimated coefficients: 2.978325 2.008538 1.028334
```