Package 'linreg'

April 16, 2026

Title Linear Regression and Model Selection Framework

Type Package

Version 0.1.0

Author Dr. Pramit Pandit [aut, cre],
Dr. Bikramjeet Ghose [aut],
Dr. Chiranjit Mazumder [aut]

Maintainer Dr. Pramit Pandit <pramitpandit@gmail.com>

Description Provides a comprehensive framework for linear regression modeling and associated statistical analysis. The package implements methods for correlation analysis, including computation of correlation matrices with corresponding significance levels and visualization via correlation heatmaps. It supports estimation of multiple linear regression models, along with automated model selection through backward elimination procedures based on statistical significance criteria. In addition, the package offers a suite of diagnostic tools to assess key assumptions of linear regression, including multicollinearity using variance inflation factors, heteroscedasticity using the Goldfeld-Quandt test, and normality of residuals using the Shapiro-Wilk test. These functionalities, as described in Draper and Smith (1998) <doi:10.1002/9781118625590>, are designed to facilitate robust model building, evaluation, and interpretation in applied statistical and data analytical contexts.

License GPL-3

Encoding UTF-8

RoxygenNote 7.3.3

Imports stats, Hmisc, corrplot, car, lmtest

NeedsCompilation no

Repository CRAN

Date/Publication 2026-04-16 18:20:18 UTC

Contents

`autoreg` *Multiple Linear Regression with Backward Elimination*

Description

Performs multiple linear regression using backward elimination based on p-value threshold and provides full model diagnostics including ANOVA, multicollinearity, heteroscedasticity, normality test, and plots.

Usage

```
autoreg(data, threshold = 0.1)
```

Arguments

`data`	A data frame containing dependent variable (y) in the first column and independent variables (x's) in remaining columns
`threshold`	Significance level for variable removal (default = 0.10)

Details

The function starts with a full model and iteratively removes the variable with the highest p-value greater than the specified threshold until all variables are significant.

Value

A list containing:

- final_model: Final regression model
- model_summary: Summary of final model
- selected_variables: Variables retained in final model
- anova_table: ANOVA table for final model
- vif: Variance Inflation Factor values (if applicable)
- gq_test: Goldfeld-Quandt test result
- shapiro_test: Shapiro-Wilk normality test result
- actual_vs_fitted: Data frame of actual vs fitted values

Examples

```
{
library(car)
library(lmtest)

set.seed(123)
n <- 40

x1 <- rnorm(n, 50, 10)
x2 <- rnorm(n, 30, 5)
x3 <- rnorm(n, 70, 15)
x4 <- rnorm(n, 20, 7)
x5 <- rnorm(n, 100, 20)
x6 <- rnorm(n, 10, 3)

y <- 0.5*x1 - 0.3*x2 + 0.2*x3 +
     0.1*x4 - 0.05*x5 + 0.3*x6 +
     rnorm(n, 0, 15)

df <- data.frame(y, x1, x2, x3, x4, x5, x6)

result <- autoreg(df, threshold = 0.10)
result$selected_variables
}
```

`CorrAnalysis` *Correlation Analysis with P-value Matrix and Heatmap*

Description

Computes the correlation matrix along with corresponding p-values and visualizes the correlations using a heatmap.

Usage

```
CorrAnalysis(data)
```

Arguments

`data`	A numeric data frame or matrix containing variables (e.g., one dependent variable y and multiple independent variables x).

Value

A list containing:

- correlation_matrix: Numeric correlation matrix
- p_value_matrix: Formatted p-value matrix (character)

`RegAnalysis` *Multiple Linear Regression Full Model Diagnostics*

Description

Fits a multiple linear regression model and provides detailed diagnostics including ANOVA table, multicollinearity, heteroscedasticity, normality test, and diagnostic plots.

Usage

```
RegAnalysis(data)
```

Arguments

`data`	A data frame containing dependent variable (y) and independent variables (x's)

Value

A list containing:

- model_summary: Summary of regression model
- anova_table: ANOVA table (SSR, SSE, SST)
- vif: Variance Inflation Factor values
- gq_test: Goldfeld-Quandt test result
- shapiro_test: Shapiro-Wilk normality test result
- actual_vs_fitted: Data frame of actual vs fitted values

Index