Package 'regkurs'

February 25, 2022

Title Kurspaket - Regressions- och tidsserieanalys

Version 0.0.1

Description Hjälpfunktioner och datamaterial för kursen Regressions- och tidsserieanalys

License `use_gpl3_license()`

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.2

Imports glmnet,
mvtnorm

Depends R (>= 2.10)

LazyData true

Suggests rmarkdown,
knitr

VignetteBuilder knitr

R topics documented:

bike 2
logisticregsimulate 2
logisticregsummary 3
regsimulate 4
regsummary 5
titanic 6

Index **8**

bike *Number of daily rides for a bike share company in Washington D.C.*

Description

A dataset containing the number of rides per day and other attributes over the course of 2 years

Usage

```
bike
```

Format

A data frame with 731 rows and 12 variables:

dteday date in YYYY-MM-DD format

season categorical variable (1="winter", 2 = "spring", 3 = "summer", 4 = "fall")

yr year (0="2011", 1 = "2012")

mnth month from 1-12 where 1 = "January"

holiday binary variable for public holidays

weekday day of the week 0-6, 0 ="Sunday"

workingday binary variable for working days (=1)

weathersit categorical variable (1="clear", 2 = "mist", 3 = "light snow")

temp continuous temperature variable, normalized between 0,1

hum continuous humidity variable, normalized between 0,1

windspeed continuous windspeed variable, normalized between 0,1 ...

Source

https://archive.ics.uci.edu/ml/datasets/bike+sharing+dataset

logisticregsimulate *Simulate from a logistic regression model*

Description

Simulates a dataset with n observation from the logistic regression model

$$\Pr(y = 1|x) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k))}$$

with covariates (x) simulated from a normal distribution with the same correlation `rho_x` between all pairs of covariates. Covariate x_j has standard deviation `sigma_x[j]`. Alternatively the covariate can follow a uniform distribution.

Usage

```
logisticregsimulate(
  n,
  betavect,
  intercept = TRUE,
  covdist = "normal",
  rho_x = 0,
  sigma_x = rep(1, length(betavect) - intercept)
)
```

Arguments

n	the number of observations in the simulated dataset.
betavect	a vector with regression coefficients c(beta_0,beta_1,...beta_k). First element is intercept if `intercept = TRUE`
intercept	if `TRUE` an intercept is added to the model.
covdist	distribution of the covariates. Options: `'normal'` or `'uniform'`.
rho_x	correlation among the covariates. Same for all covariate pairs.
sigma_x	vector with standard deviation of the covariates.

Value

dataframe with simulated data (y, X1, X2, ..., XK) (no intercept included).

Examples

```
library(regkurs)
simdata <- logisticregsimulate(n = 500, betavect = c(1, -2, 1, 0))
glmfit <- glm(y ~ X1 + X2 + X3, data = simdata, family = binomial)
logisticregsummary(glmfit, odds_ratio = F)
```

---

logisticregsummary *Summarize the results from a logistic regression analysis*

Description

Alternative to `summary.glm` to summarize a regression from `glm`. Prints a table similar to the one generated by SAS and Minitab.

Usage

```
logisticregsummary(
  glmobject,
  odds_ratio = T,
  param = T,
  conf_intervals = F,
  vif_factors = F
)
```

Arguments

glmobject	a fitted regression model from glm.
param	TRUE if parameter estimates, standard errors etc is computed.
conf_intervals	TRUE if confidence intervals for parameters.
vif_factors	TRUE if variance inflation factors are to be printed.

Value

list with two tables: param, odds_ratio

Examples

```
library(regkurs)
glmfit <- glm(survived ~ age + sex + firstclass, data = titanic, family = binomial)
logisticregsummary(glmfit)
```

---

regsimulate *Simulate from a linear regression model*

Description

Simulates a dataset with n observation from the linear regression model

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \epsilon, \epsilon \sim N(0, \sigma_\epsilon^2)$$

with covariates (x) simulated from a normal distribution with the same correlation rho_x between all pairs of covariates. Covariate x_j has standard deviation sigma_x[j]. Alternatively the covariate can follow a uniform distribution.

Usage

```
regsimulate(
  n,
  betavect,
  sigma_eps,
  intercept = TRUE,
  covdist = "normal",
  rho_x = 0,
  sigma_x = rep(1, length(betavect) - intercept)
)
```

Arguments

n	the number of observations in the simulated dataset.
betavect	a vector with regression coefficients c(beta_0,beta_1,...beta_k). First element is intercept if intercept = TRUE
sigma_eps	standard deviation of the error terms, epsilon.
intercept	if TRUE an intercept is added to the model.
covdist	distribution of the covariates. Options: 'normal' or 'uniform'.
rho_x	correlation among the covariates. Same for all covariate pairs.
sigma_x	vector with standard deviation of the covariates.

Value

dataframe with simulated data (y, X1, X2, ..., XK) (no intercept included).

Examples

```
library(regkurs)
simdata <- regsimulate(n = 500, betavect = c(1, -2, 1, 0), sigma_eps = 2)
lmfit <- lm(y ~ X1 + X2 + X3, data = simdata)
regsummary(lmfit, anova = F)
```

---

regsummary *Summarize the results from a regression analysis*

Description

Alternative to summary.lm to summarize a regression from lm. Prints a table similar to the one generated by SAS and Minitab.

Usage

```
regsummary(
  lmobject,
  anova = T,
  fit_measures = T,
  param = T,
  conf_intervals = F,
  vif_factors = F
)
```

Arguments

`lmobject`	a fitted regression model from `lm`.
`anova`	`TRUE` if an ANOVA table is computed.
`fit_measures`	`TRUE` if measures of fit (R^2 etc) is computed.
`param`	`TRUE` if parameter estimates, standard errors etc is computed.
`conf_intervals`	`TRUE` if confidence intervals for parameters.
`vif_factors`	`TRUE` if variance inflation factors are to be printed.

Value

list with three tables: param, anova and fit_measures

Examples

```
library(regkurs)
lmfit = lm(nRides ~ temp + hum + windspeed, data = bike)
regsumm = regsummary(lmfit, anova = T, conf_intervals = T, vif_factors = T)
regsumm$param
regsumm$anova
regsumm$fit_measures
```

`titanic` *Survival of passengers on the Titanic*

Description

This data set provides information on the fate of passengers on the fatal maiden voyage of the ocean liner 'Titanic', summarized according to economic status (class), sex, age and survival.

NOTE: this is not the same as the dataset Titanic (note capital T) which has more observations, but also missing values.

Usage

```
titanic
```

Format

A data frame with 887 rows and 8 variables:

name passenger name

survived 0 = no, 1 = yes

sex male/female

age age of passenger

fare ticket cost

firstclass first class ticket ...

Details

The sinking of the Titanic is a famous event, and new books are still being published about it. Many well-known facts—from the proportions of first-class passengers to the 'women and children first' policy, and the fact that that policy was not entirely successful in saving the women and children in the third class—are reflected in the survival rates for various classes of passenger.

These data were originally collected by the British Board of Trade in their investigation of the sinking. Note that there is not complete agreement among primary sources as to the exact numbers on board, rescued, or lost.

Due in particular to the very successful film 'Titanic', the last years saw a rise in public interest in the Titanic. Very detailed data about the passengers is now available on the Internet, at sites such as Encyclopedia Titanica (https://www.encyclopedia-titanica.org/).

Source

Dawson, Robert J. MacG. (1995), The 'Unusual Episode' Data Revisited. Journal of Statistics Education, 3. doi: 10.1080/10691898.1995.11910499.

Index

∗ **datasets**
    bike, 2
    titanic, 6
0,1, 2

bike, 2

logisticregsimulate, 2
logisticregsummary, 3

regsimulate, 4
regsummary, 5

titanic, 6