Package ‘allestimates’

February 5, 2020

Type Package

Title Effect Estimates from All Models

Version 0.1.6

Description Estimates and plots effect estimates from models with all possible combinations of a list of variables. It can be used for assessing treatment effects in clinical trials or risk factors in bio-medical and epidemiological research. Like Stata command 'confall' (Wang Z (2007) <doi:10.1177/1536867X0700700203> ), 'allestimates' calculates and stores all effect estimates, and plots them against p values or Akaike information criterion (AIC) values. It currently has functions for linear regression: all_lm(), logistic and Poisson regression: all_glm() and all_speedglm(), and Cox proportional hazards regression: all_cox().

License GPL-2

Encoding UTF-8

LazyData true

Imports broom, ggplot2, MASS, speedglm, survival, tibble, tidyr, utils, stringr, dplyr

Depends R (>= 2.10)

RoxygenNote 7.0.2

Suggests spelling, knitr, rmarkdown

VignetteBuilder knitr

Language en-US

NeedsCompilation no

Author Zhiqiang Wang [aut, cre]

Maintainer Zhiqiang Wang <zhiqiang.wang@menzies.edu.au>

Repository CRAN

Date/Publication 2020-02-05 16:30:12 UTC
R topics documented:

allestimates .................................................. 2
all_cox .......................................................... 3
all_glm ........................................................... 4
all_lm ............................................................. 5
all_plot ............................................................ 5
all_plot2 ........................................................... 6
all_plot_aic ....................................................... 7
all_plot_aic2 ...................................................... 8
all_speedglm ..................................................... 9
diab_df .......................................................... 10

Index 11

allestimates  Effect estimates from models with all possible combinations of variables

Description

To assess treatment effects in clinical trials and risk factors in bio-medical and epidemiological research, we use regression coefficients, odds ratios or hazard ratios as effect estimates. all_estimates allows users to quickly obtain effect estimates from models with all possible combinations of a list of variables specified by users. all_lm for linear regression, all_glm for logistic regression, all_speedglm using speedlm as a faster alternative of all_glm, and all_cox for Cox Proportional Hazards Models. Users can further use those values in a returned list of results. all_plot draws scatter plots with all effect estimate values against p values, as Stata confall command (Wang Z (2007) <doi:10.1177/1536867X0700700203>). Those plots divide estimates into four categories:

Details

- positive and significant: left-top quarter
- negative and significant: left-bottom quarter
- positive and non-significant: right-top quarter
- negative and non-significant: right-bottom quarter

all_plot2 draws multiple plots. Each of those plots indicates whether a specific variable is included or not included in models. Those effect estimates help users better understand confounding effects, uncertainty of their estimates, as well as inappropriately including variables in the models. This is a tool for calculating and exploring effect estimates from all possible models. Interpretation of the results should be in the context of other analyses and biological knowledge.
**all_cox**

### Examples

```r
? all_speedglm
? all_glm
? all_cox
? all_lm
? all_plot
? all_plot2
```

---

**all_cox**

*Estimates all possible effect estimates using Cox Proportional Hazards regression models*  

---

**Description**

Estimates hazard ratios using Proportional Hazards Regression models ("coxph" from **survival** package) from models with all possible combinations of a list of variables.

**Usage**

```r
all_cox(crude, xlist, data, na_omit = TRUE, ...)
```

**Arguments**

- `crude`: An object of `formula` for initial model, generally crude model. However, any other variables can also be included here as the initial model. The left-hand side of ~ is the outcome of interest, and the variable on the right-hand side of ~ is the exposure of the interest (either a treatment or a risk factor).
- `xlist`: A vector of a list of variable names.
- `data`: Data frame.
- `na_omit`: Remove all missing values. Default is "na_omit = TRUE".
- `...`: Further optional arguments.

**Value**

A list of all effect estimates.

**See Also**

**survival**

**Examples**

```r
vlist <- c("Age", "Sex", "Married", "BMI", "Education", "Income")
results <- all_cox(crude = "Surv(t0, t1, Endpoint) ~ Diabetes", xlist = vlist, data = diab_df)
results
```
all_glm  Estimates all possible effect estimates using glm

Description

all_glm estimates odds ratios or rate ratios using generalized linear models (glm) with all possible combinations of a list of variables (potential confounding factors).

Usage

```r
all_glm(crude, xlist, data, family = "binomial", na_omit = TRUE, ...)
```

Arguments

- `crude` An object of formula for initial model, generally crude model. However, any other variables can also be included here as the initial model.
- `xlist` A vector of a list of variable names (potential confounding factors).
- `data` Data frame.
- `family` family Description of the error distribution. Default is "binomial".
- `na_omit` Remove all missing values. Default is "na_omit = TRUE".
- `...` Further optional arguments.

Value

A list of all effect estimates.

See Also

stats

Examples

```r
diab_df$Overweight <- as.numeric(diab_df$BMI >= 25)
vlist <- c("Age", "Sex", "Income")
all_glm(crude = "Diabetes ~ Overweight", xlist = vlist, data = diab_df)
```
### all_lm

Estimates all possible effect estimates using `lm`

#### Description

`all_lm` estimates coefficients of a specific variable using linear models (`lm`) with all possible combinations of other variables (potential confounding factors).

#### Usage

```r
all_lm(crude, xlist, data, na_omit = TRUE, ...)
```

#### Arguments

- **crude**: An object of `formula` for initial model, generally crude model. However, additional variables can also be included here as the initial model.
- **xlist**: A `vector` of a list of variable names (potential confounding factors).
- **data**: `Data frame`.
- **na_omit**: Remove all missing values. Default is "na_omit = TRUE".
- **...**: Further optional arguments.

#### Value

A list of all effect estimates.

#### See Also

`lm`

#### Examples

```r
vlist <- c("Age", "Sex", "Cancer", "CVD", "Education", "Income")
all_lm(crude = "BMI ~ Married", xlist = vlist, data = diab_df)
```

---

### all_plot

Plot all effect estimates against p values

#### Description

`all_plot()` generates a scatter plot with effect estimates of all possible models against p values.
Usage

```r
all_plot(
    data,
    xlabels = c(0, 0.001, 0.01, 0.05, 0.2, 0.5, 1),
    xlim = c(0, 1),
    xlab = "P value",
    ylim = NULL,
    ylab = NULL,
    yscale_log = FALSE,
    title = NULL
)
```

Arguments

data: Object from all_cox, all_glm, all_speedglm, or all_glm, including all effect estimate values.
xlabels: Numeric vector x-axis tick labels. Default is "c(0, 0.001, 0.01, 0.05, 0.2, 0.5, 1)".
xlim: Vector of 2 numeric values for x-axis limits. Default is "c(0,1)".
xlab: Character string for x-axis name. Default is "P value".
ylim: Vector of 2 numeric values for y-axis limits.
ylab: Character string for y-axis name. Default depends on original model types.
yscale_log: TRUE or FALSE to re-scale y-axis to "log10". Default is "FALSE".
title: Character for plot title. Default is "NULL".

Value

A `ggplot2` object: scatter plot

Examples

```r
vlist <- c("Age", "Sex", "Married", "BMI", "Education", "Income")
results <- all_cox(crude = "Surv(t0, t1, Endpoint) ~ Diabetes", xlist = vlist, data = diab_df)
all_plot(results)
```

Description

`all_plot2` plots all effect estimates against p values with each specific variable in the models.

`all_plot2()` generates a panel of scatter plots with effect estimates of all possible models against p values. Each plot includes effect estimates from all models including a specific variable.
Usage

all_plot2(
  data,
  xlabels = c(0, 0.001, 0.01, 0.05, 0.2, 0.5, 1),
  xlim = c(0, 1),
  xlab = "P value",
  ylim = NULL,
  ylab = NULL,
  yscale_log = FALSE,
  title = NULL
)

Arguments

data Object from all_cox, all_glm, all_speedglm, or all_glm, including all effect estimate values.
xlabels numeric vector x-axis tick labels. Default is "c(0,0.001,0.01,0.05,0.2,0.5,1)"
xlim vector of 2 numeric values for x-axis limits. Default is "c(0,1)".
xlab Character string for x-axis name. Default is "P value".
ylim vector of 2 numeric values for y-axis limits.
ylab Character string for y-axis name. Default depends on original model types.
yscale_log TRUE or FALSE re-scale y-axis to "log10". Default is "FALSE".
title Character title. Default is "NULL".

Value

A ggplot2 object: scatter plot

Examples

vlist <- c("Age", "Sex", "Married", "BMI", "Education", "Income")
results <- all_cox(crude = "Surv(t0, t1, Endpoint) ~ Diabetes", xlist = vlist, data = diab_df)
all_plot2(results)

all_plot_aic

Draws scatter plot with all effect estimates against AIC

Description

all_plot_aic() generates a scatter plot with all effect estimates against AIC.

Usage

all_plot_aic(data, xlab = "AIC", ylab = NULL, title = NULL)
all_plot_aic2

Arguments

data	Object from all_cox, all_glm, all_speedglm, or all_glm, including all effect estimate values.
xlab	Character string for x-axis name. Default is "AIC"
ylab	Character string for y-axis name. Default depends on original model types.
title	Character for plot title. Default is "NULL".

Value

A ggplot2 object: scatter plot.

Examples

vlist <- c("Age", "Sex", "Married", "BMI", "Education", "Income")
results <- all_cox(crude = "Surv(t0, t1, Endpoint) ~ Diabetes", xlist = vlist, data = diab_df)
all_plot_aic(data = results)

all_plot_aic2	Draws multiple scatter plots of all effect estimates against AIC

Description

all_plot_aic2() draws multiple scatter plots of all effect estimates against AIC. Each plot indicates if a specific variable is included in the models.

Usage

all_plot_aic2(data, xlab = "AIC", ylab = NULL, title = NULL)

Arguments

data	Object from all_cox, all_glm, all_speedglm, or all_glm, including all effect estimate values.
xlab	Character string for x-axis name. Default is "AIC".
ylab	Character string for y-axis name. Default depends on original model types.
title	Character for plot title. Default is "NULL".

Value

A ggplot2 object: scatter plot.

Examples

vlist <- c("Age", "Sex", "Married", "BMI", "Education", "Income")
results <- all_cox(crude = "Surv(t0, t1, Endpoint) ~ Diabetes", xlist = vlist, data = diab_df)
all_plot_aic(data = results)
**all_speedglm**  

*Effect estimates from all possible models using speedglm*

**Description**

This is a faster alternative to all_glm. all_speedglm estimates odds ratios or rate ratios using generalized linear models (speedglm) with all possible combinations of a list of variables (potential confounding factors) specified in xlist argument.

**Usage**

`all_speedglm(crude, xlist, data, family = binomial(), na_omit = TRUE, ...)`

**Arguments**

- **crude**: An object of formula for initial model, generally crude model. However, any other variables can also be included here as the initial model.
- **xlist**: A vector of characters with variable names to be included in as potential confounding factors.
- **data**: Data frame.
- **family**: Description of the error distribution. Default is `binomial()`.
- **na_omit**: Remove all missing values. Default is "na_omit = TRUE".
- **...**: Further optional arguments.

**Value**

A list of all effect estimates.

**See Also**

speedglm

**Examples**

```r
vlist <- c("Age", "Sex", "Married", "Cancer", "CVD", "Education", "Income")
results <- all_speedglm(crude = "Endpoint ~ Diabetes", xlist = vlist, data = diab_df)
results$estimate
```
**Example data: Health outcomes of 2372 adults with and without diabetes**

**Description**

A data frame with 2372 rows and 14 variables with diabetes status diabetes and mortality status endpoint. For the purpose of demonstrate, assume that we are interested in the association between diabetes and endpoint. Other variables are considered as possible confounding factors. The purposes of this dataset is to illustrate those functions in `chest` and `allestimates` packages only. Therefore, we assume it is a cohort design for Cox Proportional Hazard regression, and a case-control design for logistic regression.

**Usage**

```r
diab_df
```

**Format**

A data frame with 2372 rows and 14 variables:

- **Diabetes** diabetes status 1: with diabetes 0: without diabetes
- **Endpoint** mortality status 1: reached end point, and 0: survived
- **Age** Age, in years
- **Sex** sex, 1: male, 2: Female
- **BMI** Body mass index
- **Married** marital status 1: married, 0: not
- **Smoke** smoking status 1: smoker, 0: non-smoker
- **CVD** cardiovascular disease 1: yes 0: no
- **Cancer** cancer 1: yes, 0: no
- **Education** education 1: high, 0: low
- **Income** income 1: high, 0: low
- **t0** time (age) at the start of the follow-up
- **t1** time (age) at the end of the follow-up
- **mid** matched set id, for conditional logistic regression
Index

*Topic datasets
  diab_df, 10
  all_cox, 3
  all_glm, 4
  all_lm, 5
  all_plot, 5
  all_plot2, 6
  all_plot_aic, 7
  all_plot_aic2, 8
  all_speedglm, 9
  allesimates, 2

  diab_df, 10