Package ‘chest’

January 29, 2020

Type Package

Title Change-in-Estimate Approach to Assess Confounding Effects

Version 0.3.1

Description Applies the change-in-effect estimate method to assess confounding effects in medical and epidemiological research (Greenland & Pearce (2016) <doi:10.1146/annurev-publhealth-031914-122559>). It starts with a crude model including only the outcome and exposure variables. At each of the subsequent steps, one variable which creates the largest change among the remaining variables is selected. This process is repeated until all variables have been entered into the model (Wang Z (2007) <doi:10.1177/1536867X0700700203>). Currently, the 'chest' package has functions for linear regression, logistic regression, negative binomial regression, Cox proportional hazards model and conditional logistic regression.

License GPL-2

Encoding UTF-8

LazyData true

Imports broom, ggplot2, survival, grid, speedglm, forestplot, MASS, tibble, magrittr, dplyr

Depends R (>= 2.20)

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 2020-01-29 12:20:02 UTC
Description

In clinical trials and epidemiological studies, the association between an exposure and the outcome of interest in a study can be estimated by regression coefficients, odds ratios or hazard ratios depending on the nature of study designs and outcome measurements. We use a general term effect estimate here for any of those measurements in this document. Based on those measurements, we determine if a treatment is effective (or detrimental) or a factor is a risk factor. Imbalanced distributions of other factors could bias the effect estimates, called confounding. One way to assess the confounding effect of a factor is to examine the difference in effect estimates between models with and without a specific factor. 'chest' allows users quickly calculate the changes when potential confounding factors are sequentially added to the model in a stepwise fashion. At each step, one variable which creates the largest change (%) of the effect estimate among the remaining variables is added to the model. 'chest' returns a graph and a data frame (table) with effect estimates (95% CI) and change (%) values. The package currently has the following main functions: 'chest_lm' for linear regression, 'chest_glm' for logistic regression and Poisson regression, 'chest_speedglm' using 'speedlm' as a faster alternative of 'chest_glm', 'chest_clogit' for matched logistic regression, 'chest_nb' for negative binomial regression and 'chest_cox' for Cox proportional hazards models.

References


Examples

? chest_speedglm
? chest_glm
? chest_cox
? chest_clogit
? chest_lm
Assessing confounding effects using conditional logistic regression models

Description

'chest_clogit' is used to fit many Conditional Logistic Regression models to assess confounding effects.

Usage

```
chest_clogit(
  crude,  
  xlist,   
  data,    
  method = "exact",  
  na_omit = TRUE,   
  plus = " + ",   
  indicate = FALSE,   
  ...    
)
```

Arguments

- **crude**: An object of formula for the initial model, generally crude model. However, any other variables can also be included here as the initial model.
- **xlist**: A vector of characters with all variable names of potential confounders.
- **data**: Data frame.
- **method**: See 'clogit', default is the "exact" method.
- **na_omit**: Remove all missing values, default: 'na_omit = TRUE'.
- **plus**: Change the + sign before variable names.
- **indicate**: indicate the calculation progress.
- **...**: Further optional arguments.

Value

A table with effect estimates and their changes at all steps.

See Also

`chest`  
'clogit' in 'survival'
Examples
vlist <- c("Age", "Sex", "Married", "Cancer", "CVD", "Education", "Income")
chest_clogit(crude = "Endpoint ~ Diabetes + strata(mid)",
xlist = vlist, data = diab_df)

chest_cox Assessing confounding effects using Cox Proportional Hazards regression models

Description
'chest_cox' is used to assess confounding effects using Proportional Hazards Regression Model ('coxph' from 'survival' package). It presents the effect estimates (such as hazard ratios) for the association between exposure and outcome variables by adding other variables (potential confounders) to the model sequentially. The order of variables to be added is based on the magnitudes of the changes in effect estimates.

Usage
chest_cox(
  crude,
  xlist,
  data,
  na_omit = TRUE,
  plus = " + ",
  indicate = FALSE,
  ...
)

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 of potential confounders.
data Data frame.
na_omit Remove all missing values, default: 'na_omit = TRUE'.
plus Change the + sign before variable names.
indicate indicate the progress.
...

Value
A table with effect estimates and their changes at all steps.
See Also
'
survival’

Examples

vlist <- c("Age", "Sex", "Married", "Cancer", "CVD", "Education", "Income")

chest_cox(crude = "Surv(t0, t1, Endpoint) ~ Diabetes", xlist = vlist, data = diab_df)

---

**chest_forest**

Plot effect estimates and change-in-estimate values (forestplot type)

Description

'chest_forest' plots effect estimates and change-in-estimate values with forestplot package.

Usage

chest_forest(
  data,
  var_lab = "Variables",
  est_lab = "Estimate (95% CI)",
  change_lab = "Change, %",
  digits = ".2f",
  digits_change = ".1f",
  hrzl_lines = gpar(col = "#444444"),
  plus = "+",
  ...)

Arguments

data
  Object from chest_cox, chest_glm, chest_speedglm, chest_lm, chest_clogit, or chest_nb, including effect estimate values and change-in-estimate values.

var_lab
  Character string for the column name of variables in the graph.

est_lab
  Character string for the column name of effect estimates.

change_lab
  Character string for the column name of "Changes".

digits
  Set the display format for number in the graph other than the "Change" column. Default: ".2f"

digits_change
  Set the format for the "Change" column. Default: ".1f"

hrzl_lines
  A logic to include or remove horizontal line.

plus
  Change the + sign before variable names.

...  
  Further optional arguments for forestplot.
Value

A table with effect estimates and their changes at all steps.

See Also

'forestplot'

Examples

vlist <- c("Age", "Sex", "Married", "Cancer", "CVD", "Education", "Income")
results <- chest_cox(crude = "Surv(t0, t1, Endpoint) ~ Diabetes", xlist = vlist, data = diab_df)
chest_forest(results)

chest_glm

Assessing confounding effects using Generalized Linear regression models

Description

Please note: There is a faster option: 'chest_speedglm'. 'chest_glm' is used to assess confounding effects using Generalized Linear Models, such as logistic regression and Poisson regression with 'glm'. It presents odds ratios or rate ratios for the association between exposure and outcome variables by adding other variables (potential confounders) to the model sequentially. The order of variables to be added is based on the magnitudes of the changes in effect estimates.

Usage

chest_glm(
  crude,
  xlist,
  data,
  family = "binomial",
  method = "glm.fit",
  na_omit = TRUE,
  indicate = FALSE,
  plus = "+",
  ...
)

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 all variable names of potential confounders.
data Data frame.
family Description of the error distribution. Default is "binomial".
Assessing confounding effects using Linear regression models

Description

'chest_lm' is used to assess confounding effects using Linear Regression Models. It presents linear regression coefficients as effect estimates and changes when other variables are added sequentially to the model.

Usage

```r
chest_lm(
  crude = "Endpoint ~ Diabetes", xlist = c("Age", "Sex", "Married"),
  na_omit = TRUE, data = diab_df
)
```

Value

A table with effect estimates and their changes at all steps.

See Also

'glm' 'stats'

Examples

```r
chest_glm(crude = "Endpoint ~ Diabetes", xlist = c("Age", "Sex", "Married"),
  na_omit = TRUE, data = diab_df)
```
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 all variable names of potential confounders.
- **data**: Data frame.
- **method**: The method to be used; see 'lm'.
- **na_omit**: Remove all missing values.
- **indicate**: Indicate progress
- **plus**: Change the + sign before variable names.
- **...**: Further optional arguments.

Value

A table with effect estimates and their changes at all steps.

See Also

'lm' of 'stats'

Examples

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

chest_nb
```

Description

It fits negative binomial regression models to present rate ratios for the association between exposure and outcome variables by adding other variables (potential confounders) to the model sequentially. The order of variables to be added is based on the magnitudes of the changes in effect estimates.

Usage

```r
chest_nb(
  crude,
  xlist,
  data,
  method = "glm.fit",
  na_omit = TRUE,
  indicate = FALSE,
  plus = " + ",
  ...
)
```
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 all variable names of potential confounders.
- **data**: Data frame.
- **method**: Method to detect for singularity.
- **na_omit**: Remove all missing values, default: `na_omit = TRUE`.
- **indicate**: Indicate progress.
- **plus**: Change the + sign before variable names.
- **...**: Further optional arguments.

Value

A table with effect estimates and their changes at all steps.

See Also

`glm.nb` of 'MASS'

Examples

```r
library(MASS)
df <- quine
results <- chest_nb("Days ~ Lrn", xlist = c("Sex", "Age", "Eth"), data = df)
results$data
```

------

### chest_plot

**Plot effect estimate and change-in-estimate values (ggplot type)**

Description

Plot effect estimate and change-in-estimate values (ggplot type)

Usage

```r
chest_plot(
data,
  no_values = FALSE,
  ylab = NULL,
  xlab = NULL,
  change_lab = "Change, %",
  digits = "%.2f",
  digits_change = "%.1f",
  plus = " + ",
  nudge_y = 0.4,
)```

nudge_x = NULL,
hjust = 0.5,
height = 0.06,
point_size = 3,
point_shape = 15,
vline_type = "dashed",
vline_color = "grey50",
ebar_color = "grey50",
zero = 1,
value_position = NULL,
...)

Arguments

Arguments

data Object from chest_cox, chest_glm, chest_speedglm, chest_lm, chest_clogit,
or chest_nb, including effect estimate values and change-in-estimate values.

no_values Suppress effect estimate values in plot, default is FALSE.

ylab Add y label.
xlab Add x label.

change_lab Character string for the column name of "Change" in the graph.
digits Set the display format for number in the graph other than the "Change" column. Default: "%.2f"
digits_change Set the format for the "Change" column. Default: "%.1f"

plus Change the + sign before variable names.
nudge_y Adjust vertical distance between values and point marker.
nudge_x Adjust horizontal distance between values and point marker.
hjust Adjust horizontal alignment.

height Change the height of error bars.

point_size Change point marker size.

point_shape Change point marker shape.
vline_type Change vertical line type.
vline_color Change vertical line color.
ebar_color Change error bar color.
zero x-axis coordinate for vertical non-effect line, see forestplot.

value_position Change the position of value labels.

... Further optional arguments for forestplot.

Value

a ggplot object.
chest_speedglm

See Also

'ggplot2'

Examples

vlist<-c("Age", "Sex", "Married", "Cancer", "CVD", "Education", "Income")
results <- chest_speedglm(crude = "Endpoint ~ Diabetes",
                          xlist = vlist, na_omit=TRUE, data = diab_df)
chest_plot(results)

chest_speedglm Assessing confounding effects using Generalized Linear regression models

Description

This is a faster option to 'chest_glm'. It presents the effect estimates (such as odds ratios) for
the association between exposure and outcome variables by adding other variables (potential con-
founders) to the model sequentially. The order of variables to be added is based on the magnitudes
of the changes in effect estimates. It returns a graph and a table with effect estimates, 95% confi-
dence intervals and changes (%) at different steps.

Usage

chest_speedglm(
  crude,
  xlist,
  data,
  family = binomial(),
  method = c("eigen", "Cholesky", "qr"),
  na_omit = TRUE,
  plus = " + ",
  indicate = FALSE,
  ...
)

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 all variable names of potential confounders.
  data        Data frame.
  family      Description of the error distribution. Default is "binomial".
  method      Method to detect for singularity.
  na_omit     Remove all missing values, default: 'na_omit = TRUE'.

plus  Change the + sign before variable names.
indicate  indicate progress
...  Further optional arguments for speedglm.

Value
A table with effect estimates and their changes at all steps.

See Also
'speedglm'
'glm' of 'stats'

Examples
vlist<-c("Age", "Sex", "Married", "Cancer", "CVD", "Education", "Income")
chest_speedglm(crude = "Endpoint ~ Diabetes",
xlist = vlist, na_omit=TRUE, data = diab_df)
  # adding derived terms such as an interaction between Age and Sex, and age squared:
diab_df$Age_Sex = diab_df$Age*diab_df$Sex
diab_df$Age2 = diab_df$Age^2
chest_speedglm(crude = "Endpoint ~ Diabetes", xlist = vlist, na_omit=TRUE, data = diab_df)

---

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 confounders. The purposes of this dataset is to illustrate those functions in chest package only. Therefore, we assume it is a cohort design for Cox Proportional Hazard regression, and a case-control design for logistic regression.

Usage

Usage
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
diab_df

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, 12

  chest, 2
  chest_clogit, 3
  chest_cox, 4
  chest_forest, 5
  chest_glm, 6
  chest_lm, 7
  chest_nb, 8
  chest_plot, 9
  chest_speedglm, 11

  diab_df, 12