Package ‘EpiStats’

June 7, 2021

Type Package
Title Tools for Epidemiologists
Version 1.5-1
Date 2021-06-07
Description Provides set of functions aimed at epidemiologists.
They include commands for measures of association and impact for case control studies and co-
hort studies.
They may be particularly useful for outbreak investigations and include univariate and strati-
fied analyses.
The functions for cohort studies include the CS(), CSTable() and CSInter() commands.
The functions for case control studies include the CC(), CCTable() and CCInter() commands.
License LGPL-3
Depends epiR, dplyr
Suggests knitr, rmarkdown
VignetteBuilder knitr
NeedsCompilation no
Author Jean Pierre Decorps [aut, cre]
Maintainer Jean Pierre Decorps <jean.pierre.decorps@gmail.com>
Repository CRAN
Date/Publication 2021-06-07 15:30:02 UTC

R topics documented:

CC ................................................................. 2
CCInter .......................................................... 4
CCTable ........................................................... 5
crossTable ...................................................... 7
CS ................................................................. 8
CSInter ........................................................... 10
CSTable ........................................................... 12
orderFactors ..................................................... 14
Tiramisu .......................................................... 15
CC

Univariate analysis of case control studies

Description

CC is used with case-control studies to determine the association between an exposure and an outcome. Note that all variables need to be numeric and binary and coded as "0" and "1". Point estimates and confidence intervals for the odds ratio are calculated, along with attributable or prevented fractions for the exposed and total population.

Additionally you can select if you want to display the Fisher's exact test, by specifying exact = TRUE.

If you specify full = TRUE you can easily access useful statistics from the output tables.

Usage

CC(data, cases, exposure, exact = FALSE, full = FALSE, title = "CC")

Arguments

data data.frame

cases character - Case variable

exposure character - Exposure variable

exact boolean - TRUE if you would like to display Fisher's exact p-value

full boolean - TRUE if you need to display useful statistics and values for formatting

title character - title of tables

Value

list:

df1 data.frame - two by two table

df2 data.frame - statistics

df1.align character - alignment for kable/xtable

df2.align character - alignment for kable/xtable

df1.digits integer vector - digit number displayed for kable/xtable

df2.digits integer vector - digit number displayed for kable/xtable

st list - individual statistics

The item st returns the odds ratio and its 95 percent confidence intervals, the attributable fraction among the exposed and its 95 percent confidence intervals, the attributable fraction among the population and its 95 percent confidence intervals, the Chi square value, the Chi square p-value and the Fisher's exact test p-value.
Note

You can use the lowercase command "cc" in place of "CC"

Please note also that when the outcome is frequent the odds ratio will overestimate the risk ratio (if OR>1) or underestimate the risk ratio (OR<1). If the outcome is rare, the risk ratio and the odds ratio are similar.

In a case control study, the attributable fraction among the exposed and among the population assume that the OR approximates the risk ratio.

Please interpret all measures with caution.

Author(s)

jean.pierre.decorps@gmail.com

References


See Also

CCTable, CCInter, CS, CSTable, CSInter

Examples

library(EpiStats)

# Dataset by Anja Hauri, RKI.
data(Tiramisu)
DF <- Tiramisu

# The CC command looks at the association between the outcome variable "ill"
# and an exposure "mousse"
CC(DF, "ill", "mousse")

# The option exact = TRUE provides Fisher's exact test p-values
CC(DF, "ill", "mousse", exact = TRUE)

# With the option full = TRUE you can easily use individual elements of the results:
result <- CC(DF, "ill", "mousse", full = TRUE)
result$st$odds_ratio$point_estimate
Description

CCInter is useful to determine the effects of a third variable on the association between an exposure and an outcome. CCInter produces 2 by 2 tables with stratum specific odds ratios, attributable risk among exposed and population attributable risk.

Note that the outcome and exposure variable need to be numeric and binary and coded as "0" and 1". The third variable needs to be numeric, but may have more categories, such as "0", "1" and "2".

Usage

CCInter(x, cases, exposure, by, table = FALSE, full = FALSE)

Arguments

x: data.frame
cases: string: case binary variable (0 / 1)
exposure: string: exposure binary variable (0 / 1)
by: string: stratifying variable (a factor)
table: boolean - TRUE if you need to display interaction table
full: boolean - TRUE if you need to display useful values for formatting

Details

CCInter is useful to determine the effects of a third variable on the association between an exposure and an outcome. CCInter produces 2 by 2 tables with stratum specific odds ratios, attributable risk among exposed and population attributable risk. Note that the outcome and exposure variable need to be numeric and binary and coded as "0" and 1". The third variable needs to be numeric, but may have more categories, such as "0", "1" and "2". CCInter displays a summary with the crude OR, the Mantel Haenszel adjusted OR and the result of a Woolf test for homogeneity of stratum-specific OR.

The option "full = TRUE" provides you with useful formatting information, which can be handy if you're using "markdown".

Value

list:
df1: data.frame - cross-table
df2: data.frame - statistics
df1.digits: integer vector - digit number displayed for kable/xtable
df1.align: character - alignment for kable/xtable
df2.digits: integer vector - digit number displayed for kable/xtable
df2.align: character - alignment for kable/xtable
Note
- You can use lowercas command "ccinter" instead of "CCInter" - The "by" variable (the stratifying variable) can have more than 2 levels

Author(s)
jean.pierre.decorps@gmail.com

References
ccinter for Stata by *Gilles Desve*

See Also
CC, CCTable

Examples
library(EpiStats)
data(Tiramisu)
DF <- Tiramisu

# Here you can see the association between wmousse and ill for each stratum of tira:
CCInter(DF, "ill", "wmousse", by = "tira")

# By storing the results in the object "res", you can use individual elements of the results.
# For example if you would like to view just the Mantel-Haenszel odds ratio for beer adjusted
# for tportion, you can view it by typing:
res <- CCInter(DF, "ill", "beer", "tportion", full = TRUE)
res$df2$Stats[3]

CCTable

Summary table for univariate analysis of case control studies

Description
CCTable is used for univariate analysis of case control studies with several exposures. The results are summarised in one table with one row per exposure making comparisons between exposures easier and providing a useful table for integrating into reports. Note that all variables need to be numeric and binary and coded as "0" and "1".

The results of this function contain: The name of exposure variables, the total number of cases, the number of exposed cases, the percentage of exposed among cases, the number of controls, the number of exposed controls, the percentage of exposed among controls, odds ratios, 95%CI intervals, p-values.
You can optionally choose to display the Fisher’s exact p-value instead of the Chi squared p-value, with the option `exact = TRUE`.

You can specify the sort order, with the option `sort = "or"` to order by odds ratios. The default sort order is by p-values.

The option `full = TRUE` provides you with useful formatting information, which can be handy if you’re using “markdown”.

**Usage**

```r
CCTable(x, cases, exposure = c(), exact = FALSE, sort = "pvalue", full = FALSE)
```

**Arguments**

- `x` : data.frame
- `cases` : character - cases binary variable (0 / 1)
- `exposure` : character vector - exposure variables
- `exact` : boolean - TRUE if you want the Fisher’s exact p-value instead of CHI2
- `sort` : character - [pvalue, or, pe] sort by pvalue (default) or by odds ratio, or by percent exposed
- `full` : boolean - TRUE if you need to display useful values for formatting

**Details**

The results of this function contain: The name of exposure variables, the total number of cases, the number of exposed cases, the percentage of exposed among cases, the number of controls, the number of exposed controls, the percentage of exposed among controls, odds ratios, 95%CI intervals, p-values.

You can optionally choose to display the Fisher???s exact p-value instead of the Chi squared p-value, with the option `exact = TRUE`.

You can specify the sort order, with the option `sort = "or"` to order by odds ratios. The default sort order is by p-values.

The option "full = TRUE" provides you with useful formatting information, which can be handy if you’re using “markdown”.

**Value**

```r
list :
  df : data.frame - results table
  digits : integer vector - digit number displayed for kable/xtable
  align : character - alignment for kable/xtable
```

**Note**

- You can use the lowercase command "cctable" instead of "CCTable"
crossTable

Author(s)
jean.pierre.decorps@gmail.com

References
cctable for Stata by *Gilles Desve* and *Peter Makary*.

See Also
CC, CCInter

Examples
library(EpiStats)
data(Tiramisu)
df <- Tiramisu

# You can see the association between several exposures and being ill.
cctable(df, "ill", exposure=c("sex", "wmousse", "tira", "beer", "mousse"))

# By storing results in res, you can also use individual elements of the results.
# For example if you would like to view a particular odds ratio,
# you can view it by typing (for example):
res = CCTable(df, "ill", exposure = c("sex", "wmousse", "tira", "beer", "mousse"), exact=TRUE)
res$df$\'Odds Ratio\'[1]

---

crossTable  
contingency table of 2 variables

Description

Creates a contingency table of 2 variables. Percentages are optionals by row, column or both. It can provides an optional statistic (Fisher or Chisquare).

Usage
crossTable(data, var1, var2, percent="none", statistic="none")

Arguments
data  
data.frame
var1  
character - first varname - can be unquoted
var2  
character - second varname - can be unquoted
percent  
character - "none" (default) or ("row", "col", "both") - can be unquoted
statistic  
character - "none" (default) or ("fisher", "chi2") - can be unquoted
Value

data.frame - contingency table

Author(s)

jean.pierre.decorps@gmail.com

See Also

orderFactors, CC, CS

Examples

library(EpiStats)

# Dataset by Anja Hauri, RKI.
data(Tiramisu)
DF <- Tiramisu

# Table with percentages and statistic on ordered factors
DF %<>%
  orderFactors(ill, values = c(1,0), labels = c("YES", "NO")) %>%
  orderFactors(sex, values = c("males", "females"), labels = c("Males", "Females"))

crossTable(DF, "ill", "sex", "both", "chi2")

CS

Univariate analysis of cohort study measuring risk

Description

CS analyses cohort studies with equal follow-up time per subject. The risk (the proportion of
individuals who become cases) is calculated overall and among the exposed and unexposed. Note
that all variables need to be numeric and binary and coded as "0" and "1".

Point estimates and confidence intervals for the risk ratio and risk difference are calculated, along
with attributable or preventive fractions for the exposed and the total population.

Additionally you can select if you want to display the Fisher's exact test, by specifying exact =
TRUE.

If you specify full = TRUE you can easily access useful statistics from the output tables.

Usage

CS(x, cases, exposure, exact = F, full = FALSE, title = "CS")
Arguments

x data.frame
cases character - Case variable
exposure character - Exposure variable
exact boolean - TRUE if you would like to display Fisher's exact p-value
full boolean - TRUE if you need to display useful statistics and values for formatting
title character - title of tables

Value

list:

df1 data.frame - two by two table
df2 data.frame - statistics
st list - individual statistics
df1.digits integer vector - digit number displayed for kable/xtable
df2.digits integer vector - digit number displayed for kable/xtable
df2.align character - alignment for kable/xtable

The item st returns the risk difference and its 95 percent confidence intervals, the risk ratio and its 95 percent confidence intervals, the attributable fraction among the exposed and its 95 percent confidence intervals, the attributable fraction among the population and its 95 percent confidence intervals, the Chi square value, the Chi square p-value and the Fisher's exact test p-value.

Note

You can use the lowercase command "cs" in place of "CS"

Author(s)

jean.pierre.decorps@gmail.com

References


See Also

CSTable, CSInter, CC, CCTable, CCInter
Examples

library(EpiStats)

# Dataset by Anja Hauri, RKI.
# Dataset provided with package.
data(Tiramisu)
DF <- Tiramisu

# The CS command looks at the association between the outcome variable "ill"
# and an exposure "mousse"
CS(DF, "ill", "mousse")

# The option exact = TRUE provides Fisher's exact test p-values
CS(DF, "ill", "mousse", exact = TRUE)

# With the option full = TRUE you can easily use individual elements of the results:
result <- CS(DF, "ill", "mousse", full = TRUE)
result$st$risk_ratio$point_estimate

CSInter

Stratified analysis for cohort studies measuring risk

Description

CSInter is useful to determine the effects of a third variable on the association between an exposure and an outcome. CSInter produces 2 by 2 tables with stratum specific risk ratios, attributable risk among exposed and population attributable risk. Note that the outcome and exposure variable need to be numeric and binary and coded as "0" and "1". The third variable needs to be numeric, but may have more categories, such as "0", "1" and "2".

Usage

CSInter(x, cases, exposure, by, table = FALSE, full = FALSE)

Arguments

x data.frame
cases string: illness binary variable (0 / 1)
exposure string: exposure binary variable (0 / 1)
by string: stratifying variable (a factor)
table boolean - TRUE if you need to display interaction table
full boolean - TRUE if you need to display useful values for formatting
Details

CSInter is useful to determine the effects of a third variable on the association between an exposure and an outcome. CSInter produces 2 by 2 tables with stratum specific risk ratios, attributable risk among exposed and population attributable risk. Note that the outcome and exposure variable need to be numeric and binary and coded as "0" and 1". The third variable needs to be numeric, but may have more categories, such as "0", "1" and "2".

CSInter displays a summary with the crude RR, the Mantel Haenszel adjusted RR and the result of a "Woolf" test for homogeneity of stratum-specific RR.

The option full = TRUE provides you with useful formatting information, which can be handy if you're using "markdown".

Value

list:

- df1: data.frame - cross-table
- df2: data.frame - statistics
- df1.digits: integer vector - digit number displayed for kable/xtable
- df2.digits: integer vector - digit number displayed for kable/xtable

Note

- You can use the lowercase command "csinter" instead of "CSInter" - The "by" variable (the stratifying variable) can have more than 2 levels

Author(s)

jean.pierre.decorps@gmail.com

References

csinter for Stata by *Gilles Desve*

See Also

CS, CSTable

Examples

library(EpiStats)
data(Tiramisu)
DF <- Tiramisu

# Here you can see the association between wmosse and ill for each stratum of tira:
csinter(DF, "ill", "wmousse", by = "tira")

# By storing the results in the object "res", you can use individual elements
# of the results. For example if you would like to view just the Mantel-Haenszel
CSTable

Summary table for univariate analysis of cohort studies measuring risk

Description

CSTable is used for univariate analysis of cohort studies with several exposures. The results are summarised in one table with one row per exposure making comparisons between exposures easier and providing a useful table for integrating into reports. Note that all variables need to be numeric and binary and coded as "0" and "1".

The results of this function contain: The name of exposure variables, the total number of exposed, the number of exposed cases, the attack rate among the exposed, the total number of unexposed, the number of unexposed cases, the attack rate among the unexposed, risk ratios, 95% percent confidence intervals, and p-values.

You can optionally choose to display the Fisher’s exact p-value instead of the Chi squared p-value, with the option exact = TRUE.

You can specify the sort order, with the option sort="rr" to order by risk ratios. The default sort order is by p-values.

The option full = TRUE provides you with useful formatting information, which can be handy if you’re using "markdown".

Usage

CSTable(x, cases, exposure = c(), exact = FALSE, sort = "pvalue", full = FALSE)

Arguments

x data.frame

cases string - variable containing cases (binary 0 / 1)

exposure string vector - names of variables containing exposure (binary 0 / 1)

exact boolean - TRUE if you want the Fisher’s exact p-value instead of CHI2

sort character - [pvalue, rr, ar] sort by pvalue (default) or by risk ratio, or by percent of attributable risk

full boolean - TRUE if you need to display useful values for formatting
Details
The results of this function contain: The name of exposure variables, the total number of exposed, the number of exposed cases, the attack rate among the exposed, the total number of unexposed, the number of unexposed cases, the attack rate among the unexposed, risk ratios, 95
You can optionally choose to display the Fisher’s exact p-value instead of the Chi squared p-value, with the option exact = TRUE.
You can specify the sort order, with the option sort="rr" to order by risk ratios. The default sort order is by p-values.
The option full = TRUE provides you with useful formatting information, which can be handy if you’re using "markdown".

Value
list :
  df data.frame - results table
digits integer vector - digit number displayed for kable/xtable
align character - alignment for kable/xtable

Note
- You can use the lowercase command "cstable" instead of "CSTable"

Author(s)
jean.pierre.decorps@gmail.com

References
cstable for Stata by *Gilles Desve* and *Peter Makary*

See Also
CS, CSInter

Examples
library(EpiStats)
data(Tiramisu)
df <- Tiramisu

# You can see the association between several exposures and being ill.
CSTable(df, "ill", exposure=c("sex", "wmousse", "tira", "beer", "mousse"))

# By storing results in res, you can also use individual elements of the results.
# For example if you would like to view a particular risk ratio,
# you can view it by typing (for example):
res = CSTable(df, "ill", exposure = c("sex", "wmousse", "tira", "beer", "mousse"), exact=TRUE)
orderFactors

Generates ordered factors.

Description
Generates ordered factors for a list of columns by name or by index or range.

Usage

orderFactors(data, ..., values, labels=NULL)

Arguments

data
  data.frame

... character - first varname - can be unquoted

values
  character - second varname - can be unquoted

labels
  character - NULL (default) or ("row", "col", "both") - can be unquoted

Value
data.frame - contingency table

Author(s)
jean.pierre.decorps@gmail.com

See Also
crossTable

Examples

library(EpiStats)

# Dataset by Anja Hauri, RKI.
data(Tiramisu)
DF <- Tiramisu

# Table with percentages and statistic on ordered factors
DF %<>%
  orderFactors(ill, values = c(1,0), labels = c("YES", "NO")) %>%
  orderFactors(sex, values = c("males", "females"), labels = c("Males", "Females"))
crossTable(DF, "ill", "sex", "both", "chi2")
Tiramisu

**A foodborne disease outbreak dataset**

**Description**

The dataset available with the EpiStats package is from an outbreak investigation carried out in Germany in 1998 by Anja Hauri, Robert Koch Institute.

**Usage**

`data(Tiramisu)`

**Format**

A data frame with 291 observations with the following 21 variables.

- **ill**: a numeric vector
- **dateonset**: a date
- **sex**: a factor with levels females males
- **age**: a numeric vector
- **tira**: a numeric vector
- **tportion**: a numeric vector
- **wmousse**: a numeric vector
- **dmousse**: a numeric vector
- **mousse**: a numeric vector
- **mportion**: a numeric vector
- **beer**: a numeric vector
- **uniquekey**: a numeric vector
- **redjelly**: a numeric vector
- **fruitsalad**: a numeric vector
- **tomato**: a numeric vector
- **mince**: a numeric vector
- **salmon**: a numeric vector
- **horseradish**: a numeric vector
- **chickenwin**: a numeric vector
- **roastbeef**: a numeric vector
- **pork**: a numeric vector

**References**

The dataset available with the EpiStats package is from an outbreak investigation carried out in Germany in 1998 by Anja Hauri, Robert Koch Institute. It is used in case studies by organisations including EPIET, ECDC and EpiConcept. It is provided with this package with Anja’s permission.
Examples

data(Tiramisu)

## maybe str(Tiramisu) ; plot(Tiramisu) ...
Index

* ~documentation
  CC, 2
  CCInter, 4
  CCTable, 5
  crossTable, 7
  CS, 8
  CSInter, 10
  CSTable, 12
  orderFactors, 14

* datasets
  Tiramisu, 15

CC, 2
cc (CC), 2
CCInter, 4
ccinter (CCInter), 4
CCTable, 5
cctable (CCTable), 5
crossTable, 7
CS, 8
cs (CS), 8
CSInter, 10
csinter (CSInter), 10
CSTable, 12
cstable (CSTable), 12

orderFactors, 14

Tiramisu, 15