Package ‘konfound’

April 12, 2024

Type Package

Title Quantify the Robustness of Causal Inferences

Version 0.5.1

Description Statistical methods that quantify the conditions necessary to alter inferences, also known as sensitivity analysis, are becoming increasingly important to a variety of quantitative sciences. A series of recent works, including Frank (2000) <doi:10.1177/0049124100029002001> and Frank et al. (2013) <doi:10.3102/0162373713493129> extend previous sensitivity analyses by considering the characteristics of omitted variables or unobserved cases that would change an inference if such variables or cases were observed. These analyses generate statements such as "an omitted variable would have to be correlated at xx with the predictor of interest (e.g., treatment) and outcome to invalidate an inference of a treatment effect". Or "one would have to replace pp percent of the observed data with null hypothesis cases to invalidate the inference". We implement these recent developments of sensitivity analysis and provide modules to calculate these two robustness indices and generate such statements in R. In particular, the functions konfound(), pkonfound() and mkonfound() allow users to calculate the robustness of inferences for a user's own model, a single published study and multiple studies respectively.

License MIT + file LICENSE

URL https://github.com/konfound-project/konfound

BugReports https://github.com/konfound-project/konfound/issues

Depends R (>= 2.10)

Imports broom, broom.mixed, crayon, dplyr, ggplot2, lavaan, purrr, rlang, tidyr, lme4 (>= 1.1-35.1), tibble, ggrepel, pbkrtest

Suggests covr, devtools, forcats, knitr, rmarkdown, mice, roxygen2, testthat, Matrix (>= 1.6-2)

VignetteBuilder knitr

Encoding UTF-8
LazyData  true
RoxygenNote  7.3.1
NeedsCompilation  no
Author  Joshua M Rosenberg [aut, cre],
        Ran Xu [ctb],
        Qinyun Lin [ctb],
        Spiro Maroulis [ctb],
        Sarah Narvaiz [ctb],
        Kenneth A Frank [ctb],
        Wei Wang [ctb]
Maintainer  Joshua M Rosenberg <jmrosen48@gmail.com>
Repository  CRAN
Date/Publication  2024-04-12 21:10:09 UTC

R topics documented:

  binary_dummy_data ........................................... 3
  cal_delta_star ........................................... 3
  cal_rxy .................................................. 4
  cal_rxz .................................................. 4
  cal_ryz .................................................. 5
  chisq_p .................................................. 5
  concord1 .................................................. 6
  get_kr_df ................................................ 6
  konfound ................................................. 7
  konfound_glm ............................................. 8
  konfound_glm_dichotomous .................................... 9
  konfound_lm ............................................. 10
  konfound_lmer .......................................... 11
  mkonfound ............................................... 12
  mkonfound_ex .......................................... 12
  output_df ............................................... 13
  output_print ............................................ 14
  output_table ........................................... 15
  pkonfound ............................................... 16
  plot_correlation ......................................... 18
  plot_threshold .......................................... 19
  tkonfound ............................................... 19
  tkonfound_fig ........................................... 20
  verify_reg_Gzcv ......................................... 21
  verify_reg_uncond ...................................... 22
  zzz ....................................................... 22

Index  23
binary_dummy_data

| binary_dummy_data | Binary dummy data |

Description

This data is made-up data for use in examples.

Format

A data.frame with 107 rows and 2 variables.

cal_delta_star

| cal_delta_star | Calculate delta star for sensitivity analysis |

Description

Calculate delta star for sensitivity analysis

Usage

cal_delta_star(
    FR2max,
    R2,
    R2_uncond,
    est_eff,
    eff_thr,
    var_x,
    var_y,
    est_uncond,
    rxz,
    n_obs
)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FR2max</td>
<td>maximum R2</td>
</tr>
<tr>
<td>R2</td>
<td>current R2</td>
</tr>
<tr>
<td>R2_uncond</td>
<td>unconditional R2</td>
</tr>
<tr>
<td>est_eff</td>
<td>estimated effect</td>
</tr>
<tr>
<td>eff_thr</td>
<td>effect threshold</td>
</tr>
<tr>
<td>var_x</td>
<td>variance of X</td>
</tr>
<tr>
<td>var_y</td>
<td>variance of Y</td>
</tr>
<tr>
<td>est_uncond</td>
<td>unconditional estimate</td>
</tr>
<tr>
<td>rxz</td>
<td>correlation coefficient between X and Z</td>
</tr>
<tr>
<td>n_obs</td>
<td>number of observations</td>
</tr>
</tbody>
</table>
Value
delta star value

cal_rxy
Calculate rxy based on rxyGz, rxz, and ryz

Description
Calculate rxy based on rxyGz, rxz, and ryz

Usage
cal_rxy(ryxGz, rxz, ryz)

Arguments
ryxGz correlation coefficient between Y and X given Z
rxz correlation coefficient between X and Z
ryz correlation coefficient between Y and Z

Value
rxy value

cal_rxz
Calculate R2xz based on variances and standard error

Description
Calculate R2xz based on variances and standard error

Usage
cal_rxz(var_x, var_y, R2, df, std_err)

Arguments
var_x variance of X
var_y variance of Y
R2 coefficient of determination
df degrees of freedom
std_err standard error

Value
R2xz value
**cal_ryz**

*Calculate R2yz based on ryxGz and R2*

**Description**

Calculate R2yz based on ryxGz and R2

**Usage**

```
cal_ryz(ryxGz, R2)
```

**Arguments**

- **ryxGz**: correlation coefficient between Y and X given Z
- **R2**: coefficient of determination

**Value**

R2yz value

---

**chisq_p**

*Perform a Chi-Square Test*

**Description**

‘chisq_p’ calculates the p-value for a chi-square test given a contingency table.

**Usage**

```
chisq_p(a, b, c, d)
```

**Arguments**

- **a**: Frequency count for row 1, column 1.
- **b**: Frequency count for row 1, column 2.
- **c**: Frequency count for row 2, column 1.
- **d**: Frequency count for row 2, column 2.

**Value**

P-value from the chi-square test.
**get_kr_df**

**Concord1 data**

**Description**

This data is from Hamilton (1983)

**Format**

A data.frame with 496 rows and 10 variables.

**References**


---

**get_kr_df**

*Extract Degrees of Freedom for Fixed Effects in a Linear Mixed-Effects Model*

**Description**

Extract Degrees of Freedom for Fixed Effects in a Linear Mixed-Effects Model

**Usage**

```r
get_kr_df(model_object)
```

**Arguments**

- `model_object`: The mixed-effects model object produced by lme4::lmer.

**Value**

A vector containing degrees of freedom for the fixed effects in the model.
konfound

Konfound Analysis for Various Model Types

Description

Performs sensitivity analysis on fitted models including linear models (‘lm’), generalized linear models (‘glm’), and linear mixed-effects models (‘lmerMod’). It calculates the amount of bias required to invalidate or sustain an inference, and the impact of an omitted variable necessary to affect the inference.

Usage

konfound(
  model_object,
  tested_variable,
  alpha = 0.05,
  tails = 2,
  index = "RIR",
  to_return = "print",
  two_by_two = FALSE,
  n_treat = NULL,
  switch_trm = TRUE,
  replace = "control"
)

Arguments

model_object A model object produced by ‘lm’, ‘glm’, or ‘lme4::lmer’.
tested_variable Variable associated with the coefficient to be tested.
alpha Significance level for hypothesis testing.
tails Number of tails for the test (1 or 2).
index Type of sensitivity analysis (‘RIR’ by default).
to_return Type of output to return (‘print’, ‘raw_output’, ‘table’).
two_by_two Boolean; if ‘TRUE’, uses a 2x2 table approach for ‘glm’ dichotomous variables.
n_treat Number of treatment cases (used only if ‘two_by_two’ is ‘TRUE’).
switch_trm Boolean; switch treatment and control in the analysis.
replace Replacement method for treatment cases (‘control’ by default).

Value

Depending on ‘to_return’, prints the result, returns a raw output, or a summary table.
Examples

# using lm() for linear models
m1 <- lm(mpg ~ wt + hp, data = mtcars)
konfound(m1, wt)
konfound(m1, wt, to_return = "table")

# using glm() for non-linear models
if (requireNamespace("forcats")) {
  d <- forcats::gss_cat
  d$married <- ifelse(d$marital == "Married", 1, 0)
  m2 <- glm(married ~ age, data = d, family = binomial(link = "logit"))
  konfound(m2, age)
}

# using lme4 for mixed effects (or multi-level) models
if (requireNamespace("lme4")) {
  library(lme4)
  m3 <- fm1 <- lme4::lmer(Reaction ~ Days + (1 | Subject), sleepstudy)
  konfound(m3, Days)
}

m4 <- glm(outcome ~ condition, data = binary_dummy_data, family = binomial(link = "logit"))
konfound(m4, condition, two_by_two = TRUE, n_treat = 55)

konfound_glm

Konfound Analysis for Generalized Linear Models

Description

This function performs konfound analysis on a generalized linear model object. It uses 'broom' to tidy model outputs and calculates the sensitivity of inferences. It supports analysis for a single variable or multiple variables.

Usage

konfound_glm(
  model_object,
  tested_variable_string,
  alpha,
  tails,
  index = "RIR",
  to_return
)
konfound_glm_dichotomous

Arguments

model_object The model object produced by glm.
tested_variable_string The name of the variable being tested.
alpha Significance level for hypothesis testing.
tails Number of tails for the test (1 or 2).
index Type of sensitivity analysis (‘RIR’ by default).
to_return The type of output to return.

Value

The results of the konfound analysis for the specified variable(s).

konfound_glm_dichotomous

Konfound Analysis for Generalized Linear Models with Dichotomous Outcomes

Description

This function performs konfound analysis on a generalized linear model object with a dichotomous outcome. It uses ‘broom’ to tidy model outputs and calculates the sensitivity of inferences.

Usage

konfound_glm_dichotomous(
    model_object,
    tested_variable_string,
    alpha,
    tails,
    to_return,
    n_treat,
    switch_trm,
    replace
)

Arguments

model_object The model object produced by glm.
tested_variable_string The name of the variable being tested.
alpha Significance level for hypothesis testing.
tails Number of tails for the test (1 or 2).
to_return The type of output to return.
konfound_lm

konfound lm

Konfound Analysis for Linear Models

Description

This function performs konfound analysis on a linear model object produced by lm. It calculates the sensitivity of inferences for coefficients in the model. It supports analysis for a single variable or multiple variables.

Usage

konfound_lm(
  model_object,
  tested_variable_string,
  alpha,
  tails,
  index,
  to_return
)

Arguments

model_object 
  The linear model object produced by lm.
tested_variable_string 
  The name of the variable being tested.
alpha 
  Significance level for hypothesis testing.
tails 
  Number of tails for the test (1 or 2).
index 
  Type of sensitivity analysis (‘RIR’ by default).
to_return 
  The type of output to return.

Value

The results of the konfound analysis for the specified variable(s).
konfound_lmer

Konfound Analysis for Linear Mixed-Effects Models

Description

This function performs konfound analysis on a linear mixed-effects model object produced by lme4::lmer. It calculates the sensitivity of inferences for fixed effects in the model. It supports analysis for a single variable or multiple variables.

Usage

konfound_lmer(
  model_object,
  tested_variable_string,
  test_all,
  alpha,
  tails,
  index,
  to_return
)

Arguments

model_object The mixed-effects model object produced by lme4::lmer.
tested_variable_string The name of the fixed effect being tested.
test_all Boolean indicating whether to test all fixed effects or not.
alpha Significance level for hypothesis testing.
tails Number of tails for the test (1 or 2).
index Type of sensitivity analysis (‘RIR’ by default).
to_return The type of output to return.

Value

The results of the konfound analysis for the specified fixed effect(s).
Description

Performs sensitivity analysis for multiple models, where parameters are stored in a data frame. It calculates the amount of bias required to invalidate or sustain an inference for each case in the data frame.

Usage

```
mkonfound(d, t, df, alpha = 0.05, tails = 2, return_plot = FALSE)
```

Arguments

- `d`: A data frame or tibble containing t-statistics and associated degrees of freedom.
- `t`: Column name or vector of t-statistics.
- `df`: Column name or vector of degrees of freedom associated with t-statistics.
- `alpha`: Significance level for hypothesis testing.
- `tails`: Number of tails for the test (1 or 2).
- `return_plot`: Whether to return a plot of the percent bias (default is `FALSE`).

Value

Depending on `return_plot`, either returns a data frame with analysis results or a plot.

Examples

```r
## Not run:
mkonfound_ex
str(d)
mkonfound(mkonfound_ex, t, df)
## End(Not run)
```

Description

A dataset containing t and df values from example studies from Educational Evaluation and Policy Analysis (as detailed in Frank et al., 2013): https://drive.google.com/file/d/1aGhxGjvMvEPVAgoA8rxvA97uU05TTMe/view
output_df

Usage

mkonfound_ex

Format

A data frame with 30 rows and 2 variables:

- **t**  t value
- **df**  degrees of freedom associated with the t value ...

Source

https://drive.google.com/file/d/1aGhxGjvMvEPVAg0A8rrxVA97uU05TTMe/view

<table>
<thead>
<tr>
<th>output_df</th>
<th>Output data frame based on model estimates and thresholds</th>
</tr>
</thead>
</table>

Description

Output data frame based on model estimates and thresholds

Usage

output_df(
  est_eff,
  beta_threshold,
  unstd_beta,
  bias = NULL,
  sustain = NULL,
  recase,
  obs_r,
  critical_r,
  r_con,
  itcv,
  non_linear
)

Arguments

- **est_eff**  estimated effect
- **beta_threshold**  threshold for beta
- **unstd_beta**  unstandardized beta value
- **bias**  bias to change inference
- **sustain**  sustain to change inference
- **recase**  number of cases to replace null
output_print

<table>
<thead>
<tr>
<th>obs_r</th>
<th>observed correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>critical_r</td>
<td>critical correlation</td>
</tr>
<tr>
<td>r_con</td>
<td>correlation for omitted variable</td>
</tr>
<tr>
<td>itcv</td>
<td>inferential threshold for confounding variable</td>
</tr>
<tr>
<td>non_linear</td>
<td>flag for non-linear models</td>
</tr>
</tbody>
</table>

**Value**

data frame with model information

---

**Description**

This function outputs printed text for various indices such as RIR (Robustness of Inference to Replacement) and IT (Impact Threshold for a Confounding Variable) with specific formatting like bold, underline, and italic using functions from the crayon package. It handles different scenarios based on the effect difference, beta threshold, and other parameters, providing formatted output for each case.

**Usage**

```r
output_print(
  eff_diff,
  beta_threshold,
  bias = NULL,
  sustain = NULL,
  nu,
  recase,
  obs_r,
  critical_r,
  r_con,
  itcv,
  alpha,
  index
)
```

**Arguments**

- `eff_diff` The difference in the effect size being evaluated.
- `beta_threshold` The threshold value of beta, used for statistical significance determination.
- `bias` The percentage of the estimate that could be due to bias (optional).
- `sustain` The percentage of the estimate necessary to sustain an inference (optional).
output_table

nu The hypothesized effect size used in replacement analysis.
recase The number of cases that need to be replaced to change the inference.
obs_r The observed correlation coefficient in the data.
critical_r The critical correlation coefficient for statistical significance.
r_con The correlation coefficient of an omitted variable with both the outcome and the predictor.
itcv The impact threshold for a confounding variable.
alpha The level of statistical significance.
index A character string indicating the index for which the output is generated ('RIR' or 'IT').

---

**Description**

This function takes a model object and the tested variable, tidies the model output using `broom::tidy`, calculates the impact threshold for confounding variables (ITCV) and impact for each covariate, and returns a rounded, tidy table of model outputs.

**Usage**

```r
output_table(model_object, tested_variable)
```

**Arguments**

- `model_object` A model object from which to generate the output.
- `tested_variable` The variable being tested in the model.

**Value**

A tidy data frame containing model outputs, ITCV, and impacts for covariates.
pkonfound

Perform sensitivity analysis for published studies

Description

For published studies, this command calculates (1) how much bias there must be in an estimate to invalidate/sustain an inference; (2) the impact of an omitted variable necessary to invalidate/sustain an inference for a regression coefficient.

Usage

pkonfound(
est_eff,
std_err,
n_obs,
n_covariates = 1,
alpha = 0.05,
tails = 2,
index = "RIR",
nu = 0,
n_treat = NULL,
switch_trm = TRUE,
model_type = "ols",
a = NULL,
b = NULL,
c = NULL,
d = NULL,
two_by_two_table = NULL,
test = "fisher",
replace = "control",
sdx,
sdy,
R2,
eff_thr = 0,
FR2max,
FR2max_multiplier = 1.3,
to_return = "print"
)

Arguments

est_eff the estimated effect (such as an unstandardized beta coefficient or a group mean difference)
std_err the standard error of the estimate of the unstandardized regression coefficient
n_obs the number of observations in the sample
n_covariates the number of covariates in the regression model
alpha
probability of rejecting the null hypothesis (defaults to 0.05)
tails
integer whether hypothesis testing is one-tailed (1) or two-tailed (2; defaults to 2)
index
whether output is RIR or IT (impact threshold); defaults to "RIR"
u
what hypothesis to be tested; defaults to testing whether est_eff is significantly different from 0
n_treat
the number of cases associated with the treatment condition; applicable only when model_type = "logistic"
switch_trm
whether to switch the treatment and control cases; defaults to FALSE; applicable only when model_type = "logistic"
model_type
the type of model being estimated; defaults to "ols" for a linear regression model; the other option is "logistic"
a
cell is the number of cases in the control group showing unsuccessful results
b
cell is the number of cases in the control group showing successful results
c
cell is the number of cases in the treatment group showing unsuccessful results
d
cell is the number of cases in the treatment group showing successful results
two_by_two_table
table that is a matrix or can be coerced to one (data.frame, tibble, tribble) from which the a, b, c, and d arguments can be extracted
test
whether using Fisher’s Exact Test or A chi-square test; defaults to Fisher’s Exact Test
replace
whether using entire sample or the control group to calculate the base rate; default is the control group
sdx
the standard deviation of X
sdy
the standard deviation of Y
R2
the unadjusted, original R2 in the observed function
eff_thr
unstandardized coefficient threshold to change an inference
FR2max
the largest R2, or R2max, in the final model with unobserved confounder
FR2max_multiplier
the multiplier of R2 to get R2max, default is set to 1.3
to_return
whether to return a data.frame (by specifying this argument to equal "raw_output" for use in other analyses) or a plot ("plot"); default is to print ("print") the output to the console; can specify a vector of output to return

Value

prints the bias and the number of cases that would have to be replaced with cases for which there is no effect to invalidate the inference
Examples

# using pkonfound for linear models
pkonfound(2, .4, 100, 3)
pkonfound(-2.2, .65, 200, 3)
pkonfound(.5, 3, 200, 3)

pkonfound(-0.2, 0.103, 20888, 3, n_treat = 17888, model_type = "logistic")

pkonfound(2, .4, 100, 3, to_return = "thresh_plot")
pkonfound(2, .4, 100, 3, to_return = "corr_plot")

# using pkonfound for a 2x2 table
pkonfound(a = 35, b = 17, c = 17, d = 38)
pkonfound(a = 35, b = 17, c = 17, d = 38, alpha = 0.01)
pkonfound(a = 35, b = 17, c = 17, d = 38, alpha = 0.01, switch_trm = FALSE)
pkonfound(a = 35, b = 17, c = 17, d = 38, test = "chisq")

# use pkonfound to calculate delta* and delta_exact
pkonfound(est_eff = .4, std_err = .1, n_obs = 290, sdx = 2, sdy = 6, R2 = .7,
    eff_thr = 0, FR2max = .8, index = "COP", to_return = "raw_output")
# use pkonfound to calculate rxcv and rycv when preserving standard error
pkonfound(est_eff = .5, std_err = .056, n_obs = 6174, eff_thr = .1,
    sdx = 0.22, sdy = 1, R2 = .3, index = "PSE", to_return = "raw_output")

---

plot_correlation  
Plot Correlation Diagram

Description

This function creates a plot to illustrate the correlation between different variables, specifically focusing on the confounding variable, predictor of interest, and outcome. It uses ggplot2 for graphical representation.

Usage

plot_correlation(r_con, obs_r, critical_r)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>r_con</td>
<td>Correlation coefficient related to the confounding variable.</td>
</tr>
<tr>
<td>obs_r</td>
<td>Observed correlation coefficient.</td>
</tr>
<tr>
<td>critical_r</td>
<td>Critical correlation coefficient for decision-making.</td>
</tr>
</tbody>
</table>

Value

A ggplot object representing the correlation diagram.
plot_threshold

Plot Effect Threshold Diagram

Description
This function creates a plot to illustrate the threshold of an effect estimate in relation to a specified beta threshold. It uses ggplot2 for graphical representation.

Usage
plot_threshold(beta_threshold, est_eff)

Arguments
beta_threshold The threshold value for the effect.
est_eff The estimated effect size.

Value
A ggplot object representing the effect threshold diagram.

tkonfound

Perform Sensitivity Analysis on 2x2 Tables

Description
This function performs a sensitivity analysis on a 2x2 contingency table. It calculates the number of cases that need to be replaced to invalidate or sustain the statistical inference. The function also allows switching between treatment success and failure or control success and failure based on the provided parameters.

Usage
tkonfound(a, b, c, d, alpha = 0.05, switch_trm = TRUE, test = "fisher", replace = "control", to_return = to_return)
tkonfound_fig

Arguments

a  Number of unsuccessful cases in the control group.
b  Number of successful cases in the control group.
c  Number of unsuccessful cases in the treatment group.
d  Number of successful cases in the treatment group.
alpha  Significance level for the statistical test, default is 0.05.
switch_trm  Boolean indicating whether to switch treatment row cells, default is TRUE.
test  Type of statistical test to use, either "fisher" (default) or "chisq".
replace  Indicates whether to use the entire sample or the control group for base rate calculation, default is "control".
to_return  Type of output to return, either "raw_output" or "print".

Value

Returns detailed information about the sensitivity analysis, including the number of cases to be replaced (RIR), user-entered table, transfer table, and conclusions.

tkonfound_fig

Draw Figures for Change in Effect Size in 2x2 Tables

Description

This function generates plots illustrating how the change in effect size is influenced by switching or replacing outcomes in a 2x2 table. It produces two plots: one showing all possibilities (switching) and another zoomed in the area for positive RIR (Relative Impact Ratio).

Usage

tkonfound_fig(
  a,
  b,
  c,
  d,
  thr_p = 0.05,
  switch_trm = TRUE,
  test = "fisher",
  replace = "control"
)
Arguments

a  Number of cases in the control group with unsuccessful outcomes.
b  Number of cases in the control group with successful outcomes.
c  Number of cases in the treatment group with unsuccessful outcomes.
d  Number of cases in the treatment group with successful outcomes.
thr_p  P-value threshold for statistical significance, default is 0.05.
switch_trm  Whether to switch the two cells in the treatment or control row, default is TRUE (treatment row).
test  Type of statistical test used, either "Fisher’s Exact Test" (default) or "Chi-square test".
replace  Indicates whether to use the entire sample or just the control group for calculating the base rate, default is "control".

Value

Returns two plots showing the effect of hypothetical case switches on the effect size in a 2x2 table.
verify_reg_uncond  Verify unconditional regression model

Description
Verify unconditional regression model

Usage
verify_reg_uncond(n_obs, sdx, sdy, rxy)

Arguments
n_obs  number of observations
sdx    standard deviation of X
sdy    standard deviation of Y
rxy    correlation coefficient between X and Y

Value
list of model parameters

Package Initialization Functions and Utilities

Description
These functions are used for initializing the package environment and providing utility functions for the package.
Index

* datasets
  mkonfound_ex, 12

binary_dummy_data, 3

cal_delta_star, 3
cal_rxy, 4
cal_rxz, 4
cal_ryz, 5
chisq_p, 5
concord1, 6

get_kr_df, 6

konfound, 7
konfound_glm, 8
konfound_glm_dichotomous, 9
konfound_lm, 10
konfound_lmer, 11

mkonfound, 12
mkonfound_ex, 12

output_df, 13
output_print, 14
output_table, 15

pkonfound, 16
plot_correlation, 18
plot_threshold, 19

tkonfound, 19
tkonfound_fig, 20

verify_reg_Gzcv, 21
verify_reg_uncond, 22

zzz, 22