Package ‘riskdiff’

June 3, 2025

Title Robust Risk Difference Estimation with Multiple Link Functions

Version 0.1.0

Date 2025-05-29

Description Calculates risk differences (or prevalence differences for cross-sectional data) using generalized linear models with automatic link function selection. Provides robust model fitting with fallback methods, support for stratification and adjustment variables, and publication-ready output formatting. Handles model convergence issues gracefully and provides confidence intervals using multiple approaches. Methods are based on approaches described in Mark W. Donoghoe and Ian C. Marschner (2018) ``logbin: An R Package for Relative Risk Regression Using the Log-Binomial Model'' <doi:10.18637/jss.v086.i09> for robust GLM fitting, and standard epidemiological methods for risk difference estimation as described in Kenneth J. Rothman, Sander Greenland and Timothy L. Lash (2008, ISBN:9780781755641) ``Modern Epidemiology''.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.2

Depends R (>= 3.5.0)

Imports dplyr (>= 1.0.0), purrr, tibble, rlang, scales, stringr, stats

Suggests kableExtra, testthat (>= 3.0.0), knitr, rmarkdown, ggplot2, covr

Config/testthat/edition 3

URL `https://github.com/jackmurphy2351/riskdiff`

BugReports `https://github.com/jackmurphy2351/riskdiff/issues`

LazyData true

NeedsCompilation no

Author John D. Murphy [aut, cre] (ORCID: <`https://orcid.org/0000-0002-7714-9976`>, MPH, PhD)

Maintainer John D. Murphy <jackmurphy2351@gmail.com>

Repository CRAN

Date/Publication 2025-06-03 09:20:02 UTC

Contents

`birthweight` *Simulated Birth Weight Study Data*

Description

A simulated dataset containing information about maternal factors and birth weight outcomes, used for demonstrating risk difference calculations.

Usage

```
birthweight
```

Format

A data frame with 2,500 rows and 8 variables:

id Patient identifier (1 to 2500)

low_birthweight Binary outcome: 1 = low birth weight (<2500g), 0 = normal

smoking Maternal smoking status: "No" or "Yes"

maternal_age Maternal age in years (continuous, mean ~28)

race Maternal race: "White", "Black", "Hispanic", "Other"

education Education level: "Less than HS", "HS", "Some college", "College+"

prenatal_care Prenatal care adequacy: "Adequate", "Inadequate"

parity Number of previous births: 0, 1, 2, or 3+ (capped at 3)

Details

This dataset was simulated to reflect realistic associations between maternal factors and low birth weight risk. The relationships include:

- **Smoking**: Increases risk of low birth weight
- **Maternal age**: Modest association with risk
- **Race**: Health disparities reflected in different baseline risks
- **Education**: Higher education associated with lower risk
- **Prenatal care**: Adequate care reduces risk
- **Parity**: Higher parity associated with slightly increased risk

The base rate of low birth weight is approximately 8%, which is realistic for developed countries. The effect sizes and interactions were designed to demonstrate various analysis scenarios including stratification and adjustment.

Source

Simulated data based on patterns from epidemiological literature and the National Center for Health Statistics

Examples

```
data(birthweight)
head(birthweight)

# Basic descriptive statistics
table(birthweight$smoking, birthweight$low_birthweight)

# Summary by race
with(birthweight, table(race, low_birthweight))

# Simple risk difference
rd <- calc_risk_diff(birthweight, "low_birthweight", "smoking")
print(rd)

# Create a simple summary table
cat(create_simple_table(rd, "Risk of Low Birth Weight by Smoking Status"))
```

`calc_risk_diff` *Calculate Risk Differences with Robust Model Fitting*

Description

Calculates risk differences (or prevalence differences for cross-sectional data) using generalized linear models with identity, log, or logit links. Includes support for stratification, adjustment variables, and robust model fitting.

The function addresses common convergence issues with identity link binomial GLMs by implementing a fallback strategy across multiple link functions, similar to approaches described in Donoghoe & Marschner (2018) for relative risk regression.

Usage

```
calc_risk_diff(
  data,
  outcome,
  exposure,
  adjust_vars = NULL,
  strata = NULL,
  link = "auto",
  alpha = 0.05,
  verbose = FALSE
)
```

Arguments

`data`	A data frame containing all necessary variables
`outcome`	Character string naming the binary outcome variable (must be 0/1 or logical)
`exposure`	Character string naming the exposure variable of interest
`adjust_vars`	Character vector of variables to adjust for (default: NULL)
`strata`	Character vector of stratification variables (default: NULL)
`link`	Character string specifying link function: "auto", "identity", "log", or "logit" (default: "auto")
`alpha`	Significance level for confidence intervals (default: 0.05)
`verbose`	Logical indicating whether to print diagnostic messages (default: FALSE)

Details

Statistical Approach:

The function attempts to fit models in the following order:

1. **Identity link**: Directly estimates risk differences using a binomial GLM with identity link
2. **Log link**: Estimates relative risks, then transforms to risk differences via prediction
3. **Logit link**: Estimates odds ratios, then transforms to risk differences via prediction

Convergence Strategy:

Identity link models for binary outcomes often fail to converge due to parameter constraints (fitted probabilities must be \le 1). This package implements robust starting values and automatic fallback to alternative link functions when convergence fails, ensuring reliable estimation.

Risk Difference Interpretation:

Risk differences represent absolute changes in probability. A risk difference of 0.05 means the exposed group has a 5 percentage point higher risk than the unexposed group. This is often more interpretable than relative measures (risk ratios, odds ratios) for public health decision-making.

Value

A tibble of class "riskdiff_result" containing the following columns:

exposure_var Character. Name of exposure variable analyzed

rd Numeric. Risk difference estimate (proportion scale, e.g. 0.05 = 5 percentage points)

ci_lower Numeric. Lower bound of confidence interval

ci_upper Numeric. Upper bound of confidence interval

p_value Numeric. P-value for test of null hypothesis (risk difference = 0)

model_type Character. Link function successfully used ("identity", "log", "logit", or error type)

n_obs Integer. Number of observations used in analysis

... Additional columns for stratification variables if specified

The returned object has attributes including the original function call and alpha level used. Risk differences are on the probability scale where 0.05 represents a 5 percentage point difference.

References

Donoghoe MW, Marschner IC (2018). "logbin: An R Package for Relative Risk Regression Using the Log-Binomial Model." Journal of Statistical Software, 86(9), 1-22. doi:10.18637/jss.v086.i09

Rothman KJ, Greenland S, Lash TL (2008). Modern Epidemiology, 3rd edition. Lippincott Williams & Wilkins.

Examples

```
# Simple risk difference
data(birthweight)
rd_simple <- calc_risk_diff(
  data = birthweight,
  outcome = "low_birthweight",
  exposure = "smoking"
)
print(rd_simple)

# Age-adjusted risk difference
rd_adjusted <- calc_risk_diff(
  data = birthweight,
  outcome = "low_birthweight",
  exposure = "smoking",
  adjust_vars = "maternal_age"
)
print(rd_adjusted)

# Stratified analysis
```

```
rd_stratified <- calc_risk_diff(
  data = birthweight,
  outcome = "low_birthweight",
  exposure = "smoking",
  strata = "race"
)
print(rd_stratified)
```

`create_rd_table` *Create Formatted Table of Risk Difference Results*

Description

Creates a publication-ready table of risk difference results with appropriate grouping and formatting. Requires the kableExtra package for full functionality.

Usage

```
create_rd_table(
  results,
  caption = "Risk Differences",
  include_model_type = FALSE,
  ...
)
```

Arguments

`results`	Results tibble from calc_risk_diff()
`caption`	Table caption (default: "Risk Differences")
`include_model_type`	Whether to include model type column (default: FALSE)
`...`	Additional arguments passed to kableExtra::kable()

Value

If kableExtra is available, returns a kable table object suitable for rendering in R Markdown or HTML. The table includes formatted risk differences, confidence intervals, and p-values with appropriate styling and footnotes. If kableExtra is not available, returns a formatted tibble with the same information in a basic data frame structure.

Examples

```
data(birthweight)
results <- calc_risk_diff(birthweight, "low_birthweight", "smoking")

# Basic table (works without kableExtra)
basic_table <- create_rd_table(results, caption = "Risk of Low Birth Weight")
```

```
print(basic_table)

# Enhanced table (requires kableExtra)
if (requireNamespace("kableExtra", quietly = TRUE)) {
  enhanced_table <- create_rd_table(
    results,
    caption = "Risk of Low Birth Weight by Smoking Status",
    include_model_type = TRUE
  )
  print(enhanced_table)
}
```

`create_simple_table` *Create a Simple Summary Table*

Description

Creates a simple text-based summary table that doesn't require kableExtra.

Usage

```
create_simple_table(results, title = "Risk Difference Results")
```

Arguments

`results`	Results tibble from calc_risk_diff()
`title`	Optional title for the table

Value

A formatted character vector representing the table

Examples

```
data(birthweight)
results <- calc_risk_diff(birthweight, "low_birthweight", "smoking")
cat(create_simple_table(results))
```

`format_risk_diff` *Format Risk Difference Results for Display*

Description

Formats numerical values in risk difference results for presentation, with appropriate percentage formatting and rounding.

Usage

```
format_risk_diff(results, digits = 2, p_accuracy = 0.001)
```

Arguments

`results`	Results tibble from calc_risk_diff()
`digits`	Number of decimal places for percentages (default: 2)
`p_accuracy`	Accuracy for p-values (default: 0.001)

Value

Tibble with additional formatted columns

Examples

```
data(birthweight)
results <- calc_risk_diff(birthweight, "low_birthweight", "smoking")
formatted <- format_risk_diff(results)
print(formatted)
```

`print.riskdiff_result` *Print method for riskdiff_result objects*

Description

Prints risk difference results in a formatted, readable way showing key statistics including risk differences, confidence intervals, and model types used.

Usage

```
## S3 method for class 'riskdiff_result'
print(x, ...)
```

Arguments

`x`	A riskdiff_result object from calc_risk_diff()
`...`	Additional arguments passed to print methods

Value

Invisibly returns the original riskdiff_result object (x). Called primarily for its side effect of printing formatted results to the console.

Examples

```
data(birthweight)
result <- calc_risk_diff(birthweight, "low_birthweight", "smoking")
print(result)
```

Index