Package 'infectionIndex'

November 28, 2026

Type Package

Title Disease Infection Intensity and Progress Analysis Tools

Version 0.1.0

Description Analyzes plant disease intensity and its progression over time using the infection intensity index proposed by Czermainski (1999) <doi:10.1590/S0100-204X1999000900004> and the disease index associated with McKinney (1923) <https://babel.hathitrust.org/cgi/pt?id=uc1.b3610175>. Provides generalized estimating equation (GEE) models, area under the disease progress curve (AUDPC), area under the disease progress stairs (AUDPS), outlier detection, sensitivity analysis, estimated marginal means, temporal trend analysis, and graphical visualization.

Depends R (>= 4.1)

Imports readxl, dplyr, tidyr, tidyselect, ggplot2, geepack, emmeans, readr, stats, utils

Suggests testthat (>= 3.0.0), openxlsx, knitr, rmarkdown

License GPL-3

Encoding UTF-8

VignetteBuilder knitr

NeedsCompilation no

Author Carlos Augusto Dórea Bragança [aut, cre] (Universidade Federal do Recôncavo da Bahia (UFRB)), Malena Andrade Nogueira [aut] (Universidade Federal do Recôncavo da Bahia (UFRB))

Maintainer Carlos Augusto Dórea Bragança <cbraganca@ufrb.edu.br>

Config/roxygen2/version 8.1.1

RoxygenNote 8.1.1

Repository CRAN

Date/Publication 2026-11-28 15:00:13 UTC

Contents

amaral_czermainski_I 2
gee_di 3
mckinney_I 4
plot_curves_by_trat 5
run_di_pipeline 6

Index **8**

`amaral_czermainski_I` *Czermainski infection intensity index*

Description

Calculates the generalized infection intensity index proposed by Czermainski (1999) for observations classified into ordered disease classes.

Usage

```
amaral_czermainski_I(freq_by_class, return_omega = FALSE)
```

Arguments

`freq_by_class`	Numeric vector of non-negative absolute frequencies ordered from class 0 (absence of disease) to class J (highest disease intensity).
`return_omega`	Logical. If `FALSE` (default), returns only the infection intensity index I. If `TRUE`, returns both omega and I.

Details

Let F0, ..., FJ denote the absolute frequencies in J + 1 ordered disease classes, from absence of disease (F0) to maximum disease intensity (FJ). For j = 1, ..., J, the conditional frequency is

$$f_j = (F_j + ... + F_J)/(F_{j-1} + F_j + ... + F_J).$$

Let p{j-1,J} denote the cumulative relative frequency from class j - 1 through class J:

$$p_{j-1,J} = (F_{j-1} + ... + F_J)/n.$$

The weights k_j are defined so that their cumulative values satisfy

$$k_1 + ... + k_j = asin(sqrt(j/J))/(pi/2).$$

The angular quantity omega is calculated as

$$omega = sum_{j=1}^{J} k_j sqrt(p_{j-1,J}) asin(sqrt(f_j)),$$

and the infection intensity index is

$$I = sin^2(omega).$$

Calculations are performed in radians. Therefore, when `return_omega = TRUE`, the returned value of omega is in radians.

The order of `freq_by_class` is essential: frequencies must be supplied from the lowest disease class (F0) to the highest (FJ).

Value

A numeric infection intensity index between 0 and 1. If `return_omega = TRUE`, a named numeric vector containing `omega` (in radians) and `I`.

References

Czermainski, A. B. C. (1999). Generalizacao de um indice de intensidade de infeccao em experimentos de avaliacao de doencas em plantas. *Pesquisa Agropecuaria Brasileira*, 34(9), 1545-1555. doi:10.1590/S0100204X1999000900004

Examples

```
freq <- c(19, 9, 4, 5, 6, 7)

amaral_czermainski_I(freq)

amaral_czermainski_I(freq, return_omega = TRUE)
```

---

gee_di *Fit generalized estimating equation models for disease intensity*

Description

Fits generalized estimating equation (GEE) models to angular-transformed disease intensity. Compares a model without treatment-by-time interactions against a model including these interactions. Supports single-factor and factorial experimental designs.

Usage

```
gee_di(
  data_longo,
  corstr = c("ar1", "exchangeable", "independence", "unstructured"),
  do_anova = TRUE,
  factors = "trat"
)
```

Arguments

data_longo	Long-format data frame with time, DI, DI_clip, DI_ang, and columns named in factors. Must also contain unit or block to identify experimental units.
corstr	Working correlation structure: "ar1", "exchangeable", "independence", or "unstructured".
do_anova	If TRUE, compares treatment-by-time interactions. When the standard model comparison is unavailable, uses the package's Wald test.
factors	Character vector naming treatment-factor columns. The default "trat" preserves the single-factor interface. For factorial designs, for example, use c("cultivar", "produto").

Details

The experimental unit is identified by unit; if absent, it is derived from block and the combination of factors specified in factors.

Value

Invisibly returns a list containing gee0 (model without treatment-by-time interactions), gee1 (model with interactions), comp (comparison result when do_anova = TRUE, otherwise NULL), and predictions (prediction grid). When Wald inference is numerically unreliable, such as under near-zero dispersion, its statistic and p-value may be returned as NA rather than a spurious inference.

---

mckinney_I *McKinney disease index*

Description

Calculates the disease index described by McKinney from absolute frequencies distributed among ordered disease classes.

Usage

```
mckinney_I(freq_by_class, scores = NULL)
```

Arguments

freq_by_class	Numeric vector of non-negative absolute frequencies ordered from the lowest to the highest disease class.
scores	Optional numeric vector containing the scores assigned to the disease classes. It must have the same length as freq_by_class. If NULL, scores 0:J are used.

Details

The index is calculated as

$$I = sum(F_j x_j)/(n x_m ax),$$

where F_j is the absolute frequency in class j, x_j is the score assigned to that class, n is the total number of observations, and x_max is the maximum score.

If `scores = NULL`, equally spaced scores from 0 to J are used. The returned index is expressed on a 0 to 1 scale rather than as a percentage.

Value

A numeric disease index between 0 and 1.

References

McKinney, H. H. (1923). Influence of soil temperature and moisture on infection of wheat seedlings by Helminthosporium sativum. *Journal of Agricultural Research*, 26, 195–217. https://babel.hathitrust.org/cgi/pt?id=uc1.b3610175

Examples

```
freq <- c(19, 9, 4, 5, 6, 7)
mckinney_I(freq)

mckinney_I(
  c(10, 5, 3, 2, 1),
  scores = c(0, 0.75, 1, 2, 3)
)
```

---

plot_curves_by_trat *Plot treatment-specific curves over time*

Description

Plot treatment-specific curves over time

Usage

```
plot_curves_by_trat(data)
```

Arguments

`data`	Data frame containing the columns `trat`, `time`, and `y`.

Value

A ggplot object.

`run_di_pipeline` *Run disease intensity pipeline*

Description

Analyzes disease intensity and progression from an XLSX workbook in long or wide format. Computes the Czermainski or McKinney intensity index, angular transformation, generalized estimating equation (GEE) models, predictions, area under the disease progress curve (AUDPC) or area under the disease progress stairs (AUDPS), outlier detection, and sensitivity analysis with optional refitting.

Usage

```
run_di_pipeline(
  excel_file,
  sheet_prefer = NULL,
  index = c("czermainski", "mckinney"),
  corstr = c("ar1", "exchangeable", "independence", "unstructured"),
  progress_metric = c("audpc", "audps"),
  outdir = NULL,
  save_outputs = FALSE,
  outlier_action = c("flag", "remove_all"),
  factors = "trat"
)
```

Arguments

`excel_file`	Path to the `.xlsx` workbook.
`sheet_prefer`	Preferred worksheet name; automatically detected if `NULL`.
`index`	Disease intensity index: `"czermainski"` (default) or `"mckinney"`.
`corstr`	GEE working correlation structure: `"ar1"`, `"exchangeable"`, `"independence"`, or `"unstructured"`.
`progress_metric`	Disease progression metric used for summaries, outlier detection, and subsequent analyses: `"audpc"` (default) or `"audps"`.
`outdir`	Output directory for CSV, PNG, and RDS files. Must be explicitly provided when `save_outputs = TRUE`.
`save_outputs`	Logical. If `TRUE`, writes analysis results to disk. Defaults to `FALSE`; no files are written by default.
`outlier_action`	Action for detected outliers: `"flag"` (default) marks them, while `"remove_all"` performs a sensitivity analysis after their removal.
`factors`	Character vector naming treatment-factor columns. The default `"trat"` preserves the single-factor interface. For factorial designs, for example, use `c("cultivar", "produto")`.

Details

The design may contain one or several treatment factors. Experimental units are identified by `unit`; when absent, this identifier is derived from `block` and the combination of columns specified in `factors`.

Value

A list with `out` (analysis of the complete data), `out_clean` (sensitivity-analysis refit when outliers are removed), and `outliers` (information on detected and optionally removed experimental units). For legacy single-factor designs, nonparametric tests associated with the progression metric are also provided. These tests are not automatically computed for multiple treatment factors.

Index

amaral_czermainski_I, 2

gee_di, 3

mckinney_I, 4

plot_curves_by_trat, 5

run_di_pipeline, 6