

Package ‘ILRCM’

July 27, 2025

Type Package

Title Convert Irregular Longitudinal Data to Regular Intervals and
Perform Clustering

Version 0.2.0

Maintainer Atanu Bhattacharjee <atanustat@gmail.com>

Description Convert irregularly spaced longitudinal data into regular intervals for further analysis,
and perform clustering using advanced machine learning techniques.
The package is designed for handling complex longitudinal datasets,
optimizing them for research in healthcare, demography, and other fields
requiring temporal data modeling.

Imports ggplot2, scales, rlang, dplyr

License GPL-3

Encoding UTF-8

LazyData true

Depends R (>= 3.5.0)

RoxygenNote 7.3.2

NeedsCompilation no

Author Atanu Bhattacharjee [aut, cre, ctb],
Tanmoy Majumdar [aut, ctb],
Gajendra Kumar Vishwakarma [aut, ctb]

Repository CRAN

Date/Publication 2025-07-27 21:00:08 UTC

Contents

dropplot	2
err	3
intlen	5
irr	6
lrrr	8
sdata	10
smocc	11

Index**12**

dropplot	<i>Dropout Curve and Observation Distribution for Irregular Longitudinal Data</i>
----------	---

Description

This function generates a combined plot of a dropout curve and a histogram of observation counts over time. The dropout curve shows how many subjects remain in the study over time based on their last observation time. The histogram shows how the observations are distributed across time.

Usage

```
dropplot(data, id_col, time_col, bins = 100, percentile = 90)
```

Arguments

<code>data</code>	A data frame containing the longitudinal data.
<code>id_col</code>	A character string specifying the column name for subject identifiers.
<code>time_col</code>	A character string specifying the column name for the time variable.
<code>bins</code>	Number of bins for the histogram (default is 100).
<code>percentile</code>	A numeric value between 0 and 100 specifying the cutoff for the red dropout line (default is 90).

Value

A list with two elements:

- `plot`: A ggplot object showing the dropout curve and histogram.
- `data`: A data frame with mid-points of the time bins (`'mid_time'`) and the number of observations (`'count'`) in descending order.

Examples

```
## Not run:
data(smocc) # assumes smocc is loaded with columns id and age
result <- dropplot(data = smocc, id_col = "id", time_col = "age", bins = 60, percentile = 90)
print(result$plot)
head(result$data)

## End(Not run)
```

err	<i>Convert Irregular Longitudinal Data to Regular Intervals and Perform Clustering using Excluding Repeated Responses (ERR) method</i>
-----	--

Description

This function takes irregular longitudinal data and converts it into regularly spaced intervals using linear interpolation. It then computes the relative change in the response variable between consecutive time points, clusters the data based on these changes, and provides various visualizations of the process.

Usage

```
err(data, subject_id_col, time_col, response_col, rel, interval_length)
```

Arguments

data	A data frame containing the irregular longitudinal data.
subject_id_col	A character string representing the name of the column with the subject IDs.
time_col	A character string representing the name of the column with time values.
response_col	A character string representing the name of the column with the response values.
rel	Relative change method such as SRC, CARC and SWRC.
interval_length	A numeric value indicating the length of the regular intervals to which the time values should be converted.

Details

The `err` function handles irregular longitudinal data by:

- Interpolating response values at regular time intervals.
- Calculating the relative change in the response values across time points.
- Clustering subjects based on these relative changes using alphabet labels ("a", "b", ..., "h") corresponding to different levels of deviation from the mean.
- Resolving cluster ties using a sum of squares criterion.

Visualizations of the data include plots for both the original irregular data and the regularized data, as well as histograms of time distributions and relative change trends.

Value

A list containing:

- `regular_data`: A data frame of the regularized longitudinal data.
- `regular_data_wide`: A wide-format version of the regularized data.

- `relative_change`: A data frame containing the relative changes in response values.
- `cluster_data`: A data frame with cluster assignments for each subject at each time step.
- `cluster_data_reduced`: A reduced version of `cluster_data` with only subject IDs and their final cluster assignments.
- `merged_data`: The wide-format data merged with the final cluster assignments.
- `plot_irregular`: A ggplot object showing the original irregular data.
- `plot_regular`: A ggplot object showing the regularized data.
- `plot_change`: A ggplot object showing the relative changes over time.
- `histogram_irregular`: A ggplot object showing the histogram of irregular time distribution.
- `histogram_regular`: A ggplot object showing the histogram of regular time distribution.

Author(s)

author name

References

Reference

See Also

`intlen`, `irr`, `lrrr`

Examples

```
##
data(sdata)
sdata <- sdata[1:100,]
#Using relative change method: Simple relative change (SRC)
fit1 <- err(sdata, "subject_id", "time", "response", rel="SRC", interval_length = 3)
#for showing the regularized data in long format
fit1$regular_data
fit1$regular_data_wide #for showing the regularized data in wide format
fit1$cluster_data #dataset consisting clusters for different time points
fit1$merged_data #for showing the regularized data in wide format with final cluster
fit1$plot_regular #For plotting regularized longitudinal data
fit1$plot_irregular #For plotting irregular longitudinal data
fit1$plot_change #For plotting relative change
fit1$histogram_irregular #histogram for time of irregular data
fit1$histogram_regular #histogram for time of regular data
#Using relative change method: Cumulative average relative change (CARC)
fit2<-err(sdata,"subject_id","time","response",rel="CARC",interval_length=3)
fit2$regular_data #for showing the regularized data in long format
fit2$regular_data_wide #for showing the regularized data in wide format
fit2$cluster_data #dataset consisting clusters for different time points
fit2$merged_data #for showing the regularized data in wide format with final cluster
fit2$plot_regular #For plotting regularized longitudinal data
fit2$plot_irregular #For plotting irregular longitudinal data
```

```

fit2$plot_change #For plotting relative change
fit2$histogram_irregular #histogram for time of irregular data
fit2$histogram_regular #histogram for time of regular data
#Using relative change method: Weighted sum relative change (WSRC)
fit3 <- err(sdata, "subject_id", "time", "response", rel="WSRC", interval_length = 3)
fit3$regular_data #for showing the regularized data in long format
fit3$regular_data_wide #for showing the regularized data in wide format
fit3$cluster_data #dataset consisting clusters for different time points
fit3$merged_data #for showing the regularized data in wide format with final cluster
fit3$plot_regular #For plotting regularized longitudinal data
fit3$plot_irregular #For plotting irregular longitudinal data
fit3$plot_change #For plotting relative change
fit3$histogram_irregular #histogram for time of irregular data
fit3$histogram_regular #histogram for time of regular data

```

intlen	<i>Preferred Interval Length for Regularizing Irregular Longitudinal Data</i>
--------	---

Description

This function calculates the optimal interval length for regularizing irregular longitudinal data based on the given subject ID and time columns.

Usage

```
intlen(data, subject_col, time_col)
```

Arguments

data	A data frame containing the irregular longitudinal data.
subject_col	The column name for unique subject IDs.
time_col	The column name for time points.

Details

The function calculates the optimal interval length based on the observed range of time points and the average number of measurements per subject.

Value

Computed preferred interval length.

Examples

```

sdata <- sdata[1:100,]
intlen(sdata, "subject_id", "time")

```

irr	<i>Convert Irregular Longitudinal Data to Regular Intervals and Perform Clustering using Including Repeated Responses (IRR) method</i>
-----	--

Description

This function takes irregular longitudinal data and converts it into regularly spaced intervals using linear interpolation. It then computes the relative change in the response variable between consecutive time points, clusters the data based on these changes, and provides various visualizations of the process.

Usage

```
irr(data, subject_id_col, time_col, response_col, rel, interval_length)
```

Arguments

data	A data frame containing the irregular longitudinal data.
subject_id_col	A character string representing the name of the column with the subject IDs.
time_col	A character string representing the name of the column with time values.
response_col	A character string representing the name of the column with the response values.
rel	Relative change method such as SRC, CARC and SWRC.
interval_length	A numeric value indicating the length of the regular intervals to which the time values should be converted.

Details

The `irr` function handles irregular longitudinal data by:

- Interpolating response values at regular time intervals without replacing the last responses.
- Calculating the relative change in the response values across time points.
- Clustering subjects based on these relative changes using alphabet labels ("a", "b", ..., "h") corresponding to different levels of deviation from the mean.
- Resolving cluster ties using a sum of squares criterion.

Visualizations of the data include plots for both the original irregular data and the regularized data, as well as histograms of time distributions and relative change trends.

Value

A list containing:

- `regular_data`: A data frame of the regularized longitudinal data.
- `regular_data_wide`: A wide-format version of the regularized data.

- `relative_change`: A data frame containing the relative changes in response values.
- `cluster_data`: A data frame with cluster assignments for each subject at each time step.
- `cluster_data_reduced`: A reduced version of `cluster_data` with only subject IDs and their final cluster assignments.
- `merged_data`: The wide-format data merged with the final cluster assignments.
- `plot_irregular`: A ggplot object showing the original irregular data.
- `plot_regular`: A ggplot object showing the regularized data.
- `plot_change`: A ggplot object showing the relative changes over time.
- `histogram_irregular`: A ggplot object showing the histogram of irregular time distribution.
- `histogram_regular`: A ggplot object showing the histogram of regular time distribution.

Author(s)

author name

References

Reference

See Also

`intlen`, `err`, `lrrr`

Examples

```
##
data(sdata)
sdata <- sdata[1:100,]
#' #Using relative change method: Simple relative change (SRC)
fit1 <- irr(sdata, "subject_id", "time", "response", rel="SRC", interval_length = 3)
fit1$regular_data #for showing the regularized data in long format
fit1$regular_data_wide #for showing the regularized data in wide format
fit1$cluster_data #dataset consisting clusters for different time points
fit1$merged_data #for showing the regularized data in wide format with final cluster
fit1$plot_regular #For plotting regularized longitudinal data
fit1$plot_irregular #For plotting irregular longitudinal data
fit1$plot_change #For plotting relative change
fit1$histogram_irregular #histogram for time of irregular data
fit1$histogram_regular #histogram for time of regular data
#Using relative change method: Cumulative average relative change (CARC)
fit2 <- irr(sdata, "subject_id", "time", "response", rel="CARC", interval_length = 3)
fit2$regular_data #for showing the regularized data in long format
fit2$regular_data_wide #for showing the regularized data in wide format
fit2$cluster_data #dataset consisting clusters for different time points
fit2$merged_data #for showing the regularized data in wide format with final cluster
fit2$plot_regular #For plotting regularized longitudinal data
fit2$plot_irregular #For plotting irregular longitudinal data
fit2$plot_change #For plotting relative change
```

```

fit2$histogram_irregular #histogram for time of irregular data
fit2$histogram_regular #histogram for time of regular data
#Using relative change method: Weighted sum relative change (WSRC)
fit3 <- lrrr(sdata, "subject_id", "time", "response", rel="WSRC", interval_length = 3)
fit3$regular_data #for showing the regularized data in long format
fit3$regular_data_wide #for showing the regularized data in wide format
fit3$cluster_data #dataset consisting clusters for different time points
fit3$merged_data #for showing the regularized data in wide format with final cluster
fit3$plot_regular #For plotting regularized longitudinal data
fit3$plot_irregular #For plotting irregular longitudinal data
fit3$plot_change #For plotting relative change
fit3$histogram_irregular #histogram for time of irregular data
fit3$histogram_regular #histogram for time of regular data
##

```

lrrr

Convert Irregular Longitudinal Data to Regular Intervals and Perform Clustering using Linear Regression model for replacing Repeated Responses (LRRS) method

Description

This function takes irregular longitudinal data and converts it into regularly spaced intervals using linear interpolation. It then computes the relative change in the response variable between consecutive time points, clusters the data based on these changes, and provides various visualizations of the process.

Usage

```
lrrr(data, subject_id_col, time_col, response_col, rel, interval_length)
```

Arguments

data	A data frame containing the irregular longitudinal data.
subject_id_col	A character string representing the name of the column with the subject IDs.
time_col	A character string representing the name of the column with time values.
response_col	A character string representing the name of the column with the response values.
rel	Relative change method such as SRC, CARC and SWRC.
interval_length	A numeric value indicating the length of the regular intervals to which the time values should be converted.

Details

The `lrrr` function handles irregular longitudinal data by:

- Interpolating response values at regular time intervals and replaing the last responses using linear regression model.
- Calculating the relative change in the response values across time points.
- Clustering subjects based on these relative changes using alphabet labels ("a", "b", ..., "h") corresponding to different levels of deviation from the mean.
- Resolving cluster ties using a sum of squares criterion.

Visualizations of the data include plots for both the original irregular data and the regularized data, as well as histograms of time distributions and relative change trends.

Value

A list containing:

- `regular_data`: A data frame of the regularized longitudinal data.
- `regular_data_wide`: A wide-format version of the regularized data.
- `relative_change`: A data frame containing the relative changes in response values.
- `cluster_data`: A data frame with cluster assignments for each subject at each time step.
- `cluster_data_reduced`: A reduced version of `cluster_data` with only subject IDs and their final cluster assignments.
- `merged_data`: The wide-format data merged with the final cluster assignments.
- `plot_irregular`: A ggplot object showing the original irregular data.
- `plot_regular`: A ggplot object showing the regularized data.
- `plot_change`: A ggplot object showing the relative changes over time.
- `histogram_irregular`: A ggplot object showing the histogram of irregular time distribution.
- `histogram_regular`: A ggplot object showing the histogram of regular time distribution.

Author(s)

author name

References

Reference

See Also

`intlen`, `err`, `irr`

Examples

```
##
data(sdata)
sdata <- sdata[1:100,]
#Using relative change method: Simple relative change (SRC)
fit1 <- lrrr(sdata, "subject_id", "time", "response", rel="SRC", interval_length = 3)
fit1$regular_data #for showing the regularized data in long format
fit1$regular_data_wide #for showing the regularized data in wide format
fit1$cluster_data #dataset consisting clusters for different time points
fit1$merged_data #for showing the regularized data in wide format with final cluster
fit1$plot_regular #For plotting regularized longitudinal data
fit1$plot_irregular #For plotting irregular longitudinal data
fit1$plot_change #For plotting relative change
fit1$histogram_irregular #histogram for time of irregular data
fit1$histogram_regular #histogram for time of regular data
#Using relative change method: Cumulative average relative change (CARC)
fit2 <- lrrr(sdata, "subject_id", "time", "response", rel="CARC", interval_length = 3)
fit2$regular_data #for showing the regularized data in long format
fit2$regular_data_wide #for showing the regularized data in wide format
fit2$cluster_data #dataset consisting clusters for different time points
fit2$merged_data #for showing the regularized data in wide format with final cluster
fit2$plot_regular #For plotting regularized longitudinal data
fit2$plot_irregular #For plotting irregular longitudinal data
fit2$plot_change #For plotting relative change
fit2$histogram_irregular #histogram for time of irregular data
fit2$histogram_regular #histogram for time of regular data
#Using relative change method: Weighted sum relative change (WSRC)
fit3 <- lrrr(sdata, "subject_id", "time", "response", rel="WSRC", interval_length = 3)
fit3$regular_data #for showing the regularized data in long format
fit3$regular_data_wide #for showing the regularized data in wide format
fit3$cluster_data #dataset consisting clusters for different time points
fit3$merged_data #for showing the regularized data in wide format with final cluster
fit3$plot_regular #For plotting regularized longitudinal data
fit3$plot_irregular #For plotting irregular longitudinal data
fit3$plot_change #For plotting relative change
fit3$histogram_irregular #histogram for time of irregular data
fit3$histogram_regular #histogram for time of regular data
##
```

sdata

Simulated Irregular Longitudinal Data

Description

Simulated irregular longitudinal data for 1000 patients. This dataset contains irregularly spaced time points and responses for analysis.

Usage

```
data(sdata)
```

Format

A data frame with 8631 rows and 3 variables:

subject_id ID of subjects

time Irregular time points.

response Response values at different time points.

Examples

```
data(sdata)
head(sdata)
```

smocc

SMOCC Data

Description

Longitudinal height and weight measurements during ages 0-2 years for a representative sample of 1933 Dutch children born in 1988-1989. The dataset smocc is the sample of 200 subjects from the full dataset.

Usage

```
data(smocc)
```

Format

A data frame with 1942 rows and 7 variables:

id ID, unique id of each child (numeric)

age Decimal age, 0-2.68 years (numeric)

sex Sex, "male" or "female" (character)

ga Gestational age, completed weeks (numeric)

bw Birth weight in grammes (numeric)

hgt Height measurement in cm (numeric)

hgt_z Height in SDS relative Fourth Dutch Growth Study 1997 (numeric)

Examples

```
data(smocc)
head(smocc)
```

Index

* **datasets**

sdata, [10](#)

smocc, [11](#)

dropplot, [2](#)

err, [3](#)

intlen, [5](#)

irr, [6](#)

lrrr, [8](#)

sdata, [10](#)

smocc, [11](#)