Package 'missr'

June 4, 2025

Title Classify Missing Data as MCAR, MAR, or MNAR

Version 1.0.1

Maintainer Noah William Trelawny Hellen <`noahhellen@gmail.com`>

Description Classify missing data as missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR). This step is required before handling missing data (e.g. mean imputation) so that bias is not introduced. See Little (1988) <doi:10.1080/01621459.1988.10478722> for the statistical rationale for the methods used.

License MIT + file LICENSE

URL `https://github.com/NoahHellen/missr`,
`https://noahhellen.github.io/missr/`

BugReports `https://github.com/NoahHellen/missr/issues`

Depends R (>= 3.5)

Imports norm, tibble, lifecycle

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

Encoding UTF-8

Language en-GB

LazyData true

RoxygenNote 7.3.2

NeedsCompilation no

Author Noah William Trelawny Hellen [aut, cre, cph]

Repository CRAN

Date/Publication 2025-06-04 11:20:01 UTC

Contents

`animalhealth` *Simulated animal health data (MCAR)*

Description

A toy dataset with heart rate data for various animals.

Usage

```
animalhealth
```

Format

A 200 x 2 data frame:

animal The animal of interest

hear_rate The corresponding heart rate of the animal (bpm)

`companydata` *Simulated company data (MNAR)*

Description

A toy dataset with typical company metrics across various firms.

Usage

```
companydata
```

Format

A 500 x 5 data frame:

sales Sales in the last fiscal year (USD, million)

marketing_spend Marketing spend in last fiscal year (USD, million)

product_rating Average rating across all products

employees Total employee count in last fiscal year

gross_profit Gross profit in last fiscal year (USD, million)

`healthcheck` *Simulated health check data (MAR)*

Description

A toy dataset with typical health check-up metrics for various individuals.

Usage

```
healthcheck
```

Format

A 200 x 5 data frame:

bone_mass Bone mass of individual (kg)

body_fat Body fat percentage of individual

height Height of individual (cm)

age Age of individual

rbc Red blood cell count of individual (million/mm^3)

`mar` *Missing at random (MAR) test*

Description

[Stable] `mar()` performs multiple logistic regressions to test for MAR. The null hypothesis for each is that the data are not MAR.

Usage

```
mar(data, debug = FALSE)
```

Arguments

`data`	A data frame.
`debug`	A logical value used only for unit testing.

Details

In the following, each column of M with missing data is regressed on D_obs. Each regression produces a vector of p-values (one for each variable in D_obs). The smallest p-value is the most important. This is because missing data need only be dependent on one observed variable for the data to be MAR. If each reported smallest p-value is significant, the data is MAR. See `vignette("background")` for definitions of M and D_obs.

Value

A `tibble::tibble()`:

`missing`	Column of M with missing data
`p_value`	Smallest p-value of the logistic regressions
`explanatory`	Variable corresponding to `p_value`
`p_values`	The p-values of the logistic regressions
`variables`	Variables corresponding to `p_values`
`combined`	Paired `p_values` and `variables` for easier interpretation

Examples

```
mar(healthcheck)
```

`mcar` *Little's missing completely at random (MCAR) test*

Description

[Stable] `mcar()` performs Little's MCAR test to test for MCAR. The null hypothesis is that the data is MCAR.

Usage

```
mcar(data, debug = FALSE)
```

Arguments

`data`	A data frame.
`debug`	A logical value used only for unit testing.

Details

This function reproduces the d^2 statistic in equation (5) from [1]. This statistic is used to test for MCAR. Comments reference variables from `vignette("background")` (in brackets) to improve readability and traceability.

Value

A `tibble::tibble()`:

`statistic`	The d^2 statistic
`degrees_freedom`	Degrees of freedom of chi-squared distribution
`p_val`	P-value of the test
`missing_patterns`	Number of missing patterns

Note

Code is adapted from `mcar_test()` from the naniar package using base R instead of the tidyverse.

References

[1] Little RJA. A Test of Missing Completely at Random for Multivariate Data with Missing Values. Journal of the American Statistical Association. 1988;83(404):1198-202.

Examples

```
mcar(pollutionlevels)
```

`mnar` *Missing not at random (MNAR) classification*

Description

[Stable] `mnar()` presents the statistics from `mar()` and `mcar()`. If at least one p-value in `mar()` is not significant, and the p-value in `mcar()` is significant then the data is MNAR.

Usage

```
mnar(data)
```

Arguments

`data`	A data frame

Details

There exists no formal test for MNAR data. This function therefore presents the statistics for the tests in `mar()` and `mcar()`. If the results suggest the data is neither MAR nor MCAR, one can use process of elimination to deduce that the data is MNAR.

Value

A list:

`mcar`	Results of Little's MCAR test
`mar`	Results of MAR test

Examples

```
mnar(companydata)
```

`pollutionlevels` *Simulated pollution level data (MCAR)*

Description

A toy dataset with typical pollution level metrics for various settlements.

Usage

```
pollutionlevels
```

Format

A 200 x 4 data frame:

light Light pollution of settlement (mag/arcsec^2)

visual Visual pollution of settlement (VPI)

noise Noise pollution of settlement (dB)

air Air pollution of settlement (AQI)

`testscores` *Simulated test scores data*

Description

A toy dataset with test scores of various students.

Usage

```
testscores
```

Format

A 200 x 2 data frame:

id The ID of the student

score The student's score in the test

Index