Package: hdrm (via r-universe)

October 10, 2026

Title Inference for High-Dimensional Repeated Measures

Version 1.0.0

Description Provides one-group and multiple-group tests for expectation vectors in high-dimensional repeated-measures designs. Hypotheses can be specified through projection matrices or by selecting predefined effects. The multiple-group procedures cover heterogeneous and equal covariance matrices. Centered and standardized chi-square approximations are combined with exact or subsampling-based trace estimators. The functions return p-values, test statistics, estimated degrees of freedom, and convergence parameters. For further details, see Sattler and Hichert (2025) <doi:10.48550/arXiv.2512.17478>.

URL https://github.com/Schnieboli/hdrm

BugReports https://github.com/Schnieboli/hdrm/issues

License GPL (>= 3)

Encoding UTF-8

Suggests knitr, testthat (>= 3.0.0)

Config/testthat/edition 3

Depends R (>= 3.5.0)

LazyData true

Imports Rcpp, stats, withr

LinkingTo Rcpp, RcppArmadillo

Config/roxygen2/version 8.1.0

NeedsCompilation yes

Author Nils Hichert [aut, cre, cph], Paavo Sattler [aut] (ORCID: <https://orcid.org/0000-0001-8731-0893>), Markus Pauly [ctb], Edgar Brunner [ctb], David Ellenberger [ctb]

Maintainer Nils Hichert <nils.hichert@tu-dortmund.de>

Repository https://cran.r-universe.dev

Date/Publication 2026-10-10 13:52:38 UTC

RemoteUrl https://github.com/cran/hdrm

RemoteRef HEAD

RemoteSha 1eea8adc2411e388245bf20f76886272d93b850e

Contents

birthrates 2
EEG 3
hdrm_grouped 4
hdrm_single 8

Index **11**

`birthrates` *Total fertility rates in the German federal states, 1990–2023*

Description

Annual total fertility rates for the 16 German federal states. The entries give the number of children per woman reported in table 12612-09 of the German Federal Statistical Office (Statistisches Bundesamt (Destatis), 2024).

Usage

```
birthrates
```

Format

A data frame with 34 rows and 16 variables. Rows represent the years 1990 through 2023 and columns represent the German federal states.

Source

Statistisches Bundesamt (Destatis), table 12612-09.

References

Statistisches Bundesamt (Destatis) (2024). "Statistischer Bericht – Geburten 2023, Tabelle 12612-09." Zusammengefasste Geburtenziffer nach Bundesländern (Kinder je Frau). https://www.statistischebibliothek.de/mir/recei

EEG *EEG measurements from 160 subjects*

Description

A data frame containing four quantitative EEG variables measured at ten scalp regions for each of 160 subjects. The resulting 40 repeated-measurement dimensions are stored in long format. The data originate from a study conducted by Höller et al. (2017) and are a part of the package HRM described by Happ et al. (2018)

Usage

```
EEG
```

Format

A data frame with 6,400 rows and 7 variables:

group Diagnostic group with levels "SCC+", "SCC-", "MCI", and "AD".

value Numeric EEG-derived measurement.

sex Recorded sex, encoded as "M" or "W".

subject Subject identifier.

variable EEG variable coded from 1 to 4: activity, complexity, mobility, and brain rate.

region Scalp region coded from 1 to 10: frontal, central, temporal, occipital, and parietal, each measured on the left and right.

dimension Combined variable-region dimension coded from 1 to 40.

The documentation was taken from Happ et al. (2018).

References

Happ M, Harrar SW, Bathke AC (2018). "HRM: An R Package for Analysing High-dimensional Multi-factor Repeated Measures." The R Journal, 10(1), 534–548. doi:10.32614/RJ-2018-032.

Höller Y, Bathke AC, Uhl A, Strobl N, Lang A, Bergmann J, Nardone R, Rossini F, Zauner H, Kirschner M, Jahanbekam A, Trinka E, Staffen W (2017). "Combining SPECT and Quantitative EEG Analysis for the Automated Differential Diagnosis of Disorders with Amnestic Symptoms." Frontiers in Aging Neuroscience, 9, 290. doi:10.3389/fnagi.2017.00290.

hdrm_grouped *Multiple-group inference for high-dimensional repeated measures*

Description

Implements the multiple-group procedure allowing heterogeneous covariance matrices described by Sattler and Pauly (2018) and the multiple-group procedure under equal covariance matrices described by Sattler (2021).

Usage

```
hdrm_grouped(
  data,
  hypothesis = "whole",
  group,
  cov.equal = FALSE,
  subsampling = FALSE,
  B = "1000*N",
  AM = TRUE,
  seed = NULL
)
```

Arguments

data	A matrix or a data.frame. For matrix input, subjects are represented by rows and repeated-measurement dimensions by columns. For data.frame input, data must contain columns `value`, `subject` and `dimension`, containing the observations and the subject and dimension IDs.
hypothesis	Either one of `"whole"`, `"sub"`, `"interaction"`, `"identical"`, or `"flat"`, or a named list containing projection matrices `TW` and `TS`; see Details.
group	A one-dimensional atomic vector or factor defining the group allocation. For matrix input it must contain one entry per row. For vector input it must contain one entry per measurement.
cov.equal	A single logical value specifying whether the group covariance matrices are assumed to be equal. The default is `FALSE`.
subsampling	A single logical value specifying whether the subsampling versions of all available trace estimators are used in the heterogeneous-covariance procedure; see Details.
B	A single numeric value or arithmetic character expression in `N` defining the subsampling budget. Character expressions may contain only numeric constants, `N`, parentheses, and the operators `+`, `-`, `*`, `/`, and `^`. Its interpretation depends `subsampling`; see Details.
AM	A single logical value, specifying whether the compact representation of the hypothesis matrices described by Sattler and Rosenbaum (2025) is used. It may reduce the number of rows used in the calculations without changing the resulting test. The default is `TRUE`.

seed — NULL or a single integer-valued number used to make stochastic calculations reproducible. When supplied, the seed is applied locally and the previous R random-number state is restored after the calculation.

Details

At least two groups and two repeated-measurement dimensions are required. Every group must contain at least six subjects. Missing values in data or group are not allowed and will result in an error.

The tested hypothesis has the form

$$(\boldsymbol{T}_W \otimes \boldsymbol{T}_S)\boldsymbol{\mu} = \mathbf{0}.$$

The predefined hypotheses are:

- "whole": $\boldsymbol{T}_W = \boldsymbol{P}_a$ and $\boldsymbol{T}_S = \boldsymbol{J}_d/d$; no whole-plot or group main effect.
- "sub": $\boldsymbol{T}_W = \boldsymbol{J}_a/a$ and $\boldsymbol{T}_S = \boldsymbol{P}_d$; no subplot or dimension main effect.
- "interaction": $\boldsymbol{T}_W = \boldsymbol{P}_a$ and $\boldsymbol{T}_S = \boldsymbol{P}_d$; no group-by-dimension interaction.
- "identical": $\boldsymbol{T}_W = \boldsymbol{P}_a$ and $\boldsymbol{T}_S = \boldsymbol{I}_d$; identical expectation vectors across groups.
- "flat": $\boldsymbol{T}_W = \boldsymbol{I}_a$ and $\boldsymbol{T}_S = \boldsymbol{P}_d$; a flat expectation profile in every group.

Alternatively, hypothesis may be a named list containing TW and TS. Both matrices must be finite, symmetric, idempotent projection matrices with positive rank. TW must have one row and column per analyzed group, and TS must have one row and column per repeated-measurement dimension. Small numerical deviations within the implemented tolerance are accepted.

For data.frame input, observations are sorted according to subject, dimension and group (in that order) and then converted to a matrix with the first subject in the first row. This should be taken into account, when passing a list with custom matrices to hypothesis.

When cov.equal = FALSE, the method of Sattler and Pauly (2018) is used. The third-trace estimator entering f is always computed using subsampling. Here, B is the base subsampling budget. Group-specific and pairwise subsampling estimators use B draws for each group or group pair, respectively. The joint third-trace estimator uses aB joint draws, where each draw simultaneously samples six subjects from every group and a is the number of groups. When subsampling = TRUE, the remaining available trace estimators are also replaced by their subsampling versions.

When cov.equal = TRUE, the method described in Sattler (2021) is used. The pooled third-trace estimator uses an exact total of aB six-subject draws across all groups. These draws are allocated approximately proportionally to $\binom{n_i}{6}$ using the largest-remainder method, with every group receiving at least one draw.

Even when subsampling = FALSE, f, tau, and p.value remain seed dependent because a third-trace quantity is estimated by subsampling. For heterogeneous covariance matrices with subsampling = TRUE, statistic is seed dependent as well.

B may be numeric or an arithmetic character expression involving N, such as "1000*N" or "10*(N + 1)". The result is rounded up to the next integer. Functions, assignments, indexing, and additional variable names are rejected and are never evaluated.

Upper-tail probabilities are computed directly. Reported p-values are bounded below by .Machine$double.eps; a returned value at this boundary should be interpreted as no larger than the numerical reporting threshold.

Value

A named list of class `"hdrm_grouped"` with components:

data The processed data matrix with subjects in rows, repeated-measurement dimensions in columns, and subjects ordered by group.

dim A named vector containing the repeated-measurement dimension `d`, the number of analyzed subjects `N` and the number of analyzed groups `a`.

H A named list containing the projection matrices `TW` and `TS` used for the test and `label` with the hypothesis label.

group The sorted input to `group`.

cov.equal The input value for `cov.equal`.

subsampling The input value for `subsampling`.

B The evaluated integer base budget `B`. The grouped third-trace estimators use aB draws; see Details.

AM The input value for `AM`.

seed The input value for `seed`

statistic The standardized test statistic W.

p.value The upper-tail p-value, bounded below by `.Machine$double.eps`.

f The estimated degrees of freedom.

tau The estimated convergence parameter $\tau = 1/f$.

References

Sattler P (2021). "A comprehensive treatment of quadratic-form-based inference in repeated measures designs under diverse asymptotics." Electronic Journal of Statistics, 15(1), 3611–3634. doi:10.1214/21-EJS1865.

Sattler P, Pauly M (2018). "Inference for high-dimensional split-plot-designs: A unified approach for small to large numbers of factor levels." Electronic Journal of Statistics, 12(2), 2743–2805. doi:10.1214/18-EJS1465.

Sattler P, Rosenbaum M (2025). "Choice of the hypothesis matrix for using the Anova-type-statistic." Statistics & Probability Letters, 219, 110356. doi:10.1016/j.spl.2025.110356.

Examples

```
# Long-format data ------------------------------------------------------

data("EEG")
summary(EEG)

# Heterogeneous covariance matrices; the available exact trace estimators are
# used, while the third-trace quantity is still estimated by subsampling.
hdrm_grouped(
  data = EEG,
  hypothesis = "whole",
  group = EEG$group,
```

```
  cov.equal = FALSE,
  subsampling = FALSE,
  B = "100*N",
  seed = 3141
)

# Replace the remaining available trace estimators by their subsampling
# versions. Here B is used by each subsampling-estimator invocation.
hdrm_grouped(
  data = EEG,
  hypothesis = "sub",
  group = EEG$group,
  cov.equal = FALSE,
  subsampling = TRUE,
  B = 10000,
  seed = 3141
)

# A custom projection-matrix hypothesis equivalent to hypothesis = "sub".
custom_hypothesis <- list(
  TW = matrix(1 / 4, nrow = 4, ncol = 4),
  TS = diag(40) - matrix(1 / 40, nrow = 40, ncol = 40)
)

hdrm_grouped(
  data = EEG,
  hypothesis = custom_hypothesis,
  group = EEG$group,
  B = "100*N",
  seed = 3141
)


# Wide matrix data ----------------------------------------------------------

data("birthrates")

birthrates_matrix <- t(as.matrix(birthrates))

group <- factor(
  c(1, 1, 2, 2, 1, 1, 1, 2, 1, 1, 1, 1, 2, 2, 1, 2),
  labels = c("west", "east")
)

hdrm_grouped(
  data = birthrates_matrix,
  hypothesis = "interaction",
  group = group,
  cov.equal = FALSE,
  subsampling = FALSE,
  B = "100*N",
  seed = 3141
)
```

```
# Under equal covariance matrices, B is the total pooled subsampling budget
# across groups and the value of subsampling is ignored.
hdrm_grouped(
  data = birthrates_matrix,
  hypothesis = "whole",
  group = group,
  cov.equal = TRUE,
  B = "100*N",
  seed = 3141
)

d <- ncol(birthrates_matrix)
custom_birthrates_hypothesis <- list(
  TW = matrix(1 / 2, nrow = 2, ncol = 2),
  TS = diag(d) - matrix(1 / d, nrow = d, ncol = d)
)

hdrm_grouped(
  data = birthrates_matrix,
  hypothesis = custom_birthrates_hypothesis,
  group = group,
  B = "100*N",
  seed = 3141
)
```

`hdrm_single` *One-group inference for high-dimensional repeated measures*

Description

Implements the one-group test described by Pauly et al. (2015).

Usage

```
hdrm_single(data, hypothesis = "flat", AM = TRUE)
```

Arguments

`data`	A matrix or a data.frame. For matrix input, subjects are represented by rows and repeated-measurement dimensions by columns. For data.frame input, data must contain columns `value`, `subject` and `dimension`, giving the observations and the subject and dimension IDs.
`hypothesis`	Either `"flat"` or a finite numeric projection matrix whose dimensions equal the repeated-measurement dimension.
`AM`	A single logical value, specifying whether the compact representation of the hypothesis matrix described by Sattler and Rosenbaum (2025) is used. It may reduce the number of rows used in the calculations without changing the resulting test. The default is `TRUE`.

Details

At least two repeated-measurement dimensions are required. Missing values in `data` are not allowed and will result in an error.

The predefined value `"flat"` tests

$$\boldsymbol{P}_d\boldsymbol{\mu} = \mathbf{0}.$$

Alternatively, `hypothesis` may be a finite numeric $d \times d$ projection matrix, where d is the number of dimensions. It must be symmetric, idempotent, and have positive rank. Small numerical deviations within the implemented tolerance are accepted.

For data.frame input, observations are sorted according to `subject` and `dimension` (in that order) and then converted to a matrix with the first subject in the first row. This should be taken into account, when passing a matrix to `hypothesis`.

Upper-tail probabilities are computed directly. Reported p-values are bounded below by `.Machine$double.eps`; a returned value at this boundary should be interpreted as no larger than the numerical reporting threshold.

Value

A named list of class `"hdrm_single"` with components:

data A matrix with the data used. If data was a data.frame, this is the transformed output.

dim A named numeric vector containing the repeated-measurement dimension `d` and the number of analyzed subjects `N`.

H A named list with the hypothesis-matrix `T` used for the test and the hypothesis label `label`.

AM The input value for `AM`.

f The estimated Pearson degrees of freedom.

statistic The standardized test statistic W.

tau The estimated convergence parameter $\tau = 1/f$.

p.value The upper-tail p-value, bounded below by `.Machine$double.eps`.

References

Pauly M, Ellenberger D, Brunner E (2015). "Analysis of high-dimensional one group repeated measures designs." Statistics, 49(6), 1243–1261. doi:10.1080/02331888.2015.1050022.

Sattler P, Rosenbaum M (2025). "Choice of the hypothesis matrix for using the Anova-type-statistic." Statistics & Probability Letters, 219, 110356. doi:10.1016/j.spl.2025.110356.

Examples

```
# Long-format data ------------------------------------------------------

data("EEG")
names(EEG) # already contains columns value, dimension and subject

## select only one diagnostic group for one-group analysis
EEG_single <- EEG[EEG$group == "SCC+", ]
```

```
## test whether the time profile is flat
hdrm_single(
  data = EEG_single,
  hypothesis = "flat"
)

## define hypothesis = "flat" via equivalent projection matrix
d <- nlevels(EEG_single$dimension)
flat_projection <- diag(d) -
  matrix(1 / d, nrow = d, ncol = d)

## test whether the time profile is flat via custom hypothesis matrix
hdrm_single(
  data = EEG_single,
  hypothesis = flat_projection
)


# Wide matrix data ------------------------------------------------------

data("birthrates")

## transform 'birthrates' to matrix and transpose
birthrates_matrix <- t(as.matrix(birthrates))

## test whether the time profile is flat
hdrm_single(
  data = birthrates_matrix,
  hypothesis = "flat"
)

## define hypothesis = "flat" via equivalent projection matrix
d <- ncol(birthrates_matrix)
flat_projection <- diag(d) -
  matrix(1 / d, nrow = d, ncol = d)

## test whether the time profile is flat via custom hypothesis matrix
hdrm_single(
  data = birthrates_matrix,
  hypothesis = flat_projection
)
```

Index

∗ **datasets**
    birthrates, 2
    EEG, 3

birthrates, 2

EEG, 3

hdrm_grouped, 4
hdrm_single, 8