Package 'MVT'

January 20, 2025

Type Package

Version 0.3-81

Title Estimation and Testing for the Multivariate t-Distribution

Date 2024-09-24

Maintainer Felipe Osorio <felipe.osorios@usm.cl>

Description Routines to perform estimation and inference under the multivariate
t-distribution <doi:10.1007/s10182-022-00468-2>. Currently, the following methodologies
are implemented: multivariate mean and covariance estimation, hypothesis testing
about equicorrelation and homogeneity of variances, the Wilson-Hilferty transformation,
QQ-plots with envelopes and random variate generation.

Depends R(>= 3.5.0), fastmatrix

LinkingTo fastmatrix

Imports stats, utils, graphics

License GPL-3

URL http://mvt.mat.utfsm.cl/

NeedsCompilation yes

LazyLoad yes

Author Felipe Osorio [aut, cre] (<https://orcid.org/0000-0002-4675-5201>)

Repository CRAN

Date/Publication 2024-09-24 11:40:02 UTC

Contents

---

companies   *Financial data*

Description

Data extracted from Standard & Poor's Compustat PC Plus. This dataset has been used to illustrate some influence diagnostic techniques.

Usage

```
data(companies)
```

Format

A data frame with 26 observations on the following 3 variables.

book book value in dollars per share at the end of 1992.

net net sales in millions of dollars in 1992.

ratio sales to assets ratio in 1992.

Source

Hadi, A.S., and Nyquist, H. (1999). Frechet distance as a tool for diagnosing multivariate data. *Linear Algebra and Its Applications* **289**, 183-201.

Hadi, A.S., and Son, M.S. (1997). Detection of unusual observations in regression and multivariate data. In: A. Ullah, D.E.A. Giles (Eds.) *Handbook of Applied Economic Statistics*. Marcel Dekker, New York. pp. 441-463.

---

cork	*Cork borings*

Description

Measurements of the weight of cork borings taken from the north (N), east (E), south (S), and west (W) directions of 28 trees. It is of interest to compare the bark thickness (and hence weight) in the four directions.

Usage

```
data(cork)
```

Format

A data frame with 28 observations on the following 4 variables.

N north.
E east.
S south.
W west.

Source

Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979). *Multivariate Analysis.* Academic Press, London.

---

envelope.student	*QQ-plot with simulated envelopes*

Description

Constructs a normal QQ-plot using a Wilson-Hilferty transformation for the estimated Mahalanobis distances obtained from the fitting procedure.

Usage

```
envelope.student(object, reps = 50, conf = 0.95, plot.it = TRUE)
```

Arguments

object	an object of class `'studentFit'` representing the fitted model.
reps	number of simulated point patterns to be generated when computing the envelopes. The default number is 50.
conf	the confidence level of the envelopes required. The default is to find 95% confidence envelopes.
plot.it	if TRUE it will draw the corresponding plot, if FALSE it will only return the computed values.

Value

A list with the following components :

transformed   a vector with the z-scores obtained from the Wilson-Hilferty transformation.

envelope   a matrix with two columns corresponding to the values of the lower and upper
               pointwise confidence envelope.

References

Atkinson, A.C. (1985). *Plots, Transformations and Regression*. Oxford University Press, Oxford.

Osorio, F., Galea, M., Henriquez, C., Arellano-Valle, R. (2023). Addressing non-normality in
multivariate analysis using the t-distribution. *AStA Advances in Statistical Analysis* **107**, 785-813.

See Also

WH.student

Examples

```
data(PSG)
fit <- studentFit(~ manual + automated, data = PSG, family = Student(eta = 0.25))
envelope.student(fit, reps = 500, conf = 0.95)
```

---

equicorrelation.test   *Equicorrelation test*

Description

Performs several test for testing that the covariance matrix follows an equicorrelation (or compound
symmetry) structure. Likelihood ratio test, score, Wald and gradient can be used as a test statistic.

Usage

```
equicorrelation.test(object, test = "LRT")
```

Arguments

object   object of class 'studentFit' representing the fitted model.

test   test statistic to be used. One of "LRT" (default), "Wald", "score" or "gradient".

Value

A list of class 'equicorrelation.test' with the following elements:

statistic	value of the statistic, i.e. the value of either Likelihood ratio test, Wald, score or gradient test.
parameter	the degrees of freedom for the test statistic, which is chi-square distributed.
p.value	the p-value for the test.
estimate	the estimated covariance matrix.
null.value	the hypothesized value for the covariance matrix.
method	a character string indicating what type of test was performed.
null.fit	a list representing the fitted model under the null hypothesis.
data	name of the data used in the test.

References

Sutradhar, B.C. (1993). Score test for the covariance matrix of the elliptical t-distribution. *Journal of Multivariate Analysis* **46**, 1-12.

Examples

```
data(examScor)
fit <- studentFit(examScor, family = Student(eta = .25))
fit

z <- equicorrelation.test(fit, test = "LRT")
z
```

---

examScor   *Open/Closed book data*

Description

Dataset from Mardia, Kent and Bibby on 88 students who took examinations in five subjects. The first two subjects were tested with closed book exams and the last three were tested with open book exams.

Usage

```
data(examScor)
```

Format

A data frame with 88 observations on the following 5 variables.

mechanics mechanics, closed book exam.

vectors vectors, closed book exam.

algebra algebra, open book exam.

analysis analysis, open book exam.

statistics statistics, open book exam.

Source

Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979). *Multivariate Analysis*. Academic Press, London.

---

homogeneity.test   *Test of variance homogeneity of correlated variances*

Description

Performs several test for testing equality of $p \geq 2$ correlated variables. Likelihood ratio test, score, Wald and gradient can be used as a test statistic.

Usage

```
homogeneity.test(object, test = "LRT")
```

Arguments

object	object of class `'studentFit'` representing the fitted model.
test	test statistic to be used. One of "LRT" (default), "Wald", "score" or "gradient".

Value

A list of class 'homogeneity.test' with the following elements:

statistic	value of the statistic, i.e. the value of either Likelihood ratio test, Wald, score or gradient test.
parameter	the degrees of freedom for the test statistic, which is chi-square distributed.
p.value	the p-value for the test.
estimate	the estimated covariance matrix.
null.value	the hypothesized value for the covariance matrix.
method	a character string indicating what type of test was performed.
null.fit	a list representing the fitted model under the null hypothesis.
data	name of the data used in the test.

References

Harris, P. (1985). Testing the variance homogeneity of correlated variables. *Biometrika* **72**, 103-107.

Modarres, R. (1993). Testing the equality of dependent variables. *Biometrical Journal* **7**, 785-790.

Examples

```
data(examScor)
fit <- studentFit(examScor, family = Student(eta = .25))
fit

z <- homogeneity.test(fit, test = "LRT")
z
```

---

kurtosis.student   *Mardia's multivariate kurtosis coefficient*

Description

This function computes the kurtosis of a multivariate distribution and estimates the kurtosis parameter for the t-distribution using the method of moments.

Usage

```
kurtosis.student(x)
```

Arguments

x   vector or matrix of data with, say, p columns.

Value

A list with the following components :

kurtosis	returns the value of Mardia's multivariate kurtosis.
kappa	returns the excess kurtosis related to a multivariate t-distribution.
eta	estimated shape (kurtosis) parameter using the methods of moments, only valid if $0 \leq \eta < 1/4$.

References

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519-530.

Osorio, F., Galea, M., Henriquez, C., Arellano-Valle, R. (2023). Addressing non-normality in multivariate analysis using the t-distribution. *AStA Advances in Statistical Analysis* **107**, 785-813.

Examples

```
data(companies)
kurtosis.student(companies)
```

---

mStudent   *Multivariate t distribution*

Description

These functions provide the density and random number generation from the multivariate Student-t distribution.

Usage

```
dmt(x, mean = rep(0, nrow(Sigma)), Sigma = diag(length(mean)), eta = 0.25, log = FALSE)
rmt(n = 1, mean = rep(0, nrow(Sigma)), Sigma = diag(length(mean)), eta = 0.25)
```

Arguments

x	vector or matrix of data.
n	the number of samples requested.
mean	a vector giving the means of each variable
Sigma	a positive-definite covariance matrix
eta	shape parameter (must be in [0,1/2)). Default value is 0.25
log	logical; if TRUE, the logarithm of the density function is returned.

Details

A random vector $\boldsymbol{X} = (X_1, \ldots, X_p)^T$ has a multivariate t distribution, with a $\boldsymbol{\mu}$ mean vector, covariance matrix $\boldsymbol{\Sigma}$, and $0 \leq \eta < 1/2$ shape parameter, if its density function is given by

$$f(\boldsymbol{x}) = K_p(\eta)|\boldsymbol{\Sigma}|^{-1/2} \left\{ 1 + c(\eta)(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \right\}^{-\frac{1}{2\eta}(1+\eta p)}.$$

where

$$K_p(\eta) = \left(\frac{c(\eta)}{\pi} \right)^{p/2} \frac{\Gamma(\frac{1}{2\eta}(1 + \eta p))}{\Gamma(\frac{1}{2\eta})},$$

with $c(\eta) = \eta/(1 - 2\eta)$. This parameterization of the multivariate t distribution is introduced mainly because $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ correspond to the mean vector and covariance matrix, respectively.

The function rmt is an interface to C routines, which make calls to subroutines from LAPACK. The matrix decomposition is internally done using the Cholesky decomposition. If Sigma is not non-negative definite then there will be a warning message.

This parameterization of the multivariate-*t* includes the normal distribution as a particular case when eta = 0.

Value

If x is a matrix with n rows, then dmt returns a $n \times 1$ vector considering each row of x as a copy from the multivariate *t* distribution.

If n = 1, then rmt returns a vector of the same length as mean, otherwise a matrix of n rows of random vectors.

References

Fang, K.T., Kotz, S., Ng, K.W. (1990). *Symmetric Multivariate and Related Distributions*. Chapman & Hall, London.

Gomez, E., Gomez-Villegas, M.A., Marin, J.M. (1998). A multivariate generalization of the power exponential family of distributions. *Communications in Statistics - Theory and Methods* **27**, 589-600.

Examples

```
# covariance matrix
Sigma <- matrix(c(10,3,3,2), ncol = 2)
Sigma

# generate the sample
y <- rmt(n = 1000, Sigma = Sigma)

# scatterplot of a random bivariate t sample with mean vector
# zero and covariance matrix 'Sigma'
par(pty = "s")
plot(y, xlab = "", ylab = "")
title("bivariate t sample (eta = 0.25)", font.main = 1)
```

MVT.control   *Set control parameters*

Description

Allows users to set control parameters for the estimation routine available in MVT.

Usage

```
MVT.control(maxiter = 2000, tolerance = 1e-6, fix.shape = FALSE)
```

Arguments

maxiter	maximum number of iterations. The default is 2000.
tolerance	the relative tolerance in the iterative algorithm.
fix.shape	whether the shape parameter should be kept fixed in the fitting processes. The default is fix.shape = FALSE.

Value

A list of control arguments to be used in a call to `studentFit`.

A call to `MVT.control` can be used directly in the `control` argument of the call to `studentFit`.

Examples

```
ctrl <- MVT.control(maxiter = 500, tol = 1e-04, fix.shape = TRUE)
data(PSG)
studentFit(~ manual + automated, data = PSG, family = Student(eta = 0.25),
  control = ctrl)
```

PSG   *Transient sleep disorder*

Description

Clinical study designed to compare the automated and semi-automated scoring of Polysomnographic (PSG) recordings used to diagnose transient sleep disorders. The study considered 82 patients who were given a sleep-inducing drug (Zolpidem 10 mg). Measurements of latency to persistent sleep (LPS: lights out to the beginning of 10 consecutive minutes of uninterrupted sleep) were obtained using six different methods.

Usage

```
data(PSG)
```

Format

A data frame with 82 observations on the following 3 variables.

manual fully manual scoring.

automated automated scoring by the Morpheus software.

partial Morpheus automated scoring with manual review.

Source

Svetnik, V., Ma, J., Soper, K.A., Doran, S., Renger, J.J., Deacon, S., Koblan, K.S. (2007). Evaluation of automated and semi-automated scoring of polysomnographic recordings from a clinical trial using zolpidem in the treatment of insomnia. *SLEEP* **30**, 1562-1574.

Student   *Family object for the multivariate t-distribution*

Description

Provide a convenient way to specify the details of the model used by function `studentFit`.

Usage

```
Student(eta = .25)
```

Arguments

eta   shape parameter for the multivariate t-distribution, must be confined to `[0,1/2)`.

Details

`Student` is a generic function to create info about the t-distribution which is passed to the estimation algorithm.

Examples

```
MyFmly <- Student(eta = .4)
MyFmly
```

---

studentFit   *Estimation of mean and covariance using the multivariate t-distribution*

Description

Estimates the mean vector and covariance matrix assuming the data came from a multivariate t-distribution: this provides some degree of robustness to outlier without giving a high breakdown point.

Usage

```
studentFit(x, data, family = Student(eta = .25), covStruct = "UN", subset, na.action,
control)
```

Arguments

x	a formula or a numeric matrix or an object that can be coerced to a numeric matrix.
data	an optional data frame (or similar: see [model.frame](#)), used only if x is a formula. By default the variables are taken from environment(formula).
family	a description of the error distribution to be used in the model. By default the multivariate t-distribution with 0.25 as shape parameter is considered (using eta = 0 allows to tackle the multivariate normal distribution).
covStruct	a character string specifying the type of covariance structure. The options available are: "UN" (unstructured) general covariance matrix with no additional structure (default), "CS" (compound symmetry) corresponding to a constant correlation or equicorrelation, "DIAG" (diagonal) representing a diagonal positive-definite matrix, "HOMO" (homogeneous) meaning a covariance matrix with homogeneous variances.
subset	an optional expression indicating the subset of the rows of data that should be used in the fitting process.
na.action	a function that indicates what should happen when the data contain NAs.
control	a list of control values for the estimation algorithm to replace the default values returned by the function [MVT.control](#).

Value

A list with class 'studentFit' containing the following components:

call	a list containing an image of the studentFit call that produced the object.
family	the [Student](#) object used, with the estimated shape parameters (if requested).
center	final estimate of the location vector.
Scatter	final estimate of the scale matrix.
logLik	the log-likelihood at convergence.
numIter	the number of iterations used in the iterative algorithm.
weights	estimated weights corresponding to the assumed heavy-tailed distribution.
distances	estimated squared Mahalanobis distances.
eta	final estimate of the shape parameter, if requested.

Generic function print show the results of the fit.

References

Kent, J.T., Tyler, D.E., Vardi, Y. (1994). A curious likelihood identity for the multivariate t-distribution. *Communications in Statistics: Simulation and Computation* **23**, 441-453.

Lange, K., Little, R.J.A., Taylor, J.M.G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association* **84**, 881-896.

Osorio, F., Galea, M., Henriquez, C., Arellano-Valle, R. (2023). Addressing non-normality in multivariate analysis using the *t*-distribution. *AStA Advances in Statistical Analysis* **107**, 785-813.

See Also

cov, cov.rob and cov.trob in package **MASS**.

Examples

```
data(PSG)
fit <- studentFit(~ manual + automated, data = PSG, family = Student(eta = 0.25))
fit
```

WH.student	*Wilson-Hilferty transformation*

Description

Returns the Wilson-Hilferty transformation of random variables with F distribution.

Usage

```
WH.student(x, center, cov, eta = 0)
```

Arguments

x	object of class `'studentFit'` from which is extracted the estimated Mahalanobis distances of the fitted model. Also x can be a vector or matrix of data with, say, p columns.
center	mean vector of the distribution or second data vector of length p. Not required if x have class `'studentFit'`.
cov	covariance matrix (p by p) of the distribution. Not required if x have class `'studentFit'`.
eta	shape parameter of the multivariate t-distribution. By default the multivariate normal (eta = 0) is considered.

Details

Let F the following random variable:

$$F = \frac{D^2/p}{1 - 2\eta}$$

where D^2 denotes the squared Mahalanobis distance defined as

$$D^2 = (x - \mu)^T \Sigma^{-1} (x - \mu)$$

Thus the Wilson-Hilferty transformation is given by

$$z = \frac{(1 - \frac{2\eta}{9})F^{1/3} - (1 - \frac{2}{9p})}{(\frac{2\eta}{9}F^{2/3} + \frac{2}{9p})^{1/2}}$$

and z is approximately distributed as a standard normal distribution. This is useful, for instance, in the construction of QQ-plots.

For eta = 0, we obtain

$$z = \frac{F^{1/3} - (1 - \frac{2}{9p})}{(\frac{2}{9p})^{1/2}}$$

which is the Wilson-Hilferty transformation for chi-square variables.

References

Osorio, F., Galea, M., Henriquez, C., Arellano-Valle, R. (2023). Addressing non-normality in multivariate analysis using the *t*-distribution. *AStA Advances in Statistical Analysis* **107**, 785-813.

Wilson, E.B., and Hilferty, M.M. (1931). The distribution of chi-square. *Proceedings of the National Academy of Sciences of the United States of America* **17**, 684-688.

See Also

cov, mahalanobis, envelope.student

Examples

```
data(companies)
x <- companies
z <- WH.student(x, center = colMeans(x), cov = cov(x))
par(pty = "s")
qqnorm(z, main = "Transformed distances Q-Q plot")
abline(c(0,1), col = "red", lwd = 2)
```

WindSpeed   *Wind speed data*

Description

This dataset consists of 278 hourly average wind speed in the Pacific North-West of the United States collected at three meteorological towers approximately located on a line and ordered from west to east: Goodnoe Hills (gh), Kennewick (kw), and Vansycle (vs). The data were collected from 25 February to 30 November 2003 recorded at midnight, a time when wind speeds tend to peak.

Usage

```
data(WindSpeed)
```

Format

A data frame with 278 observations on the following 3 variables.

gh Goodnoe Hills.

kw Kennewick.

vs Vansycle.

Source

Azzalini, A., Genton, M.G. (2008). Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review* **76**, 106-129.

Index