Package 'CCP'

January 20, 2025

Type Package

Title Significance Tests for Canonical Correlation Analysis (CCA)

Version 1.2

Date 2022-04-20

Author Uwe Menzel

Maintainer Uwe Menzel <uwemenzel@gmail.com>

Description Significance tests are provided for canonical correlation analysis, including asymptotic tests and a Monte Carlo method.

License GPL

LazyLoad yes

Repository CRAN

Date/Publication 2022-04-21 21:50:02 UTC

NeedsCompilation no

Contents

CCP-package   *Significance Tests for Canonical Correlation Analysis*

Description

The package provides functions to test for statistical significance of canonical correlation coefficients, including asymptotic methods and a Monte Carlo approach. Additionally, functions are available to plot the asymptotic distributions and the permutation distribution, respectively. The main user functions are `p.asym`, `p.perm`, `plt.asym`, and `plt.perm`.

Author(s)

Uwe Menzel <uwemenzel@gmail.com>

CCP-internal   *Internal functions for CCP package*

Description

Internal functions for CCP package

Usage

```
Hotelling.stat(rho, N, p, q)
HotellingLawleyTrace(rho, p, q)
p.Roy(rho, N, p, q)
Pillai.stat(rho, N, p, q)
PillaiBartlettTrace(rho, p, q)
RaoF.stat(rho, N, p, q)
WilksLambda(rho, p, q)
```

Arguments

rho	vector containing the canonical correlation coefficients.
N	number of observations for each variable.
p	number of independent variables.
q	number of dependent variables.

Details

These functions are not intended to be called by the user.

---

p.asym	*Asymptotic tests for the statistical significance of canonical correlation coefficients*

Description

This function runs asymptotic tests to assign the statistical significance of canonical correlation coefficients. F-approximations of Wilks' Lambda, the Hotelling-Lawley Trace, the Pillai-Bartlett Trace, or of Roy's Largest Root can be used as a test statistic.

Usage

```
p.asym(rho, N, p, q, tstat = "Wilks")
```

Arguments

rho	vector containing the canonical correlation coefficients.
N	number of observations for each variable.
p	number of independent variables.
q	number of dependent variables.
tstat	test statistic to be used. One of "Wilks" (default), "Hotelling", "Pillai", or "Roy".

Details

Canonical correlation analysis (CCA) measures the degree of linear relationship between two sets of variables. The number of correlation coefficients calculated in CCA is equal to the number of variables in the smaller set: $m = min(p, q)$. The coefficients are arranged in descending order of magnitude: $rho[1] > rho[2] > rho[3] > ... > rho[m]$. Except for tstat = "Roy", the function p.asym calculates m p-values for each test statistic: the first p-value is calculated including *all* canonical correlation coefficients, the second p-value is calculated by excluding $rho[1]$, the third p-value is calculated by excluding $rho[1]$ and $rho[2]$ etc., therewith allowing assessment of the statistical significance of each individual correlation coefficient. On principle, Roy's Largest Root takes only $rho[1]$ into account, hence one p-value is calculated only.

Value

stat	value of the statistic, i.e. the value of either Wilks' Lambda, the Hotelling-Lawley Trace, the Pillai-Bartlett Trace, or Roy's Largest Root.
approx	value of the corresponding F-approximation for the statistic.
df1	numerator degrees of freedom for the F-approximation.
df2	denominator degrees of freedom for the F-approximation.
p.value	p-value

Note

Usage of asymptotic approximations requires multivariate normality of the variables, or a large number of observations. Canonical correlation is sensitive to outliers. The F-approximation for Roy's Largest Root is an upper bound, and the significance level is therefore optimistically small. The canonical correlation coefficients are statistically significant if Wilks' Lambda is *smaller* than a critical value.

Author(s)

Uwe Menzel <uwemenzel@gmail.com>

References

Wilks, S. S. (1935) On the independence of k sets of normally distributed statistical variables. *Econometrica*, **3** 309–326.

Rao, C. R. (1973) *Linear Statistical Inference and It's Applications (2nd ed.).* John Wiley and Sons, New York, 533–543, 555–556.

Pillai, K. C. W. (1956) On the distribution of the largest or the smallest root of a matrix in multivariate analysis. *Biometrika*, **43** 122–127.

Muller, K. E. and Peterson B. L. (1984) Practical Methods for computing power in testing the multivariate general linear hypothesis. *Computational Statistics & Data Analysis*, **2** 143–158.

Anderson, T. W. (1984) *An introduction to Multivariate Statistical Analysis.* John Wiley and Sons, New York.

See Also

See the function cancor or the CCA package for the calculation of canonical correlation coefficients.

Examples

```
## Load the CCP package:
library(CCP)



## Simulate example data:
X <- matrix(rnorm(150), 50, 3)
Y <- matrix(rnorm(250), 50, 5)


## Calculate canonical correlations:
rho <- cancor(X,Y)$cor

## Define number of observations,
## and number of dependent and independent variables:
N = dim(X)[1]
p = dim(X)[2]
q = dim(Y)[2]
```

```
## Calculate p-values using F-approximations of some test statistics:
p.asym(rho, N, p, q, tstat = "Wilks")
p.asym(rho, N, p, q, tstat = "Hotelling")
p.asym(rho, N, p, q, tstat = "Pillai")
p.asym(rho, N, p, q, tstat = "Roy")

## Plot the F-approximation for Wilks' Lambda,
## considering 3, 2, or 1 canonical correlation(s):
res1 <- p.asym(rho, N, p, q)
plt.asym(res1,rhostart=1)
plt.asym(res1,rhostart=2)
plt.asym(res1,rhostart=3)
```

---

p.perm	*Permutation test for the significance of canonical correlation coefficients*

Description

This function runs a permutation test to assign the statistical significance of canonical correlation coefficients. Wilks' Lambda, the Hotelling-Lawley Trace, the Pillai-Bartlett Trace, or of Roy's Largest Root can be used as a test statistic.

Usage

```
p.perm(X, Y, nboot = 999, rhostart = 1, type = "Wilks")
```

Arguments

X	array containing the independent variables, with N rows (number of observations) and p columns (number of independent variables).
Y	array containing the dependent variables, with N rows (number of observations) and q columns (number of dependent variables).
nboot	number of permutation resamples calculated.
rhostart	index of the largest canonical correlation coefficient included in the calculation of the test statistic (see $Details$).
type	test statistic to be used. One of "Wilks" (default), "Hotelling", "Pillai", or "Roy".

Details

Permutation tests are based on resampling of the original data without replacement. To test the hypothesis of no correlation between two sets (X, Y) of variables, the values of one variable (Y) are randomly reassigned. Permutation tests do not require specific population distributions of the variables such as normal distribution. Canonical correlation analysis (CCA) calculates $m = min(p, q)$ correlation coefficients, see `p.asym`. The coefficients are arranged in descending order of magnitude: $rho[1] > rho[2] > rho[3] > ... > rho[m]$. In p.perm, the parameter $rhostart$ determines how many correlation coefficients are included in the calculation of the test statistic: choosing

$rhostart = 1$, all canonical correlations are included, choosing $rhostart = 2$, the biggest canonical correlation ($rho[1]$) is excluded, choosing $rhostart = 3$, both $rho[1]$ and $rho[2]$ are excluded, etc. On principle, Roy's Largest Root takes only $rho[1]$ into account, hence only $rhostart = 1$ can be chosen.

Value

stat0	original value of the statistic (without resampling).
stat	*nboot* values of the statistic, one for each permutation resampling.
nexcess	number of permutation resamplings that resulted in a more extreme value of the statistic than $stat0$.
p.value	p-value, derived from $nexcess$.

Note

Permutation tests do not require a specific distribution of the variables. Tests based on random resampling do generally not yield the same results when repeated. The canonical correlation coefficients are statistically significant if Wilks' Lambda is *smaller* than a critical value.

Author(s)

Uwe Menzel <uwemenzel@gmail.com>

References

Efron, B. and Tibshirani, R. J. (1994) *An Introduction to the Bootstrap*, Chapman & Hall, New York.

Moore, D. S. amd McCabe, G. P. (2006) *Introduction to the Practice of Statistics, Chapter 14*, W. H. Freeman, New York.

See Also

See the function [cancor](#) or the CCA package for the calculation of canonical correlation coefficients.

Examples

```
## Load the CCP package:
library(CCP)


## Simulate example data:
X <- matrix(rnorm(150), 50, 3)
Y <- matrix(rnorm(250), 50, 5)



## Run permutation test using Wilks Lambda (default) as test statistic;
## include different numbers of canonical correlations:
p.perm(X, Y, nboot = 999, rhostart = 1)
p.perm(X, Y, nboot = 999, rhostart = 2)
```

```
p.perm(X, Y, nboot = 999, rhostart = 3)


## Run permutation tests based on different test statistics:
p.perm(X, Y, nboot = 999, rhostart = 1, type = "Wilks")
p.perm(X, Y, nboot = 999, rhostart = 1, type = "Hotelling")
p.perm(X, Y, nboot = 999, rhostart = 1, type = "Pillai")
p.perm(X, Y, nboot = 999, rhostart = 1, type = "Roy")

## Plot the permutation distribution
## with the value of the original statistic marked:
out <- p.perm(X, Y, nboot = 999, rhostart = 3, type = "Hotelling")
plt.perm(out)
```

---

plt.asym   *Plot asymptotic distributions for test statistics*

Description

This function plots asymptotic distributions used to test the statistical significance of canonical correlation coefficients, see function `p.asym`.

Usage

```
plt.asym(p.asym.out, rhostart = 1)
```

Arguments

p.asym.out   output of p.asym, see example below.

rhostart   index of the largest canonical correlation coefficient included in the calculation
                of the test statistic, see function `p.asym`.

Details

Depending on what type of statistic was chosen in p.asym, an F-approximation for this statistic is plotted. The statistic is one of: Wilks' Lambda, Hotelling-Lawley Trace, Pillai-Bartlett Trace, or Roy's Largest Root. The value of the test statistic calculated from the canonical correlation coefficients is plotted as a vertical line; thus the area located below the curve and to the right of the vertical line corresponds to the p-value. The vertical line is *not* visible if the value of the test statistic is in the far tail of the distribution, resulting in a p-value which is (close to) zero. The numerical value of the test statistic, the numerator and denominator degrees of freedom of the F-distribution, and the p-value are plotted on the bottom of the figure.

Author(s)

Uwe Menzel <uwemenzel@gmail.com>

See Also

See the function p.asym for the calculation of the p-values.

Examples

```
## Load the CCP package:
library(CCP)



## Simulate  example data:
X <- matrix(rnorm(150), 50, 3)
Y <- matrix(rnorm(250), 50, 5)



## Calculate canonical correlations,
## using the function "cancor" from the "stats" package:
rho <- cancor(X,Y)$cor

## Define number of observations,
## and number of dependent and independent variables:
N = dim(X)[1]
p = dim(X)[2]
q = dim(Y)[2]


## Plot the F-approximation for Wilks' Lambda,
## considering 3, 2, or 1 canonical correlation(s):
res1 <- p.asym(rho, N, p, q)
plt.asym(res1,rhostart=1)
plt.asym(res1,rhostart=2)
plt.asym(res1,rhostart=3)


## Plot the F-approximation for the Hotelling-Lawley Trace,
## considering 3, 2, or 1 canonical correlation(s):
res2 <- p.asym(rho, N, p, q, tstat="Hotelling")
plt.asym(res2,rhostart=1)
plt.asym(res2,rhostart=2)
plt.asym(res2,rhostart=3)
```

---

plt.perm   *Plot permutation distributions for test statistics*

Description

This function plots permutation distributions for test statistics that are used to assign the statistical significance of canonical correlation coefficients, see function p.perm.

Usage

```
plt.perm(p.perm.out)
```

Arguments

p.perm.out   output of `p.perm`, see example below.

Details

Depending on what type of statistic was chosen in `p.perm`, a permutation distribution of this statistic is shown. The statistic is one of: Wilks' Lambda, Hotelling-Lawley Trace, Pillai-Bartlett Trace, or Roy's Largest Root. These test statistics can be used to assign significance levels to canonical correlation coefficients, for details see `p.perm`. The value corresponding to the "original" test statistic (calculated using the canonical correlation coefficients resulting from unpermuted data) is plotted as a red, dotted vertical line; thus the area of the histogram outside this line determines the approximate p-value. The vertical line is *not* visible if the value corresponding to the original test statistic is in the far tail of the histogram, yielding a p-value which is (close to) zero. The numerical value corresponding to the original test statistic is plotted in the subtitle of the graph, as well as the calculated p-value. The grey vertical line represents the mean of the permutation distribution.

Author(s)

Uwe Menzel <uwemenzel@gmail.com>

See Also

See the function `p.perm` for the calculation of the p-values.

Examples

```
## Load the CCP package:
library(CCP)



## Simulate example data:
X <- matrix(rnorm(150), 50, 3)
Y <- matrix(rnorm(250), 50, 5)



## Calculate canonical correlations:
rho <- cancor(X,Y)$cor

## Define number of observations,
## and number of dependent and independent variables:
N = dim(X)[1]
p = dim(X)[2]
q = dim(Y)[2]
```

```
## Plot the permutation distribution of an F approximation
## for Wilks Lambda, considering 3 and 2 canonical correlations:
out1 <- p.perm(X, Y, nboot = 999, rhostart = 1)
plt.perm(out1)
out2 <- p.perm(X, Y, nboot = 999, rhostart = 2)
plt.perm(out2)


## Plot the permutation distribution of an F approximation
## for the Pillai-Bartlett Trace,
## considering 3, 2, and 1 canonical correlation(s):
res1 <- p.perm(X, Y, nboot = 999, rhostart = 1, type = "Pillai")
plt.perm(res1)
res2 <- p.perm(X, Y, nboot = 999, rhostart = 2, type = "Pillai")
plt.perm(res2)
res3 <- p.perm(X, Y, nboot = 999, rhostart = 3, type = "Pillai")
plt.perm(res3)
```

Index