Package 'npmv'

July 22, 2025

Type Package

Title Nonparametric Comparison of Multivariate Samples

Version 2.4.1

Date 2025-1-7

Maintainer Amanda Ellis <arelli4@uky.edu>

Description Performs analysis of one-way multivariate data, for small samples using Nonparametric techniques. Using approximations for ANOVA Type, Wilks' Lambda, Lawley Hotelling, and Bartlett Nanda Pillai Test statics, the package compares the multivariate distributions for a single explanatory variable. The comparison is also performed using a permutation test for each of the four test statistics. The package also performs an all-subsets algorithm regarding variables and regarding factor levels.

Imports Formula, methods

License GPL-2

RoxygenNote 7.3.2

NeedsCompilation no

Author Woodrow Burchett [aut],
Amanda Ellis [aut, cre]

Repository CRAN

Date/Publication 2025-01-07 19:40:02 UTC

Contents

---

npmv-package	*Nonparametric Comparison of Multivariate Samples*

Description

Performs analysis of one-way multivariate data using nonparametric techniques developed since 2008. Allows for small samples and ordinal variables, or even mixture of the different variable types ordinal, quantitative, binary. Using F-approximations for ANOVA Type, Wilks' Lambda Type, Lawley Hotelling Type, and Bartlett Nanda Pillai Type test statics, as well as a permutation test for each, the package compares the multivariate distributions of the different samples. Also computes nonparametric relative effects and produces plots.

Details

This package provides the R functions nonpartest and ssnonpartest to compute nonparametric test statistics. The function nonpartest computes the global nonparametric test statistics, their permutation test analogs, and calculates nonparametric relative effects. The function ssnonpartest performs an all-subset algorithm to determine which variables cause significant effects, and between which factor levels. See the examples below for some basic uses and look in the help pages for each function for a much more detailed look.

Warning

The nonparametric methods implemented in the code have been developed for complete data with no missing values.The code automatically produces a warning if there is missing data.

Under certain conditions, the matrices H and G are singular (See literature for explanation of H and G), for example when the number of response variables exceeds the sample size. When this happens, only the ANOVA type statistic can be computed. The code automatically produces a warning if H or G are singular.

Author(s)

Woodrow Burchett, Amanda Ellis, Arne Bathke

Maintainer: Amanda Ellis <arelli4@uky.edu>

References

Arne C. Bathke , Solomon W. Harrar, and Laurence V. Madden. "How to compare small multivariate samples using nonparametric tests," Computational Statistics and Data Analysis 52 (2008) 4951-4965

Woodrow W. Burchett, Amanda R. Ellis, Solomon W. Harrar, Arne C. Bathke (2017). "Nonparametric Inference for Multivariate Data: The R Package npmv.," Journal of Statistical Software, 76(4), 1-18.

Brunner E, Domhof S, Langer F (2002), Nonparametric Analysis of Longitudinal Data in Factorial Experiments. Wiley, New York.

Chunxu Liu, Arne C. Bathke, Solomon W. Harrar. "A nonparametric version of Wilks' lambda-Asymptotic results and small sample approximations" Statistics and Probability Letters 81 (2011) 1502-1506

Horst, L.E., Locke, J., Krause, C.R., McMahaon, R.W., Madden, L.V., Hoitink, H.A.J., 2005. Suppression of Botrytis blight of Begonia by Trichoderma hamatum 382 in peat and compost-amended potting mixes. Plant Disease 89, 1195-1200.

Examples

```
data(sberry)
nonpartest(weight|bot|fungi|rating~treatment,sberry,permreps=1000)
ssnonpartest(weight|bot|fungi|rating~treatment,sberry,test=c(1,0,0,0),alpha=.05,
            factors.and.variables=TRUE)
```

---

nonpartest	*Nonparametric Comparison of Multivariate Samples*

Description

Performs analysis of one-way multivariate data using nonparametric techniques developed since 2008. Allows for small samples and ordinal variables, or even mixture of the different variable types ordinal, quantitative, binary. Using F-approximations for ANOVA Type, Wilks' Lambda Type, Lawley Hotelling Type, and Bartlett Nanda Pillai Type test statics, as well as a permutation test for each, the package compares the multivariate distributions of the different samples. Also computes nonparametric relative effects.

Usage

```
nonpartest(formula,data,permtest=TRUE,permreps=10000,plots=TRUE,
          tests=c(1,1,1,1),releffects=TRUE,...)
```

Arguments

formula	an object of class "formula", with a single explanatory variable and multiple response variables (or one that can be coerced to that class).
data	an object of class "data.frame", containing the variables in the formula.
permtest	logical. If TRUE the p-values for the permutation test are returned
permreps	number of replications in permutation test
plots	logical. If TRUE box plots are produced for each response variable versus treatment
tests	vector of zeros and ones which specifies which test statistics are to be calculated. A 1 corresponds to the test statistics which are to be returned
releffects	logical. If TRUE the relative effects are returned
...	Graphical parameters to be passed to the boxplot function.

Details

The nonparametric methods implemented in the code have been developed for complete data with no missing values. The code automatically produces a warning if there is missing data.

Value

Returns a list of 2 data frames if relative effects are turned on, otherwise returns a single data frame. First data frame consist of p-values for test statistics and permutation test (if permutation test is turned on), second data frame consist of relative effects for each response variable.

Note

We define (for simplicity, only the formula for the balanced case is given here, the unbalanced case is given in the literature):

$H = (1/(a-1)) * \sum_{i=1}^{a} n(\bar{R}_{i.} - \bar{R}_{..})(\bar{R}_{i.} - \bar{R}_{..})'$

$G = (1/(N-1)) * \sum_{i=1}^{a} \sum_{j=1}^{n} (R_{ij} - \bar{R}_{i.})(R_{ij} - \bar{R}_{i.})'$

The ANOVA Type statistic is given by: $T_A = tr(H)/tr(G)$

The distribution of T_A is approximated by an F distribution with $\hat{f}_1$ and $\hat{f}_2$ where: $\hat{f}_1 = \frac{tr(G)^2}{tr(G^2)}$ and $\hat{f}_2 = \frac{a^2}{((a-1)\sum_{i=1}^{a} \frac{1}{n_i - 1})} * \hat{f}_1$

The Lawley Hotelling Type statistic is given by: $U = tr[(a-1)H((N-a)G)^{-1}]$ Using the McKeon approximation the distribution of U is approximated by a "stretched" F distribution with degrees freedom K and D where:

$K = p(a-1)$ and

$D = 4 + (K+2)/(B-1)$ and

$B = ((N-p-2)(N-a-1))/((N-a-p)(N-a-p-n))$

The Bartlett Nanda Pillai Type statistic is given by: $V = tr\{(a-1)H[(a-1)H + (N-a)G]^{-1}\}$ McKeon approximated the distribution of $((V/\gamma)/\nu_1)/((1-V/\gamma)/\nu_2)$ using an F distribution with degrees freedom ν_1 and ν_2 where: $\gamma = min(a-1, p)$

$\nu_1 = (p(a-1))/(\gamma(N-1)) * [(\gamma(N-a+\gamma-p)(N+2)(N-1))/((N-a)(N-p)) - 2]$

$\nu_2 = (N-a+\gamma-p)/(N) * [(\gamma(N-a+\gamma-p)(N-1))/((N-a)(N-p)) - 2]$

The Wilks' Lambda Type Statistic is given by

$\lambda = det(((N-a) * G)/((N-a) * G + (a-1) * H))$

The F approximation statistic is given by

$F_\lambda = [(1 - \lambda^{1/t})/(\lambda^{1/t})](df_2/df_1)$

where

$df_1 = p(a-1)$ and $df_2 = rt - (p(a-1) - 2)/2$ and $r = (N-a) - (p - (a-1) + 1)/2$.

If $p(a-1) = 2$ then t=1, else $t = \sqrt{(p^2(a-1)^2 - 4)/(p^2 + (a-1)^2 - 5)}$

Note that regarding the above formula, there is a typo in the article Liu, Bathke, Harrar (2011).

Warning: The nonparametric methods implemented in the code have been developed for complete data with no missing values. The code automatically produces a warning if there is missing data.

Under certain conditions, the matrices H and G are singular (See literature for explanation of H and G), for example when the number of response variables exceeds the sample size. When this happens, only the ANOVA type statistic can be computed. The code automatically produces a warning if H or G are singular.

Author(s)

Woodrow Burchett, Amanda Ellis, Arne Bathke

References

Arne C. Bathke , Solomon W. Harrar, and Laurence V. Madden. "How to compare small multivariate samples using nonparametric tests," Computational Statistics and Data Analysis 52 (2008) 4951-4965

Brunner E, Domhof S, Langer F (2002), Nonparametric Analysis of Longitudinal Data in Factorial Experiments. Wiley, New York.

Chunxu Liu, Arne C. Bathke, Solomon W. Harrar. "A nonparametric version of Wilks' lambda-Asymptotic results and small sample approximations" Statistics and Probability Letters 81 (2011) 1502-1506

Horst, L.E., Locke, J., Krause, C.R., McMahaon, R.W., Madden, L.V., Hoitink, H.A.J., 2005. Suppression of Botrytis blight of Begonia by Trichoderma hamatum 382 in peat and compost-amended potting mixes. Plant Disease 89, 1195-1200.

Examples

```
data(sberry)
nonpartest(weight|bot|fungi|rating~treatment,sberry,permreps=1000)
```

---

sberry   *Strawberry Fungal Data*

Description

The strawberry data set is a multivariate response data set that gives the measurements of weight, the percent of Botrytis, percent of other fungal species and the rating of symptoms from Phomopsis leaf blight, for 4 plots of strawberries each treated with one of 4 treatments. Three of the treatments were different chemicals, and one control.

Usage

```
data(sberry)
```

Format

sberry is a data frame with 16 cases (rows) and 6 variables (columns) named replication, weight, bot, fungi, and rating.

Details

A study was conducted in a commercial farm to evaluate the effects of three different fungicides (pesticides) on the control of fruit and foliar diseases of strawberry. A section of a 4-year-old strawberry planting was divided into 16 3-meter long single-row plots, and four treatments were randomly assigned to four plots each: sprayed with Kocide 2000 WG five times; sprayed with Elevate 50 WG plus Switch 62.5 WG four times; sprayed with V-10135 20 WP (experimental fungicide from Valent Corp.) three times; or not sprayed (control). All fruit were harvested and visually evaluated for symptoms of the fungus-caused disease grey mold (also known as Botrytis fruit rot), and symptoms of other fruit rots (caused by various fungal species). Total weight of all harvested fruit was determined. The percent of fruit with symptoms of Botrytis and other species was determined for each plot. Finally, the severity of symptoms on the foliage (leaflets) of Phomopsis leaf blight (another fungal-caused disease) was assessed with a 0-3 ordinal scale, where 0 represents disease free and 3 represents 40% or more of the foliage covered by lesions. Thirty leaflets were measured in each plot, and the median value of these measurements was determined.

Source

Horst, L.E., Locke, J., Krause, C.R., McMahaon, R.W., Madden, L.V., Hoitink, H.A.J., 2005. Suppression of Botrytis blight of Begonia by Trichoderma hamatum 382 in peat and compost-amended potting mixes. Plant Disease 89, 1195-1200.

Examples

```
data(sberry)
```

---

ssnonpartest	*Nonparametric Comparison of Multivariate Samples Using Subset algorithm*

Description

Performs detailed analysis of one-way multivariate data using nonparametric techniques developed since 2008. Allows for small samples and ordinal variables, or even mixture of the different variable types ordinal, quantitative, binary. Using F-approximations for ANOVA Type, Wilks' Lambda Type, Lawley Hotelling Type, and Bartlett Nanda Pillai Type test statics. The function compares the multivariate distributions of the different samples using a subset algorithm to determine which of the variables cause significant results, and which factor levels differ significantly from one another. The algorithm follows the closed multiple testing principle for factor levels, and adjusts p-values for subset testing of variables. In both cases, the global alpha-level is maintained at the prespecified level. When testing which subsets of factor levels produce significant results, the closure principle (Marcus, Peritz, Gabriel 1976, Sonnemann 2008) can be applied since the family of hypotheses is closed under intersections. When testing variables, the family of hypotheses is not closed under intersection. Therefore, in order to control the global (maximum overall) type I error rate, the following procedure is carried out: the global test involving all p variables is conducted at level alpha. At the steps where subsets of q<p variables are tested (first q=p-1, then q=p-2, etc. until q=1), the alpha-level is adjusted by factor (p choose q).

Usage

```
ssnonpartest(formula,data,alpha=.05,test=c(0,0,0,1),factors.and.variables=FALSE)
```

Arguments

formula	an object of class "formula", with a single explanatory variable and multiple response variables (or one that can be coerced to that class).
data	an object of class "data.frame", containing the variables in the formula.
alpha	numerical. Gives the global level of significance at which hypothesis test are to be performed.
test	vector of zeros and ones which specifies which test statistic is to be calculated. A 1 corresponds to the test statistic which is to be returned. Only one test statistic can be specified. Default is for Wilks' Lambda type statistic to be calculated. The order of the test statistics is: ANOVA type, Lawley Hotelling type (McKeon's F approximation), Bartlett-Nanda-Pillai type (Muller's F approximation), and Wilks' Lambda type.
factors.and.variables	logical. If TRUE subset algorithm is ran both by factor levels and by variable. Default is FALSE.

Details

The nonparametric methods implemented in the code have been developed for complete data with no missing values. The code automatically produces a warning if there is missing data.

Value

Returns the subsections which are significant.

Note

We define (for simplicity, only the formula for the balanced case is given here, the unbalanced case is given in the literature):

$H = (1/(a-1)) * \sum_{i=1}^{a} n(\bar{R}_{i.} - \bar{R}_{..})(\bar{R}_{i.} - \bar{R}_{..})'$

$G = (1/(N-1)) * \sum_{i=1}^{a} \sum_{j=1}^{n} (R_{ij} - \bar{R}_{i.})(R_{ij} - \bar{R}_{i.})'$

The ANOVA Type statistic is given by: $T_A = tr(H)/tr(G)$

The distribution of T_A is approximated by an F distribution with $\hat{f}_1$ and $\hat{f}_2$ where: $\hat{f}_1 = \frac{tr(G)^2}{tr(G^2)}$ and $\hat{f}_2 = \frac{a^2}{((a-1)\sum_{i=1}^{a} \frac{1}{n_i-1})} * \hat{f}_1$

The Lawley Hotelling Type statistic is given by: $U = tr[(a-1)H((N-a)G)^{-1}]$ Using the McKeon approximation the distribution of U is approximated by a "stretched" F distribution with degrees freedom K and D where:

$K = p(a-1)$ and

$D = 4 + (K+2)/(B-1)$ and

$B = ((N-p-2)(N-a-1))/((N-a-p)(N-a-p-n))$

The Bartlett Nanda Pillai Type statistic is given by: $V = tr\{(a-1)H[(a-1)H + (N-a)G]^{-1}\}$
McKeon approximated the distribution of $((V/\gamma)/\nu_1)/((1-V/\gamma)/\nu_2)$ using an F distribution with degrees freedom ν_1 and ν_2 where: $\gamma = min(a-1, p)$

$\nu_1 = (p(a-1))/(\gamma(N-1)) * [(\gamma(N-a+\gamma-p)(N-2)(N-1))/((N-a)(N-p)) - 2]$

$\nu_2 = (N-a+\gamma-p)/(N) * [(\gamma(N-a+\gamma-p)(N-1))/((N-a)(N-p)) - 2]$

The Wilks' Lambda Type Statistic is given by

$\lambda = det(((N-a) * G)/((N-a) * G + (a-1) * H)$

The F approximation statistic is given by

$F_\lambda = [(1 - \lambda^{1/t})/(\lambda^{1/t})](df_2/df_1)$

where

$df_1 = p(a-1)$ and $df_2 = rt - (p(a-1) - 2)/2$ and $r = (N-a) - (p - (a-1) + 1)/2$.

If $p(a-1) = 2$ then t=1, else $t = \sqrt{(p^2(a-1)^2 - 4)/(p^2 + (a-1)^2 - 5)}$

Note that regarding the above formula, there is a typo in the article Liu, Bathke, Harrar (2011).

Warning: The nonparametric methods implemented in the code have been developed for complete data with no missing values. The code automatically produces a warning if there is missing data.

Under certain conditions, the matrices H and G are singular (See literature for explanation of H and G), for example when the number of response variables exceeds the sample size. When this happens, only the ANOVA type statistic can be computed. The code automatically produces a warning if H or G are singular.

Author(s)

Woodrow Burchett, Amanda Ellis, Arne Bathke

References

Bathke AC, Harrar SW, Madden LV (2008). How to compare small multivariate samples using nonparametric tests. Computational Statistics and Data Analysis 52, 4951-4965

Brunner E, Domhof S, Langer F (2002). Nonparametric Analysis of Longitudinal Data in Factorial Experiments. Wiley, New York.

Liu C, Bathke AC, Harrar SW (2011). A nonparametric version of Wilks' lambda - Asymptotic results and small sample approximations. Statistics and Probability Letters 81, 1502-1506

Horst LE, Locke J, Krause CR, McMahaon RW, Madden LV, Hoitink HAJ (2005). Suppression of Botrytis blight of Begonia by Trichoderma hamatum 382 in peat and compost-amended potting mixes. Plant Disease 89, 1195-1200.

Marcus R, Peritz E, Gabriel KR (1976). On closed test procedures with special reference to ordered analysis of variance. Biometrika 63(3), 655-660.

Sonnemann E (2008). General solutions to multiple testing problems. Translation of "Sonnemann E (1982). Allgemeine Losungen multipler Testprobleme. EDV in Medizin und Biologie 13(4), 120-128". Biometrical Journal 50, 641-656.

Examples

```
data(sberry)
ssnonpartest(weight|bot|fungi|rating~treatment,sberry,test=c(1,0,0,0),alpha=.05,
            factors.and.variables=TRUE)
```

Index