Package 'RATest'

January 23, 2018

Type Package

Title Randomization Tests

Description A collection of randomization tests, data sets and examples. The current version focuses on two testing problems and their implementation in empirical work. First, it facilitates the empirical researcher to test for particular hypotheses, such as comparisons of means, medians, and variances from k populations using robust permutation tests, which asymptotic validity holds under very weak assumptions, while retaining the exact rejection probability in finite samples when the underlying distributions are identical. Second, the description and implementation of a permutation test for testing the continuity assumption of the baseline covariates in the sharp regression discontinuity design (RDD) as in Canay and Kamat (2017) <https://goo.gl/UZFqt7>. More specifically, it allows the user to select a set of covariates and test the aforementioned hypothesis using a permutation test based on the Cramer-von Miss test statistic. Graphical inspection of the empirical CDF and histograms for the variables of interest is also supported in the package.

Version 0.1.2

Maintainer Mauricio Olivares-Gonzalez <maurice.olivares@gmail.com>

Depends R (>= 3.00), ggplot2(>= 2.2.1), gridExtra

Imports stats, quantreg

License GPL (>= 2)

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

Author Mauricio Olivares-Gonzalez [aut, cre],
Ignacio Sarmiento-Barbieri [aut]

Collate 'CvM.stat.R' 'H.cdf.R' 'RDperm.R' 'RPT.R' 'lee2008.R'
'plot.RDperm.R' 'summary.RDperm.R' 'summary.RPT.R'

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2018-01-23 04:16:27 UTC

R topics documented:

`CvM.stat` *Cramer - von Mises statistics*

Description

Calculates the Cramer-von Mises test statistic

$$T(S_n) = \frac{1}{2q} \sum_{i=1}^{2q} \left(H_n^-(S_{n,i}) - H_n^+(S_{n,i}) \right)^2$$

where $H_n^-(\cdot)$ and $H_n^+(\cdot)$ are the empirical CDFs of the the sample of baseline covariates close to the cutoff from the left and right, respectively. See equation (12) in Canay and Kamat (2017).

Usage

```
CvM.stat(Sn)
```

Arguments

`Sn`	Numeric. The pooled sample of induced order statistics. The first column of S can be viewed as an independent sample of W conditional on Z being close to zero from the left. Similarly, the second column of S can be viewed as an independent sample of W conditional on Z being close to the cutoff from the right. See section 3 in Canay and Kamat (2017).

Value

Returns the numeric value of the Cramer - von Mises test statistic.

Author(s)

Maurcio Olivares Gonzalez

Ignacio Sarmiento Barbieri

References

Canay, I and Kamat V, (2017) Approximate Permutation Tests and Induced Order Statistics in the Regression Discontinuity Design. http://faculty.wcas.northwestern.edu/~iac879/wp/RDDPermutations.pdf

---

`H.cdf` *Regression Discontinuity Design Permutation test*

Description

Calculates the empirical CDF of the sample of W conditional on Z being close to the cutoff from either the left or right. Given the induced order for the baseline covariates

$$W_{[q]}^-, W_{[q-1]}^-, \ldots \leq W_{[1]}^-$$

or

$$W_{[1]}^+, W_{[2]}^+, \ldots, W_{[q]}^+$$

, this function will calculate either

$$H_n^-(t) = \frac{1}{q}\sum_{i=1}^{q} I\{W_{[i]}^- \leq t\}$$

or

$$H_n^+(t) = \frac{1}{q}\sum_{i=1}^{q} I\{W_{[i]}^+ \leq t\}$$

depending on the argument of the function. See section 3 in Canay & Kamat (2017).

Usage

```
H.cdf(W, t)
```

Arguments

`W`	Numeric. The sample of induced order statistics. The input can be either $\{W_{[q]}^-, W_{[q-1]}^-, \ldots, W_{[1]}^-\}$ or $\{W_{[1]}^+, W_{[2]}^+, \ldots, W_{[q]}^+\}$.
`t`	Numeric. The scalar needed for the calculation of the CDF.

Value

Numeric. For a sample $W = (w_1, \ldots, w_n)$, returns the fraction of observations less or equal to t.

Author(s)

Maurcio Olivares Gonzalez

Ignacio Sarmiento Barbieri

References

Canay, I and Kamat V, (2017) Approximate Permutation Tests and Induced Order Statistics in the Regression Discontinuity Design. http://faculty.wcas.northwestern.edu/~iac879/wp/RDDPermutations.pdf

---

`lee2008` *Dataset used in Lee (2008)*

Description

Randomized experiments from non-random selection in U.S. House elections

Format

A data frame with 6558 observations and two variables:

demsharenext Democrat vote share election t+1

difdemshare Running variable. Diff. democratic share

demshareprev Democrat vote share t-1

demwinprev Democrat win t-1

demofficeexp Democrat political experience t

othofficeexp Oppositions political experience t

demelectexp Democrat electoral experience t

othelectexp Oposition electoral experience t

Source

Mostly Harmless Econometrics Data Archive: http://economics.mit.edu/faculty/angrist/data1/mhe

References

Lee, D. (2008) Randomized experiments from non-random selection in U.S. House elections, *Journal of Econometrics*, 142, 675-697

`plot.RDperm` *Plot RDperm*

Description

Plots a histogram and empirical cdf

Usage

```
## S3 method for class 'RDperm'
plot(x, w, plot.class = "both", ...)
```

Arguments

`x`	Object of class "RDperm"
`w`	Character. Name of variable to be plotted
`plot.class`	Character. Can be: "both" for a histogram and cdf plot, "hist" for a histogram or "cdf" for only the cdf plot
`...`	Additional ggplot2 controls

Author(s)

Maurcio Olivares Gonzalez

Ignacio Sarmiento Barbieri

References

Canay, I and Kamat V, (2017) Approximate Permutation Tests and Induced Order Statistics in the Regression Discontinuity Design. http://faculty.wcas.northwestern.edu/~iac879/wp/RDDPermutations.pdf

Examples

```
## Not run:
permtest<-RDperm(W=c("demshareprev","demwinprev"),z="difdemshare",data=lee2008)
plot(permtest,w="demshareprev")

## End(Not run)
```

RDperm *Regression Discontinuity Design Permutation Test*

Description

A permutation test for continuity of covariates in Sharp Regression Discontinuity Design as described in Canay and Kamat (2017).

Usage

```
RDperm(W, z, data, n.perm = 499, q_type = 10, cutoff = 0,
  test.statistic = "CvM")
```

Arguments

W	Character. Vector of covariates names. The procedure will test the null hypothesis of continuity of the distribution of each element in W at the cutoff.
z	Character. Running variable name. This is the scalar random variable that defines, along with the cutoff, the treatment assignment rule in the sharp regression discontinuity design.
data	Data.frame.
n.perm	Numeric. Number of permutations needed for the stochastic approximation of the p-values. See remark 3.2 in Canay and Kamat (2017). The default is B=499.
q_type	A fixed and small (relative to the sample size) natural number that will define the q closest values of the order statistic of Z to the right and to the left of the cutoff. The default, 'rot', value is given by the feasible rule of thumb in footnote 4 of Canay and Kamat (2017), section 3.1. If 'arot', it calls for the Rule of Thumb described in equation (15) of Canay and Kamat (2017), section 3.1. The default option grows at a slower rate than the optional rule of thumb, but adds a larger constant.
cutoff	Numeric. The scalar defining the threshold of the running variable.
test.statistic	Character. A rank test statistic satisfying rank invariance. The default is a Cramer-von Mises test statistic.

Value

The functions summary and plot are used to obtain and print a summary and plot of the estimated regression discontinuity. The object of class RDperm is a list containing the following components:

results	Matrix. Test Statistic, P-values and Q
test.statistic	Test Statistic
q_type	Type of Q used in the calculations, can be either, "Defined by User", the "Rule of Thumb" or the "Alternative Rule of Thumb".
n_perm	number of permutations

rv	Character. Running variable name
Z	Vector. Running Variable
cutoff	cutoff
data	data set
S	Matrix. Pooled sample of induced order statistics
S_perm	List. Permutations of the induced order statistic.

Author(s)

Maurcio Olivares Gonzalez

Ignacio Sarmiento Barbieri

References

Canay, I and Kamat V, (2017) Approximate Permutation Tests and Induced Order Statistics in the Regression Discontinuity Design. http://faculty.wcas.northwestern.edu/~iac879/wp/RDDPermutations.pdf

Examples

```
permtest<-RDperm(W=c("demshareprev"),z="difdemshare",data=lee2008)
summary(permtest)
## Not run:
permtest<-RDperm(W=c("demshareprev","demwinprev"),z="difdemshare",data=lee2008)
summary(permtest)

## End(Not run)
```

RPT *Robust Permutation Test*

Description

This function considers the general problem of inference from the permutation distribution when comparing parameters from k populations. The test statistics will be based on the difference of estimators that are asymptotically linear. For illustrative purposes we will consider here the 2 sample case, but the function works for k-samples.

Difference of means: Here, the null hypothesis is of the form $H_0 : \mu(P) - \mu(Q) = 0$, and the corresponding test statistic is given by

$$T_{m,n} = \frac{N^{1/2}(\bar{X}_m - \bar{Y}_n)}{\sqrt{\frac{N}{m}\sigma_m^2(X_1, \ldots, X_m) + \frac{N}{n}\sigma_n^2(Y_1, \ldots, Y_n)}}$$

where $\bar{X}_m$ and $\bar{Y}_n$ are the sample means from population P and population Q, respectively, and $\sigma_m^2(X_1, \ldots, X_m)$ is a consistent estimator of $\sigma^2(P)$ when $X_1, \ldots, X_m$ are i.i.d. from P. Assume consitency also under Q.

Difference of medians: Let F and G be the CDFs corresponding to P and Q, and denote $\theta(F)$ the median of F i.e. $\theta(F) = \inf\{x : F(x) \geq 1/2\}$. Assume that F is continuously differentiable at $\theta(P)$ with derivative F' (and the same with F replaced by G). Here, the null hypothesis is of the form $H_0 : \theta(P) - \theta(Q) = 0$, and the corresponding test statistic is given by

$$T_{m,n} = \frac{N^{1/2}\left(\theta(\hat{P}_m) - \theta(\hat{Q})\right)}{\hat{\upsilon}_{m,n}}$$

where $\hat{\upsilon}_{m,n}$ is a consistent estimator of $\upsilon(P, Q)$:

$$\upsilon(P, Q) = \frac{1}{\lambda}\frac{1}{4(F'(\theta))^2} + \frac{1}{1-\lambda}\frac{1}{4(G'(\theta))^2}$$

Choices of $\hat{\upsilon}_{m,n}$ may include the kernel estimator of Devroye and Wagner (1980), the bootstrap estimator of Efron (1992), or the smoothed bootstrap Hall et al. (1989) to list a few. For further details, see Chung and Romano (2013). Current implementation uses the bootstrap estimator of Efron (1992)

Difference of variances: Here, the null hypothesis is of the form $H_0 : \sigma^2(P) - \sigma^2(Q) = 0$, and the corresponding test statistic is given by

$$T_{m,n} = \frac{N^{1/2}(\hat{\sigma}_m^2(X_1, \ldots, X_,) - \hat{\sigma}_n^2(Y_1, \ldots, Y_n))}{\sqrt{\frac{N}{m}(\hat{\mu}_{4,x} - \frac{(m-3)}{(m-1)}(\hat{\sigma}_m^2)^2) + \frac{N}{n}(\hat{\mu}_{4,y} - \frac{(n-3)}{(n-1)}(\hat{\sigma}_y^2)^2)}}$$

where $\hat{\mu}_{4,m}$ the sample analog of $E(X-\mu)^4$ based on an iid sample $X_1, \ldots, X_m$ from P. Similarly for $\hat{\mu}_{4,n}$.

Usage

```
RPT(formula, data, test = "means", n.perm = 499, na.action)
```

Arguments

`formula`	a formula object, in which the response variable is on the left of a ~ operator, and the groups on the right.
`data`	a data.frame containing the named variables needed for the formula. If this argument is missing, then the variables in the formula should be on the search list.
`test`	testing problem. It admits "means" if the objective is to test for difference of Means, "medians" for difference of Medians, and "variances" for difference of Variances. In the case the user is interested in testing for difference of medians, the Efron (1992) bootstrap estimator is used to estimate the variances (For further details, see Chung and Romano (2013))
`n.perm`	Numeric. Number of permutations needed for the stochastic approximation of the p-values. The default is n.perm=499.
`na.action`	a function to filter missing data. This is applied to the model.frame . The default is na.omit, which deletes observations that contain one or more missing values.

Value

An object of class "RPT" is a list containing at least the following components:

description	Type of test, can be Difference of Means, Medians, or Variances.
n_populations	Number of grups.
N	Sample Size.
T.obs	Observed test statistic.
pvalue	P-value.
T.perm	Vector. Test statistics calculated from the permutations of the data.
n_perm	Number of permutations.
parameters	Estimated parameters.
sample_sizes	Sample size of groups.

Author(s)

Maurcio Olivares Gonzalez

Ignacio Sarmiento Barbieri

References

Chung, E. and Romano, J. P. (2013). Exact and asymptotically robust permutation tests. The Annals of Statistics, 41(2):484–507.

Chung, E. and Romano, J. P. (2016). Asymptotically valid and exact permutation tests based on two-sample u-statistics. Journal of Statistical Planning and Inference, 168:97–105.

Devroye, L. P. and Wagner, T. J. (1980). The strong uniform consistency of kernel density estimates. In Multivariate Analysis V: Proceedings of the fifth International Symposium on Multivariate Analysis, volume 5, pages 59–77.

Efron, B. (1992). Bootstrap methods: another look at the jackknife. In Breakthroughs in statistics, pages 569–593. Springer. Hall, P., DiCiccio, T. J., and Romano, J. P. (1989). On smoothing and the bootstrap. The Annals of Statistics, pages 692–704.

Examples

```
## Not run:
male<-rnorm(50,1,1)
female<-rnorm(50,1,2)
dta<-data.frame(group=c(rep(1,50),rep(2,50)),outcome=c(male,female))
rpt.var<-RPT(dta$outcome~dta$group,test="variances")
summary(rpt.var)


## End(Not run)
```

summary.RDperm *Summarizing Regression Discontinuity Design Permutation Test*

Description

summary method for class "RDPerm"

Usage

```
## S3 method for class 'RDperm'
summary(object, digits = max(3, getOption("digits") - 3),
  ...)
```

Arguments

object	an object of class "RDperm", usually a result of a call to RDperm
digits	number of digits to display
...	unused

Value

summary.RDperm returns an object of class "summary.RDperm" which has the following components

results	Matrix with the Test Statistic, P-values and Q used

Author(s)

Maurcio Olivares Gonzalez

Ignacio Sarmiento Barbieri

summary.RPT *Summarizing Robust Permutation Tests*

Description

summary method for class "RPT"

Usage

```
## S3 method for class 'RPT'
summary(object, digits = max(3, getOption("digits") - 3), ...)
```

Arguments

object	an object of class "RPT", usually a result of a call to RPT
digits	number of digits to display
...	unused

Value

summary.RPT returns an object of class "summary.RPT" which has the following components:

T.obs	Observed test statistic.
pvalue	P-value.

Author(s)

Maurcio Olivares Gonzalez

Ignacio Sarmiento Barbieri

Index