Package 'QDComparison'

October 12, 2022

Type Package

Title Modern Nonparametric Tools for Two-Sample Quantile and
Distribution Comparisons

Version 3.0

Date 2019-06-24

Author David Jungreis, Subhadeep Mukhopadhyay

Maintainer David Jungreis <dbjungreis@gmail.com>

Description
Allows practitioners to determine (i) if two univariate distributions (which can be continuous, discrete, or even mixed) are equal, (ii) how two distributions differ (shape differences, e.g., location, scale, etc.), and (iii) where two distributions differ (at which quantiles), all using nonparametric LP statistics. The primary reference is Jungreis, D. (2019, Technical Report).

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2019-06-24 04:50:04 UTC

Depends R (>= 3.5.0)

R topics documented:

QDComparison-package | *Modern Nonparametric Tools for Two-Sample Quantile and Distribution Comparisons*

Description

Allows practitioners to determine (i) if two univariate distributions (which can be continuous, discrete, or even mixed) are equal, (ii) how two distributions differ (shape differences, e.g., location, scale, etc.), and (iii) where two distributions differ (at which quantiles), all using nonparametric LP statistics.

Author(s)

David Jungreis, Subhadeep Mukhopadhyay

Maintainer: David Jungreis <dbjungreis@gmail.com>

References

Jungreis, D., (2019) "Unification of Continuous, Discrete, and Mixed Distribution Two-Sample Testing with Inferences in the Quantile Domain"

Mukhopadhyay, S., (2013) "Nonparametric Inference for High Dimensional Data,"" Ph.D. diss., Texas A&M University, College Station, Texas.

Mukhopadhyay, S. and Parzen, E. (2014), "LP Approach to Statistical Modeling", arXiv:1405.2601.

Examples

```
x <- c(rep(0,200),rep(1,200))
y <- c(rnorm(200,0,1),rnorm(200,1,1))
L <- LP.QDC(x,y)
L$pval
```

Depression | *Jackson's CESD Depression Scores*

Description

The data come from Jackson's (2009) depression data, used by Wilcox (2014).

Usage

```
data(Depression)
```

Format

A data frame with 372 observations on the following 2 variables.

x A binary indicator variable: 0 for control, 1 for intervention (received therapy)

y The response variable: CESD score (higher means more depressed)

References

Jackson, J., Mandel, D., Blanchard, J., Carlson, M., Cherry, B., Azen, S., Chou, C.P.,Jordan-Marsh, M., Forman, T., White, B., et al. (2009), "Confronting challenges in intervention research with ethnically diverse older adults: the USC Well Elderly II trial," Clinical Trials, 6, 90-101.

Wilcox, R. R., Erceg-Hurn, D. M., Clark, F., and Carlson, M. (2014), "Comparing two independent groups via the lower and upper quantiles," Journal of Statistical Computation and Simulation, 84, 1543-1551.

Examples

```
data(Depression)
## maybe str(Depression)
y <- Depression[,2]
x <- Depression[,1]
hist(y[x==1])
```

Earnings1978   *LaLonde's 1978 Earnings Data*

Description

These data come from LaLonde's (1986) National Supported Work Demonstration (NSW) Data (Dehejia-Wahha Sample (1999)), used by Firpo (2007).

Usage

```
data(Earnings1978)
```

Format

A data frame with 445 observations on the following 2 variables.

x A binary indicator variable: 0 for control, 1 for intervention (received job training)

y The response variable: earnings in 1978

Source

http://users.nber.org/~rdehejia/data/nswdata2.html

References

Dehejia, R. H. and Wahba, S. (1999), "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs," Journal of the American Statistical Association, 94, 1053-1062.

Firpo, S. (2007), "Efficient semiparametric estimation of quantile treatment effects," Econometrica, 75, 259-276.

LaLonde, R. J. (1986), "Evaluating the econometric evaluations of training programs with experimental data," The American Economic Review, 604-620.

Examples

```
data(Earnings1978)
## maybe str(Earnings1978)
y <- Earnings1978[,2]
x <- Earnings1978[,1]
hist(y[x==1])
```

---

eLP.poly   *A function to compute the LP basis functions*

Description

Given a random sample from X (which can be discrete, continuous, or even mixed), this function computes the empirical LP-basis functions.

Usage

```
eLP.poly(x,m)
```

Arguments

x	The random samples
m	Number of basis functions to compute

Value

LP basis functions

Author(s)

Subhadeep Mukhopadhyay

References

Jungreis, D., (2019) "Unification of Continuous, Discrete, and Mixed Distribution Two-Sample Testing with Inferences in the Quantile Domain"

Mukhopadhyay, S. and Parzen, E. (2014), "LP Approach to Statistical Modeling", arXiv:1405.2601.

Examples

```
x <- c(rep(0,200),rep(1,200))
m <- 6
eLP.poly(x,m)
```

Fundraising	*Gneezy's Fundraising Data with a Gift Wage*

Description

These data come from Gneezy's (2006) fundraising experiment, on which Goldman (2018) performed quantile treatment effect analysis. These data correspond to the "pre-lunch" period.

Usage

```
data(Fundraising)
```

Format

A data frame with 23 observations on the following 2 variables.

x A binary indicator variable: 0 for control, 1 for intervention (gift wage)

y The response variable: dollars raised

Source

Gneezy, U. and List, J. A. (2006), "Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments," Econometrica, 74, 1365-1384.

References

Gneezy, U. and List, J. A. (2006), "Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments," Econometrica, 74, 1365-1384.

Goldman, M. and Kaplan, D. M. (2018), "Comparing distributions by multiple testing across quantiles or CDF values," Journal of Econometrics, Volume 206, Issue 1, 143-166.

Examples

```
data(Fundraising)
## maybe str(Fundraising)
y <- Fundraising[,2]
x <- Fundraising[,1]
hist(y[x==1])
```

LP.QDC   *The main function for two-sample quantile and distribution comparison*

Description

This function runs the entire quantile and distribution comparison, giving LP comoments, LP coefficients, LPINFOR test statistic, p-value, estimated comparison density with null-band, and intervals where the comparison density is above or below the null band

Usage

```
LP.QDC(x,y,m=6,smooth="TRUE",method="BIC",alpha=0.05,
       B=1000,spar=NA,plot="TRUE",inset=-0.2)
```

Arguments

x	Indicator variable denoting group membership
y	Response variable with measured values
m	Number of LP comoments and LP coefficients to be calculated, default: 6
smooth	If smoothing should be applied, default: TRUE
method	Smoothing method as AIC or BIC, default: BIC
alpha	Alpha-level for confidence bands, default: 0.05
B	Number of permutations of the x labels, default: 1000
spar	"spar" in "smooth.spline" of upper and lower bounds of confidence bands, default: NA, let smooth.splines function figure it out
plot	Should plotting be performed, default: TRUE
inset	Graphical parameter that controls where the color legend is plotted below x-axis, default: -0.2

Value

A list containing:

band	y-values of the upper and lower bounds of the confidence band
d.hat	y-values of the comparison density
sparL	"spar" value in "smooth.spline" of lower bound of the null band
sparU	"spar" value in "smooth.spline" of upper bound of the null band
out.above	Quantile intervals where group 1 dominates the pooled distribution
out.below	Quantile intervals where group 0 dominates the pooled distribution
LP.comoment.0	LP comoments, conditioned on X=0
LP.coef.0	LP coefficients, conditioned on X=0

LP.comoment.1	LP comoments, conditioned on X=1
LP.coef.1	LP coefficients, conditioned on X=1
LPINFOR	Test statistics value
pval	The p-value for testing equality of two distributions F0=F1

Author(s)

David Jungreis

Subhadeep Mukhopadhyay

References

Jungreis, D., (2019) "Unification of Continuous, Discrete, and Mixed Distribution Two-Sample Testing with Inferences in the Quantile Domain"

Mukhopadhyay, S. and Parzen, E. (2014), "LP Approach to Statistical Modeling", arXiv:1405.2601.

Examples

```
x <- c(rep(0,200),rep(1,200))
y <- c(rnorm(200,0,1),rnorm(200,1,1))
L <- LP.QDC(x,y)
L$pval
```

LP.XY	*A function to compute LP comoments, LP coefficients, LPINFOR test statistic, and a p-value of distribution equality*

Description

This function computes LP comoments, LP coefficients, LPINFOR test statistic, and the corresponding p-value of for testing equality of two distributions.

Usage

```
LP.XY(x,y,m=6,smooth="TRUE",method="BIC")
```

Arguments

x	Indicator variable denoting group membership
y	Response variable with measured values
m	Number of LP comoments and LP coefficients to be calculated, default: 6
smooth	If smoothing should be applied, default: TRUE
method	Smoothing method, default: BIC

Value

A list containing:

`LP.comoment.0`	LP comoments, conditioned on X=0
`LP.coef.0`	LP coefficients, conditioned on X=0
`LP.comoment.1`	LP comoments, conditioned on X=1
`LP.coef.1`	LP coefficients, conditioned on X=1
`LPINFOR`	Test statistics value
`pval`	The p-value for testing equality of two distributions F0=F1

Author(s)

Subhadeep Mukhopadhyay

David Jungreis

References

Jungreis, D., (2019) "Unification of Continuous, Discrete, and Mixed Distribution Two-Sample Testing with Inferences in the Quantile Domain"

Mukhopadhyay, S. and Parzen, E. (2014), "LP Approach to Statistical Modeling", arXiv:1405.2601.

Examples

```
x <- c(rep(0,200),rep(1,200))
y <- c(rnorm(200,0,1),rnorm(200,1,1))
L <- LP.XY(x,y)
L$pval
```

Microfinance	*Informal Borrowing in Neighborhoods of Hyderabad, India*

Description

These data come from Banerjee's (2015) informal borrowing observations.

Usage

```
data(Microfinance)
```

Format

A data frame with 6811 observations on the following 2 variables.

x A binary indicator variable: 0 for control, 1 for intervention (access to microfinance)

y The response variable: rupees informally borrowed

Source

https://www.aeaweb.org/articles?id=10.1257/app.20130533

References

Banerjee, A., Duflo, E., Glennerster, R., and Kinnan, C. (2015), "The miracle of microfinance? Evidence from a randomized evaluation," American Economic Journal: Applied Economics, 7, 22-53.

Examples

```
data(Microfinance)
## maybe str(Microfinance)
y <- Microfinance[,2]
x <- Microfinance[,1]
# Remove the 0s (as Banerjee (2015) appears to have done)
ind <- which(y==0)
x <- x[-ind]
y <- y[-ind]
hist(y[x==0])
```

NMES   *National Medicare Expenditure Survey (NMES) Data on Cost of Hospitalizations*

Description

These data come from Venturini's (2015) study of hospital costs for patients with smoking and non-smoking diseases.

Usage

```
data(NMES)
```

Format

A data frame with 9416 observations on the following 2 variables.

x A binary indicator variable: 0 for non-smoking disease, 1 for smoking disease

y The response variable: cost of a hospital stay, in dollars

References

Dominici, F., Cope, L., Naiman, D. Q., and Zeger, S. L. (2005), "Smooth quantile ratio estimation," Biometrika, 92, 543-557.

Dominici, F. and Zeger, S. L. (2005), "Smooth quantile ratio estimation with regression: estimating medical expenditures for smoking-attributable diseases," Biostatistics, 6, 505-519.

Johnson, E., Dominici, F., Griswold, M., and Zeger, S. L. (2003), "Disease cases and their medical costs attributable to smoking: an analysis of the national medical expenditure survey," Journal of Econometrics, 112, 135-151.

Venturini, S., Dominici, F., Parmigiani, G., et al. (2015), "Generalized quantile treatment effect: A flexible Bayesian approach using quantile ratio smoothing," Bayesian Analysis, 10, 523-552.

Examples

```
data(NMES)
## maybe str(NMES)
y <- NMES[,2]
x <- NMES[,1]
# Remove the 0s (as Venturini (2015) notes was necessary)
ind <- which(y==0)
x <- x[-ind]
y <- y[-ind]
hist(y[x==0])
```

Index