Package 'randChecks'

May 9, 2026

Type Package

Title Covariate Balance Checks: Randomization Tests and Graphical Diagnostics

Version 0.2.1

Author Zach Branson

Maintainer Zach Branson <zach@stat.cmu.edu>

Description Provides randomization tests and graphical diagnostics for assessing randomized assignment and covariate balance for a binary treatment variable. See Branson (2021) <doi:10.48550/arXiv.1804.08760> for details.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

NeedsCompilation no

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2021-03-18 23:20:02 UTC

Contents

asIfRandPlot . . . 2
asIfRandTest . . . 5
getCovMeanDiffs . . . 9
getMD . . . 10
getStandardizedCovMeanDiffs . . . 11
lalonde . . . 13
lalonde.matched.card . . . 14
lalonde.matched.ps . . . 15
lalondeMatches . . . 16
lovePlot . . . 17
lovePlotCompare . . . 21

Index **23**

`asIfRandPlot` *Graphical Diagnostic of As-If Randomization for Different Assignment Mechanisms*

Description

`asIfRandPlot` produces a plot showing the distribution of the Mahalanobis distance for different assignment mechanisms, along with the observed Mahalanobis distance. If the observed Mahalanobis distance is well within the range of a particular distribution, then that suggests that a particular assignment mechanism holds. This function supports the following assignment mechanisms:

- Complete randomization ("complete"): Corresponds to random permutations of the indicator across units.
- Block randomization ("blocked"): Corresponds to random permutations of the indicator within blocks of units.
- Constrained-differences randomization ("constrained diffs"): Corresponds to random permutations of the indicator across units, conditional on the standardized covariate mean differences being below some threshold.
- Constrained-Mahalanobis randomization ("constrained md"): Corresponds to random permutations of the indicator across units, conditional on the Mahalanobis being below some threshold.
- Blocked Constrained-differences randomization ("blocked constrained diffs"): Corresponds to random permutations of the indicator within blocks of units, conditional on the standardized covariate mean differences being below some threshold.
- Blocked Constrained-Mahalanobis randomization ("blocked constrained md"): Corresponds to random permutations of the indicator within blocks of units, conditional on the Mahalanobis being below some threshold.

Usage

```
asIfRandPlot(X.matched, indicator.matched,
  assignment = c("complete"),
  subclass = NULL, threshold = NULL,
  perms = 1000,
  X.full = NULL, indicator.full = NULL)
```

Arguments

`X.matched` A covariate matrix (rows correspond to subjects/units; columns correspond to covariates) for the matched dataset.

`indicator.matched` A vector of 1s and 0s (e.g., denoting treatment and control) for the matched dataset.

assignment	A vector of assignment mechanisms that the user wants to visualize; the user can test one assignment mechanism or multiple. The possible choices are "complete", "blocked", "constrained diffs", "constrained md", "blocked constrained diffs", and "blocked constrained md". See Description for more details on these assignment mechanisms.
subclass	A vector denoting the subclass/block for each subject/unit. This must be specified only if one of the blocked assignment mechanisms are used.
threshold	The threshold used within the constrained assignment mechanisms; thus, this must be specified only if one of the constrained assignment mechanisms are used. This can be a single number or a vector of numbers (e.g., if one wants to use a different threshold for each covariate when testing constrained-differences randomization).
perms	The number of permutations used within the randomization test. A larger number requires more computation time but results in a more consistent p-value.
X.full	A covariate matrix (rows correspond to subjects/units; columns correspond to covariates) for the full, unmatched dataset if available.
indicator.full	A vector of 1s and 0s (e.g., denoting treatment and control) for the full, unmatched dataset if available.

Details

The arguments X.full and indicator.full (i.e., the covariate matrix and indicator for the full, unmatched dataset) are only used to correctly define the standardized covariate mean differences and Mahalanobis distance. Technically, the covariate mean differences should be standardized by the pooled variance within the full, unmatched dataset, instead of within the matched dataset. If X.full and indicator.full are unspecified, the pooled variance within the matched dataset is used for standardization instead. This distinction rarely leads to large differences in the resulting standardized covariate mean differences, and so researchers should feel comfortable only specifying X.matched and indicator.matched if only a matched dataset is available. Furthermore, if one wants to make this plot for a full, unmatched dataset, then they should only specify X.matched and indicator.matched.

Value

A plot showing the distribution of the Mahalanobis distance for different assignment mechanisms, along with the observed Mahalanobis distance. Also returns a p-value for each assignment mechanism - this is simply the area of the distribution more extreme than the observed Mahalanobis distance. This is the same as `asIfRandTest()` using the Mahalanobis distance as a test statistic.

Author(s)

Zach Branson

Examples

```
#This loads the classic Lalonde (1986) dataset,
#as well as two matched datasets:
#one from 1:1 propensity score matching,
#and one from cardinality matching, where
```

```
#the standardized covariate mean differences are all below 0.1.
data("lalondeMatches")

#obtain the covariates for these datasets
X.lalonde = subset(lalonde, select = -c(treat))
X.matched.ps = subset(lalonde.matched.ps, select = -c(treat,subclass))
X.matched.card = subset(lalonde.matched.card, select = -c(treat,subclass))
#the treatment indicators are
indicator.lalonde = lalonde$treat
indicator.matched.ps = lalonde.matched.ps$treat
indicator.matched.card = lalonde.matched.card$treat

#the subclass for the matched datasets are
subclass.matched.ps = lalonde.matched.ps$subclass
subclass.matched.card = lalonde.matched.card$subclass


#The following lines of code create diagnostic plots assessing
#whether the treatment follows different assignment mechanisms.

#Note that the following examples only use 100 permutations
#to approximate the randomization distribution.
#In practice, we recommend setting perms = 1000 or more;
#in these examples we use perms = 50 to save computation time.

#Assessing complete randomization for the full dataset
#Here, complete randomization clearly does not hold,
#because the observed Mahalanobis distance is far outside
#the complete randomization distribution.
asIfRandPlot(X.matched = X.lalonde, indicator.matched = indicator.lalonde, perms = 50)

#Assessing complete and block (paired) randomization for
#the propensity score matched dataset
#Again, complete and block randomization appear to not hold
#because the observed Mahalanobis distance is far outside
#the randomization distributions.
asIfRandPlot(X.matched = X.matched.ps, indicator.matched = indicator.matched.ps,
X.full = X.lalonde, indicator.full = indicator,
   assignment = c("complete", "blocked"),
   subclass = lalonde.matched.ps$subclass,
   perms = 50)

#Assessing three assignment mechanisms for the
#cardinality matched dataset:
# 1) complete randomization
# 2) blocked (paired) randomization
# 3) constrained-MD randomization
#Note that the Mahalanobis distance is approximately a chi^2_K distribution,
#where K is the number of covariates. In the Lalonde data, K = 8.
#Thus, the threshold can be chosen as the quantile of the chi^2_8 distribution.
#This threshold constrains the Mahalanobis distance to be below the 25-percent quantile:
a = qchisq(p = 0.25, df = 8)
#Then, we can assess these three assignment mechanisms with the plot below.
```

```
#Here, these assignment mechanisms seem plausible,
#because the observed Mahalanobis distance is well
#within the randomization distributions.
asIfRandPlot(X.matched = X.matched.card, indicator.matched = indicator.matched.card,
X.full = X.lalonde, indicator.full = indicator,
   assignment = c("complete", "blocked", "constrained md"),
   subclass = lalonde.matched.card$subclass,
   threshold = a,
   perms = 50)
```

`asIfRandTest` *As-If Randomization Test for Different Assignment Mechanisms: Global Tests and Covariate-by-Covariate Tests*

Description

`asIfRandTest` computes p-values testing whether an indicator follows a given assignment mechanism, based on observed covariates. This function supports the following assignment mechanisms:

- Complete randomization ("complete"): Corresponds to random permutations of the indicator across units.
- Block randomization ("blocked"): Corresponds to random permutations of the indicator within blocks of units.
- Constrained-differences randomization ("constrained diffs"): Corresponds to random permutations of the indicator across units, conditional on the standardized covariate mean differences being below some threshold.
- Constrained-Mahalanobis randomization ("constrained md"): Corresponds to random permutations of the indicator across units, conditional on the Mahalanobis being below some threshold.
- Blocked Constrained-differences randomization ("blocked constrained diffs"): Corresponds to random permutations of the indicator within blocks of units, conditional on the standardized covariate mean differences being below some threshold.
- Blocked Constrained-Mahalanobis randomization ("blocked constrained md"): Corresponds to random permutations of the indicator within blocks of units, conditional on the Mahalanobis being below some threshold.

The null hypothesis is that the assignment mechanism holds. A large p-value does not prove that the assumption holds, but a small p-value implies that the assumption doesn't hold. These p-values are exact, in the sense that they only rely on permutations within the data and not asymptotic approximations.

In addition to specifying different assignment mechanisms, the user can specify two different test statistics:

- The Mahalanobis distance ("mahalanobis"). This acts as a global test statistic, and thus only one p-value is computed.
- The standardized covariate mean differences ("diffs"). This acts as a covariate-by-covariate test statistic, and thus a p-value for each covariate is computed.

Usage

```
asIfRandTest(X.matched, indicator.matched,
  assignment = c("complete"),
  statistic = "mahalanobis",
  subclass = NULL, threshold = NULL,
  perms = 1000,
  X.full = NULL, indicator.full = NULL)
```

Arguments

`X.matched`	A covariate matrix (rows correspond to subjects/units; columns correspond to covariates) for the matched dataset.
`indicator.matched`	A vector of 1s and 0s (e.g., denoting treatment and control) for the matched dataset.
`assignment`	A vector of assignment mechanisms that the user wants to test; the user can test one assignment mechanism or multiple. The possible choices are "complete", "blocked", "constrained diffs", "constrained md", "blocked constrained diffs", and "blocked constrained md". See Description for more details on these assignment mechanisms.
`statistic`	The test statistic used in the randomization test. The choices are either "mahalanobis" (the Mahalanobis distance) or "diffs" (the standardized covariate mean differences). The former runs a global test and provides one p-value; the latter runs covariate-by-covariate tests and provides a p-value for each covariate.
`subclass`	A vector denoting the subclass/block for each subject/unit. This must be specified only if one of the blocked assignment mechanisms are used.
`threshold`	The threshold used within the constrained assignment mechanisms; thus, this must be specified only if one of the constrained assignment mechanisms are used. This can be a single number or a vector of numbers (e.g., if one wants to use a different threshold for each covariate when testing constrained-differences randomization).
`perms`	The number of permutations used within the randomization test. A larger number requires more computation time but results in a more consistent p-value.
`X.full`	A covariate matrix (rows correspond to subjects/units; columns correspond to covariates) for the full, unmatched dataset if available.
`indicator.full`	A vector of 1s and 0s (e.g., denoting treatment and control) for the full, unmatched dataset if available.

Details

The arguments X.full and indicator.full (i.e., the covariate matrix and indicator for the full, unmatched dataset) are only used to correctly define the standardized covariate mean differences and Mahalanobis distance. Technically, the covariate mean differences should be standardized by the pooled variance within the full, unmatched dataset, instead of within the matched dataset. If X.full and indicator.full are unspecified, the pooled variance within the matched dataset is used for standardization instead. This distinction rarely leads to large differences in the resulting standardized

covariate mean differences, and so researchers should feel comfortable only specifying X.matched and indicator.matched if only a matched dataset is available. Furthermore, if one wants to run this test for a full, unmatched dataset, then they should only specify X.matched and indicator.matched.

Value

p-values assessing as-if randomization of an indicator for different assignment mechanisms. If the Mahalanobis distance is used as a test statistic, then a vector of p-values is reported is reported (one for each assignment mechanism). If the standardized covariate mean differences are used as a test statistic, then a table of p-values is reported, where the rows correspond to assignment mechanisms and the columns correspond to covariates.

Author(s)

Zach Branson

Examples

```
#This loads the classic Lalonde (1986) dataset,
#as well as two matched datasets:
#one from 1:1 propensity score matching,
#and one from cardinality matching, where
#the standardized covariate mean differences are all below 0.1.
data("lalondeMatches")

#obtain the covariates for these datasets
X.lalonde = subset(lalonde, select = -c(treat))
X.matched.ps = subset(lalonde.matched.ps, select = -c(treat,subclass))
X.matched.card = subset(lalonde.matched.card, select = -c(treat,subclass))
#the treatment indicators are
indicator.lalonde = lalonde$treat
indicator.matched.ps = lalonde.matched.ps$treat
indicator.matched.card = lalonde.matched.card$treat

#the subclass for the matched datasets are
subclass.matched.ps = lalonde.matched.ps$subclass
subclass.matched.card = lalonde.matched.card$subclass

#Note that the following examples only use 100 permutations
#to approximate the randomization distribution.
#In practice, we recommend setting perms = 1000 or more;
#in these examples we use perms = 50 to save computation time.

#testing complete randomization for the full dataset
#using the Mahalanobis distance.
#We reject complete randomization in this test.
asIfRandTest(X.matched = X.lalonde, indicator.matched = indicator.lalonde, perms = 50)
#testing complete randomization for the full dataset
#using standardized covariate mean differences.
#We reject complete randomization for most covariates:
asIfRandTest(X.matched = X.lalonde, indicator.matched = indicator.lalonde,
   statistic = "diffs",
```

```
   perms = 50)

#testing complete randomization and block (paired) randomization
#for the propensity score matched dataset
#using the Mahalanobis distance.
#We reject both assignment mechanisms in this test.
asIfRandTest(X.matched = X.matched.ps, indicator.matched = indicator.matched.ps,
X.full = X.lalonde, indicator.full = indicator.lalonde,
   assignment = c("complete", "blocked"),
subclass = lalonde.matched.ps$subclass,
perms = 50)
#testing complete randomization and block (paired) randomization
#for the propensity score matched dataset
#using the standardized covariate mean differences.
#We reject these assignment mechanisms for
#the race covariates (hispan and black):
asIfRandTest(X.matched = X.matched.ps, indicator.matched = indicator.matched.ps,
X.full = X.lalonde, indicator.full = indicator.lalonde,
   assignment = c("complete", "blocked"),
   subclass = lalonde.matched.ps$subclass,
statistic = "diffs",
perms = 50)

#testing three assignment mechanisms for
#the cardinality matched dataset:
# 1) complete randomization
# 2) blocked (paired) randomization
# 3) constrained-MD randomization
#Note that the Mahalanobis distance is approximately a chi^2_K distribution,
#where K is the number of covariates. In the Lalonde data, K = 8.
#Thus, the threshold can be chosen as the quantile of the chi^2_8 distribution.
#This threshold constrains the Mahalanobis distance to be below the 25-percent quantile:
a = qchisq(p = 0.25, df = 8)
#First we'll run the test using the Mahalanobis distance.
#We fail to reject for the first two assignment mechanisms,
#but reject the third.
asIfRandTest(X.matched = X.matched.card, indicator.matched = indicator.matched.card,
X.full = X.lalonde, indicator.full = indicator.lalonde,
   assignment = c("complete", "blocked", "constrained md"),
   subclass = lalonde.matched.card$subclass,
   threshold = a,
   perms = 50)
   #Now we'll run the test using the standardized covariate mean differences.
   #Interestingly, you fail to reject for all three assignment mechanisms
   #for all covariates:
   asIfRandTest(X.matched = X.matched.card, indicator.matched = indicator.matched.card,
   X.full = X.lalonde, indicator.full = indicator.lalonde,
   assignment = c("complete", "blocked", "constrained md"),
   subclass = lalonde.matched.card$subclass,
   threshold = a,
   statistic = "diffs",
   perms = 50)
```

getCovMeanDiffs *Covariate Mean Differences*

Description

getCovMeanDiffs computes the covariate mean differences between a treatment and control group.

Usage

```
getCovMeanDiffs(X, indicator)
```

Arguments

X	A covariate matrix (rows correspond to subjects/units; columns correspond to covariates).
indicator	A vector of 1s and 0s (e.g., denoting treatment and control).

Value

The covariate mean differences between a treatment and control group, defined as treatment minus control.

Author(s)

Zach Branson

See Also

See also lalondeMatches for details about the Lalonde and matched datasets.

Examples

```
#This loads the classic Lalonde (1986) dataset,
#as well as two matched datasets:
#one from 1:1 propensity score matching,
#and one from cardinality matching, where
#the standardized covariate mean differences are all below 0.1.
data("lalondeMatches")

#obtain the covariates for these datasets
X.lalonde = subset(lalonde, select = -c(treat))
X.matched.ps = subset(lalonde.matched.ps, select = -c(treat,subclass))
X.matched.card = subset(lalonde.matched.card, select = -c(treat,subclass))
#the treatment indicators are
indicator.lalonde = lalonde$treat
indicator.matched.ps = lalonde.matched.ps$treat
indicator.matched.card = lalonde.matched.card$treat

#the covariate mean differences are:
```

```
getCovMeanDiffs(X = X.lalonde, indicator = indicator.lalonde)
getCovMeanDiffs(X = X.matched.ps, indicator = indicator.matched.ps)
getCovMeanDiffs(X = X.matched.card, indicator = indicator.matched.card)
```

`getMD` *Mahalanobis Distance*

Description

`getMD` computes the Mahalanobis distance of the covariate means between a treatment and control group.

Usage

```
getMD(X.matched, indicator.matched,
covX.inv = NULL,
X.full = NULL)
```

Arguments

`X.matched`	A covariate matrix (rows correspond to subjects/units; columns correspond to covariates) for the matched dataset.
`indicator.matched`	A vector of 1s and 0s (e.g., denoting treatment and control) for the matched dataset.
`covX.inv`	The inverse of X's covariance matrix. Almost always this should be set to NULL, and `getMD` will compute the inverse of the covariance matrix automatically.
`X.full`	A covariate matrix (rows correspond to subjects/units; columns correspond to covariates) for the full, unmatched dataset if available.

Details

The argument X.full (i.e., the covariate matrix for the full, unmatched dataset) is only used to correctly define the Mahalanobis distance after matching. Technically, the Mahalanobis distance should be standardized by the covariance matrix within the full, unmatched dataset, instead of within the matched dataset. If X.full is unspecified, the covariance matrix within the matched dataset is used instead. This distinction rarely leads to large differences in the resulting distance, and so researchers should feel comfortable only specifying X.matched and indicator.matched if only a matched dataset is available. Furthermore, if one wants to compute the Mahalanobis distance for a full, unmatched dataset, then they should only specify X.matched and indicator.matched.

Value

The Mahalanobis distance of the covariate means between a treatment and control group.

Author(s)

Zach Branson

References

Mahalanobis, P. C. (1936). On the generalized distance in statistics. National Institute of Science of India, 1936.

See Also

See also `lalondeMatches` for details about the Lalonde and matched datasets.

Examples

```
#This loads the classic Lalonde (1986) dataset,
#as well as two matched datasets:
#one from 1:1 propensity score matching,
#and one from cardinality matching, where
#the standardized covariate mean differences are all below 0.1.
data("lalondeMatches")

#obtain the covariates for these datasets
X.lalonde = subset(lalonde, select = -c(treat))
X.matched.ps = subset(lalonde.matched.ps, select = -c(treat,subclass))
X.matched.card = subset(lalonde.matched.card, select = -c(treat,subclass))
#the treatment indicators are
indicator.lalonde = lalonde$treat
indicator.matched.ps = lalonde.matched.ps$treat
indicator.matched.card = lalonde.matched.card$treat

#the Mahalanobis distance for each dataset is:
getMD(X.matched = X.lalonde, indicator.matched = indicator.lalonde)
getMD(X.matched = X.matched.ps, indicator.matched = indicator.matched.ps,
X.full = X.lalonde)
getMD(X.matched = X.matched.card, indicator.matched = indicator.matched.card,
X.full = X.lalonde)
```

`getStandardizedCovMeanDiffs` *Standardized Covariate Mean Differences*

Description

`getStandardizedCovMeanDiffs` computes the standardized covariate mean differences between a treatment and control group, defined as treatment minus control. The standardized covariate mean differences are defined as the covariate mean differences divided by the square-root of the pooled variance between groups.

Usage

```
getStandardizedCovMeanDiffs(X.matched, indicator.matched,
X.full = NULL, indicator.full = NULL)
```

Arguments

X.matched	A covariate matrix (rows correspond to subjects/units; columns correspond to covariates) for the matched dataset.
indicator.matched	A vector of 1s and 0s (e.g., denoting treatment and control) for the matched dataset.
X.full	A covariate matrix (rows correspond to subjects/units; columns correspond to covariates) for the full, unmatched dataset if available.
indicator.full	A vector of 1s and 0s (e.g., denoting treatment and control) for the full, unmatched dataset if available.

Details

The arguments X.full and indicator.full (i.e., the covariate matrix and indicator for the full, unmatched dataset) are only used to correctly define the standardized covariate mean differences. Technically, the covariate mean differences should be standardized by the pooled variance within the full, unmatched dataset, instead of within the matched dataset. If X.full and indicator.full are unspecified, the pooled variance within the matched dataset is used for standardization instead. This distinction rarely leads to large differences in the resulting standardized covariate mean differences, and so researchers should feel comfortable only specifying X.matched and indicator.matched if only a matched dataset is available. Furthermore, if one wants to compute the standardized mean differences for a full, unmatched dataset, then they should only specify X.matched and indicator.matched.

Value

The standardized covariate mean differences between a treatment and control group, defined as treatment minus control.

Author(s)

Zach Branson

See Also

See also `lalondeMatches` for details about the Lalonde and matched datasets.

Examples

```
#This loads the classic Lalonde (1986) dataset,
#as well as two matched datasets:
#one from 1:1 propensity score matching,
#and one from cardinality matching, where
#the standardized covariate mean differences are all below 0.1.
data("lalondeMatches")
```

```
#obtain the covariates for these datasets
X.lalonde = subset(lalonde, select = -c(treat))
X.matched.ps = subset(lalonde.matched.ps, select = -c(treat,subclass))
X.matched.card = subset(lalonde.matched.card, select = -c(treat,subclass))
#the treatment indicators are
indicator.lalonde = lalonde$treat
indicator.matched.ps = lalonde.matched.ps$treat
indicator.matched.card = lalonde.matched.card$treat

#the standardized covariate mean differences
#for these three datasets are:
getStandardizedCovMeanDiffs(
X.matched = X.lalonde,
indicator.matched = indicator.lalonde)
getStandardizedCovMeanDiffs(
X.matched = X.matched.ps,
indicator.matched = indicator.matched.ps,
X.full = X.lalonde,
indicator.full = indicator.lalonde)
getStandardizedCovMeanDiffs(
X.matched = X.matched.card,
indicator.matched = indicator.matched.card,
X.full = X.lalonde,
indicator.full = indicator.lalonde)
```

`lalonde` *Lalonde (1986) Data*

Description

Data from Lalonde (1986).

Usage

```
data(lalondeMatches)
```

Format

The full Lalonde (1986) dataset, containing 614 units (rows) and 9 variables (columns). The columns are:

- `treat`: A binary treatment variable. Equal to 1 if treated in the National Supported Work Demonstration; equal to 0 otherwise.
- `age`: age in years.
- `educ`: years of education.
- `black`: an indicator variable, equal to 1 only if the subject is black.
- `hispan`: an indicator variable, equal to 1 only if the subject is hispanic.

- `married`: an indicator variable, equal to 1 only if the subject is married.
- `nodegree`: an indicator variable, equal to 1 only if the subject does not have a degree.
- `re74`: earnings in 1974.
- `re75`: earnings in 1975.

All of the columns except `treat` are covariates; in these datasets, the outcome variable is not provided.

References

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

Examples

```
data(lalonde)
lalonde
```

`lalonde.matched.card` *A Cardinality Matched Dataset for the Lalonde (1986) Data*

Description

Data from Lalonde (1986) and two matched datasets: One where optimal 1:1 propensity score matching was used, and one where cardinality matching was used, with the balance constraint that all standardized covariate mean differences be below 0.1.

Usage

```
data(lalondeMatches)
```

Format

240 units (rows) and 10 variables (columns). The columns are:

- `treat`: A binary treatment variable. Equal to 1 if treated in the National Supported Work Demonstration; equal to 0 otherwise.
- `age`: age in years.
- `educ`: years of education.
- `black`: an indicator variable, equal to 1 only if the subject is black.
- `hispan`: an indicator variable, equal to 1 only if the subject is hispanic.
- `married`: an indicator variable, equal to 1 only if the subject is married.
- `nodegree`: an indicator variable, equal to 1 only if the subject does not have a degree.
- `re74`: earnings in 1974.
- `re75`: earnings in 1975.
- `subclass`: The subclass denoting the pairs within the matched dataset.

Details

The cardinality matched datset was produced using the designmatch R package.

References

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

Examples

```
data(lalondeMatches)
lalonde.matched.card
```

---

lalonde.matched.ps *A 1:1 Propensity Score Matched Dataset for the Lalonde (1986) Data*

Description

A optimal 1:1 propensity score matched dataset for the Lalonde (1986) dataset.

Usage

```
data(lalondeMatches)
```

Format

370 units (rows) and 10 variables (columns). The columns are:

- treat: A binary treatment variable. Equal to 1 if treated in the National Supported Work Demonstration; equal to 0 otherwise.
- age: age in years.
- educ: years of education.
- black: an indicator variable, equal to 1 only if the subject is black.
- hispan: an indicator variable, equal to 1 only if the subject is hispanic.
- married: an indicator variable, equal to 1 only if the subject is married.
- nodegree: an indicator variable, equal to 1 only if the subject does not have a degree.
- re74: earnings in 1974.
- re75: earnings in 1975.
- subclass: The subclass denoting the pairs within the matched dataset.

Details

The optimal 1:1 propensity score matched dataset was produced using the MatchIt R package. The propensity scores were estimated using logistic regression, where treat was the outcome and the other variables were the covariates (within no interactions included).

References

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

Examples

```
data(lalondeMatches)
lalonde.matched.ps
```

`lalondeMatches` *Lalonde (1986) Data and Two Matched Datasets*

Description

Data from Lalonde (1986) and two matched datasets: One where optimal 1:1 propensity score matching was used, and one where cardinality matching was used, with the balance constraint that all standardized covariate mean differences be below 0.1.

Usage

```
data(lalondeMatches)
```

Format

Three data frames:

- `lalonde`: 614 units (rows) and 9 variables (columns). This is the full Lalonde (1986) dataset.
- `lalonde.matched.ps`: 370 units (rows) and 10 variables (columns). This is the 1:1 propensity score matched dataset.
- `lalonde.matched.card`: 240 units (rows) and 10 variables (columns). This is the cardinality matched dataset.

All three data frames have these 9 columns:

- `treat`: A binary treatment variable. Equal to 1 if treated in the National Supported Work Demonstration; equal to 0 otherwise.
- `age`: age in years.
- `educ`: years of education.
- `black`: an indicator variable, equal to 1 only if the subject is black.
- `hispan`: an indicator variable, equal to 1 only if the subject is hispanic.
- `married`: an indicator variable, equal to 1 only if the subject is married.
- `nodegree`: an indicator variable, equal to 1 only if the subject does not have a degree.
- `re74`: earnings in 1974.
- `re75`: earnings in 1975.

All of the columns except `treat` are covariates; in these datasets, the outcome variable is not provided.

Meanwhile, `lalonde.matched.ps` and `lalonde.matched.card` have one additional column, `subclass`, denoting the pairs for those matched datasets.

Details

The optimal 1:1 propensity score matched dataset was produced using the `MatchIt` `R` package. The propensity scores were estimated using logistic regression, where `treat` was the outcome and the other variables were the covariates (within no interactions included).

The cardinality matched datset was produced using the `designmatch` `R` package.

References

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

Examples

```
data(lalondeMatches)
```

`lovePlot` *Love Plot of Standardized Covariate Mean Differences (along with Permutation Quantiles)*

Description

`lovePlot` produces a Love plot displaying the standardized covariate mean differences (produced by `getStandardizedCovMeanDiffs()`). This function can also produce permutation quantiles for different assignment mechanisms - if a standardized covariate mean difference is outside these quantiles, then that is evidence that the assignment mechanism does not hold. This function supports the following assignment mechanisms:

- Complete randomization ("complete"): Corresponds to random permutations of the indicator across units.
- Block randomization ("blocked"): Corresponds to random permutations of the indicator within blocks of units.
- Constrained-differences randomization ("constrained diffs"): Corresponds to random permutations of the indicator across units, conditional on the standardized covariate mean differences being below some threshold.
- Constrained-Mahalanobis randomization ("constrained md"): Corresponds to random permutations of the indicator across units, conditional on the Mahalanobis being below some threshold.
- Blocked Constrained-differences randomization ("blocked constrained diffs"): Corresponds to random permutations of the indicator within blocks of units, conditional on the standardized covariate mean differences being below some threshold.

- Blocked Constrained-Mahalanobis randomization ("blocked constrained md"): Corresponds to random permutations of the indicator within blocks of units, conditional on the Mahalanobis being below some threshold.

Usage

```
lovePlot(X.matched, indicator.matched,
  permQuantiles = FALSE,
  assignment = "complete",
  subclass = NULL, threshold = NULL,
  alpha = 0.15, perms = 1000,
  X.full = NULL, indicator.full = NULL)
```

Arguments

`X.matched`	A covariate matrix (rows correspond to subjects/units; columns correspond to covariates) for the matched dataset.
`indicator.matched`	A vector of 1s and 0s (e.g., denoting treatment and control) for the matched dataset.
`permQuantiles`	Display permutation quantiles? TRUE or FALSE.
`assignment`	An assignment mechanism that the user wants to visualize. The possible choices are "complete", "blocked", "constrained diffs", "constrained md", "blocked constrained diffs", and "blocked constrained md". See Description for more details on these assignment mechanisms.
`subclass`	A vector denoting the subclass/block for each subject/unit. This must be specified only if one of the blocked assignment mechanisms are used.
`threshold`	The threshold used within the constrained assignment mechanisms; thus, this must be specified only if one of the constrained assignment mechanisms are used. This can be a single number or a vector of numbers (e.g., if one wants to use a different threshold for each covariate when testing constrained-differences randomization).
`alpha`	The alpha-level of the permutation quantiles, where the lower quantile is the alpha/2 quantile and the upper quantile is the 1-alpha/2 quantile. For example, if alpha = 0.15 (the default), then `lovePlot()` will display the 7.5-percent and 92.5-percent quantiles of the standardized covariate mean differences.
`perms`	The number of permutations used to compute the permutation quantiles. A larger number requires more computation time but results in a more consistent p-value.
`X.full`	A covariate matrix (rows correspond to subjects/units; columns correspond to covariates) for the full, unmatched dataset if available.
`indicator.full`	A vector of 1s and 0s (e.g., denoting treatment and control) for the full, unmatched dataset if available.

Details

The arguments X.full and indicator.full (i.e., the covariate matrix and indicator for the full, unmatched dataset) are only used to correctly define the standardized covariate mean differences. Technically, the covariate mean differences should be standardized by the pooled variance within the full, unmatched dataset, instead of within the matched dataset. If X.full and indicator.full are unspecified, the pooled variance within the matched dataset is used for standardization instead. This distinction rarely leads to large differences in the resulting standardized covariate mean differences, and so researchers should feel comfortable only specifying X.matched and indicator.matched if only a matched dataset is available. Furthermore, if one wants to make a Love plot for a full, unmatched dataset, then they should only specify X.matched and indicator.matched.

Value

A Love plot displaying the standardized covariate mean differences. Can also produce permutation quantiles for different assignment mechanisms.

Author(s)

Zach Branson

Examples

```
#This loads the classic Lalonde (1986) dataset,
#as well as two matched datasets:
#one from 1:1 propensity score matching,
#and one from cardinality matching, where
#the standardized covariate mean differences are all below 0.1.
data("lalondeMatches")

#obtain the covariates for these datasets
X.lalonde = subset(lalonde, select = -c(treat))
X.matched.ps = subset(lalonde.matched.ps, select = -c(treat,subclass))
X.matched.card = subset(lalonde.matched.card, select = -c(treat,subclass))
#the treatment indicators are
indicator.lalonde = lalonde$treat
indicator.matched.ps = lalonde.matched.ps$treat
indicator.matched.card = lalonde.matched.card$treat

#the subclass for the matched datasets are
subclass.matched.ps = lalonde.matched.ps$subclass
subclass.matched.card = lalonde.matched.card$subclass

#The following code will display a classic Love plot
#(with a dot for each standardized covariate mean difference).
#Note that, for the full dataset, we only specify X.matched and indicator.matched.
lovePlot(X.matched = X.lalonde, indicator.matched = indicator.lalonde)
lovePlot(X.matched = X.matched.ps, indicator.matched = indicator.matched.ps,
X.full = X.lalonde, indicator.full = indicator.lalonde)
lovePlot(X.matched = X.matched.card, indicator.matched = indicator.matched.card,
X.full = X.lalonde, indicator.full = indicator.lalonde)
```

```
#The following lines of code create Love plots assessing
#whether indicator.data follows different assignment mechanisms by
#plotting the permutation quantiles

#Note that the following examples only use 100 permutations
#to approximate the randomization distribution.
#In practice, we recommend setting perms = 1000 or more;
#in these examples we use perms = 50 to save computation time.

#Assessing complete randomization for the full dataset
#Here we conclude complete randomization doesn't hold
#because the standardized covariate mean differences
#are almost all outside the quantiles.
lovePlot(X.matched = X.lalonde, indicator.matched = indicator.lalonde,
permQuantiles = TRUE,
perms = 50)

#assessing block (paired) randomization for
#the 1:1 propensity score matched dataset
#Many of the standardized covariate mean differences
#are within the permutation quantiles,
#but the race covariates (hispan and black)
#are outside these quantiles.
lovePlot(X.matched = X.matched.ps, indicator.matched = indicator.matched.ps,
X.full = X.lalonde, indicator.full = indicator.lalonde,
   permQuantiles = TRUE,
   perms = 50,
   assignment = "blocked", subclass = subclass.matched.ps)

#assessing block (paired) randomization for
#the cardinality matched dataset
#All of the standardized covariate mean differences
#are within the permutation quantiles
lovePlot(X.matched = X.matched.card, indicator.matched = indicator.matched.card,
X.full = X.lalonde, indicator.full = indicator.lalonde,
   permQuantiles = TRUE,
   perms = 50,
   assignment = "blocked", subclass = subclass.matched.card)

#assessing constrained randomization,
#where the Mahalanobis distance is constrained.
#Note that the Mahalanobis distance is approximately
#a chi^2_K distribution, where K is the number of covariates.
#In the Lalonde data, K = 8.
#Thus, the threshold can be chosen as the quantile of the chi^2_8 distribution.
#This threshold constrains the Mahalanobis distance to be below the 25-percent quantile:
a = qchisq(p = 0.25, df = 8)
#Then, the corresponding Love plot and permutation quantiles are:
lovePlot(X.matched = X.matched.card, indicator.matched = indicator.matched.card,
X.full = X.lalonde, indicator.full = indicator.lalonde,
   permQuantiles = TRUE,
   perms = 50,
   assignment = "constrained md",
```

```
threshold = a)
```

`lovePlotCompare` *Love Plot of Standardized Covariate Mean Differences (along with Permutation Quantiles)*

Description

`lovePlotCompare` produces a Love plot displaying the standardized covariate mean differences (produced by `getStandardizedCovMeanDiffs()`) for two different datasets. The dataset with smaller covariate mean differences is deemed the "more balanced" dataset; this is particularly useful when comparing a full dataset to a matched dataset.

Usage

```
lovePlotCompare(X1, indicator1, X2, indicator2, dataNames = c("Dataset1", "Dataset2"),
X.full = NULL, indicator.full = NULL)
```

Arguments

`X1`	A covariate matrix (rows correspond to subjects/units; columns correspond to covariates) for one dataset.
`indicator1`	A vector of 1s and 0s (e.g., denoting treatment and control) for one dataset.
`X2`	A covariate matrix (rows correspond to subjects/units; columns correspond to covariates) for another dataset.
`indicator2`	A vector of 1s and 0s (e.g., denoting treatment and control) for another dataset.
`dataNames`	A two-length vector denoting the names of the datasets (used in the legend of the plot).
`X.full`	A covariate matrix (rows correspond to subjects/units; columns correspond to covariates) for the full, unmatched dataset if available.
`indicator.full`	A vector of 1s and 0s (e.g., denoting treatment and control) for the full, unmatched dataset if available.

Details

Note that the covariate matrices X1 and X2 have to have the same number of columns and should correspond to the same covariates. However, they do not have to have the same number of rows (i.e., the same number of subjects/units).

Furthermore, the arguments X.full and indicator.full (i.e., the covariate matrix and indicator for the full, unmatched dataset) are only used to correctly define the standardized covariate mean differences. Technically, the covariate mean differences should be standardized by the pooled variance within the full, unmatched dataset, instead of within the matched dataset. If X.full and indicator.full are unspecified, the pooled variance within the matched dataset is used for standardization instead. This distinction rarely leads to large differences in the resulting standardized covariate mean differences, and so researchers should feel comfortable only specifying X1, X2, indicator1, and indicator2 if a full, unmatched dataset is not available.

Value

A Love plot displaying the standardized covariate mean differences for two datasets.

Author(s)

Zach Branson

Examples

```
#This loads the classic Lalonde (1986) dataset,
#as well as two matched datasets:
#one from 1:1 propensity score matching,
#and one from cardinality matching, where
#the standardized covariate mean differences are all below 0.1.
data("lalondeMatches")

#obtain the covariates for these datasets
X.lalonde = subset(lalonde, select = -c(treat))
X.matched.ps = subset(lalonde.matched.ps, select = -c(treat,subclass))
X.matched.card = subset(lalonde.matched.card, select = -c(treat,subclass))
#the treatment indicators are
indicator.lalonde = lalonde$treat
indicator.matched.ps = lalonde.matched.ps$treat
indicator.matched.card = lalonde.matched.card$treat

#The following code will display a classic Love plot
#(with a dot for each standardized covariate mean difference),
#where there are differently-colored dots for each dataset.

#full lalonde dataset vs ps matched dataset
lovePlotCompare(X1 = X.lalonde, indicator1 = indicator.lalonde,
  X2 = X.matched.ps, indicator2 = indicator.matched.ps,
  X.full = X.lalonde, indicator.full = indicator.lalonde,
  dataNames = c("unmatched", "ps matched"))

#ps vs card
lovePlotCompare(X1 = X.matched.ps, indicator1 = indicator.matched.ps,
  X2 = X.matched.card, indicator2 = indicator.matched.card,
  X.full = X.lalonde, indicator.full = indicator.lalonde,
  dataNames = c("ps matched", "card matched"))
```

Index

∗ **Covariate mean difference.**
getCovMeanDiffs, 9
getStandardizedCovMeanDiffs, 11
∗ **Love plot**
lovePlot, 17
lovePlotCompare, 21
∗ **Mahalanobis distance**
asIfRandPlot, 2
asIfRandTest, 5
getMD, 10
lovePlot, 17
∗ **block randomization**
asIfRandPlot, 2
asIfRandTest, 5
lovePlot, 17
∗ **complete randomization**
asIfRandPlot, 2
asIfRandTest, 5
lovePlot, 17
∗ **constrained randomization**
asIfRandPlot, 2
asIfRandTest, 5
lovePlot, 17
∗ **covariate mean difference**
lovePlot, 17
lovePlotCompare, 21
∗ **datasets**
lalonde, 13
lalonde.matched.card, 14
lalonde.matched.ps, 15
lalondeMatches, 16
∗ **lalonde**
lalonde, 13
lalonde.matched.card, 14
lalonde.matched.ps, 15
lalondeMatches, 16
∗ **randomization test**
asIfRandPlot, 2
asIfRandTest, 5
lovePlot, 17

asIfRandPlot, 2
asIfRandTest, 5

getCovMeanDiffs, 9
getMD, 10
getStandardizedCovMeanDiffs, 11

lalonde, 13
lalonde.matched.card, 14
lalonde.matched.ps, 15
lalondeMatches, 16
lovePlot, 17
lovePlotCompare, 21