Package 'DPI'

June 18, 2025

Title The Directed Prediction Index

Version 2025.6

Date 2025-06-17

Maintainer Han-Wu-Shuang Bao <baohws@foxmail.com>

Description The Directed Prediction Index ('DPI') is
a simulation-based and conservative method
for quantifying the relative endogeneity (relative dependence)
of outcome (Y) versus predictor (X) variables
in multiple linear regression models.
By comparing the proportion of variance explained (R-squared)
between the Y-as-outcome model and the X-as-outcome model
while controlling for a sufficient number
of potential confounding variables,
it suggests a more plausible influence direction
from a more exogenous variable (X) to a more endogenous variable (Y).
Methodological details are provided at
<https://psychbruce.github.io/DPI/>.

License GPL-3

Encoding UTF-8

URL https://psychbruce.github.io/DPI/

BugReports https://github.com/psychbruce/DPI/issues

Depends R (>= 4.0.0)

Imports glue, crayon, cli, ggplot2, qgraph

Suggests bruceR

RoxygenNote 7.3.2

NeedsCompilation no

Author Han-Wu-Shuang Bao [aut, cre] (ORCID:
<https://orcid.org/0000-0003-3043-710X>)

Repository CRAN

Date/Publication 2025-06-18 08:50:05 UTC

Contents

`cor_network` *Correlation network plot.*

Description

Correlation network plot.

Usage

```
cor_network(
  data,
  index = c("cor", "pcor"),
  show.value = TRUE,
  show.insig = FALSE,
  show.cutoff = FALSE,
  faded = FALSE,
  text.size = 1.2,
  node.group = NULL,
  node.color = NULL,
  edge.color.pos = "#0571B0",
  edge.color.neg = "#CA0020",
  edge.color.insig = "#EEEEEEEE",
  title = NULL,
  file = NULL,
  width = 6,
  height = 4,
  dpi = 500,
  ...
)
```

Arguments

`data`	Data.

index	Type of graph: "cor" (raw correlation network) or "pcor" (partial correlation network). Defaults to "cor".
show.value	Show correlation coefficients and their significance on edges. Defaults to TRUE.
show.insig	Show edges with insignificant correlations ($p > 0.05$). Defaults to FALSE. To change significance level, please set the alpha parameter (defaults to alpha=0.05).
show.cutoff	Show cut-off values of correlations. Defaults to FALSE.
faded	Transparency of edges according to the effect size of correlation. Defaults to FALSE.
text.size	Scalar on the font size of variable labels. Defaults to 1.2.
node.group	A list that indicates which nodes belong together, with each element of list as a vector of integers identifying the column numbers of variables that belong together.
node.color	A vector with a color for each element in node.group, or a color for each node.
edge.color.pos	Color for (significant) positive correlations. Defaults to "#0571B0" (blue in ColorBrewer's RdBu palette).
edge.color.neg	Color for (significant) negative correlations. Defaults to "#CA0020" (red in ColorBrewer's RdBu palette).
edge.color.insig	Color for insignificant correlations. Defaults to "#EEEEEEEE" (transparent grey).
title	Plot title.
file	File name of saved plot (".png" or ".pdf").
width, height	Width and height (in inches) of saved plot. Defaults to 6 and 4.
dpi	Dots per inch (figure resolution). Defaults to 500.
...	Other parameters passed to qgraph.

Value

Invisibly return a qgraph object.

Examples

```
cor_network(airquality)
cor_network(airquality, show.insig=TRUE)
```

---

data_random *Generate random data.*

Description

Generate random data.

Usage

```
data_random(k, n, seed = NULL)
```

Arguments

k	Number of variables.
n	Number of observations (cases).
seed	Random seed for replicable results. Defaults to NULL.

Value

Return a data.frame of random data.

Examples

```
d = data_random(k=5, n=100, seed=1)
cor_network(d)
```

DPI *The Directed Prediction Index (DPI).*

Description

The Directed Prediction Index (DPI) is a simulation-based and conservative method for quantifying the *relative endogeneity* (relative dependence) of outcome (Y) vs. predictor (X) variables in multiple linear regression models. By comparing the proportion of variance explained (R-squared) between the Y-as-outcome model and the X-as-outcome model while controlling for a sufficient number of potential confounding variables, it suggests a more plausible influence direction from a more exogenous variable (X) to a more endogenous variable (Y). Methodological details are provided at https://psychbruce.github.io/DPI/.

Usage

```
DPI(
  model,
  y,
  x,
  data = NULL,
  k.cov = 1,
  n.sim = 1000,
  seed = NULL,
  progress,
  file = NULL,
  width = 6,
  height = 4,
  dpi = 500
)
```

Arguments

model	Model object (lm).
y	Dependent (outcome) variable.
x	Independent (predictor) variable.
data	[Optional] Defaults to NULL. If data is specified, then model will be ignored and a linear model lm({y} ~ {x} + .) will be fitted inside. This is helpful for exploring all variables in a dataset.
k.cov	Number of random covariates (simulating potential omitted variables) added to each simulation sample. • Defaults to 1. Please also test different k.cov values as robustness checks (see DPI_curve). • If k.cov > 0, the raw data (without bootstrapping) are used, with k.cov random variables appended, for simulation. • If k.cov = 0 (not suggested), bootstrap samples (resampling with replacement) are used for simulation.
n.sim	Number of simulation samples. Defaults to 1000.
seed	Random seed for replicable results. Defaults to NULL.
progress	Show progress bar. Defaults to FALSE (if n.sim < 5000).
file	File name of saved plot (".png" or ".pdf").
width, height	Width and height (in inches) of saved plot. Defaults to 6 and 4.
dpi	Dots per inch (figure resolution). Defaults to 500.

Value

Return a data.frame of simulation results:

- DPI
  - t.beta.xy^2 * (R2.Y - R2.X)
- t.beta.xy
  - *t* value for coefficient of X predicting Y (always equal to *t* value for coefficient of Y predicting X) when controlling for all other covariates
- df.beta.xy
  - residual degree of freedom (df) of t.beta.xy
- r.partial.xy
  - partial correlation (always with the same *t* value as t.beta.xy) between X and Y when controlling for all other covariates
- delta.R2
  - R2.Y - R2.X
- R2.Y
  - R^2 of regression model predicting Y using X and all other covariates
- R2.X
  - R^2 of regression model predicting X using Y and all other covariates

Examples

```
model = lm(Ozone ~ ., data=airquality)
DPI(model, y="Ozone", x="Solar.R", seed=1)
DPI(data=airquality, y="Ozone", x="Solar.R", k.cov=10, seed=1)
```

`DPI_curve` *The DPI curve analysis.*

Description

The DPI curve analysis.

Usage

```
DPI_curve(
  model,
  y,
  x,
  data = NULL,
  k.covs = 1:10,
  n.sim = 1000,
  seed = NULL,
  file = NULL,
  width = 6,
  height = 4,
  dpi = 500
)
```

Arguments

`model`	Model object (`lm`).
`y`	Dependent (outcome) variable.
`x`	Independent (predictor) variable.
`data`	[Optional] Defaults to `NULL`. If `data` is specified, then `model` will be ignored and a linear model `lm({y} ~ {x} + .)` will be fitted inside. This is helpful for exploring all variables in a dataset.
`k.covs`	An integer vector of number of random covariates (simulating potential omitted variables) added to each simulation sample. Defaults to `1:10` (producing DPI results for `k.cov=1~10`). For details, see `DPI`.
`n.sim`	Number of simulation samples. Defaults to `1000`.
`seed`	Random seed for replicable results. Defaults to `NULL`.
`file`	File name of saved plot (`".png"` or `".pdf"`).
`width, height`	Width and height (in inches) of saved plot. Defaults to `6` and `4`.
`dpi`	Dots per inch (figure resolution). Defaults to `500`.

Value

Return a data.frame of DPI curve results.

Examples

```
model = lm(Ozone ~ ., data=airquality)
DPIs = DPI_curve(model, y="Ozone", x="Solar.R", seed=1)
plot(DPIs)  # ggplot object
```

`plot.dpi` *[S3 method] Plot DPI results.*

Description

[S3 method] Plot DPI results.

Usage

```
## S3 method for class 'dpi'
plot(x, file = NULL, width = 6, height = 4, dpi = 500, ...)
```

Arguments

`x`	A data.frame (of new class `dpi`) returned from `DPI`.
`file`	File name of saved plot (`".png"` or `".pdf"`).
`width, height`	Width and height (in inches) of saved plot. Defaults to `6` and `4`.
`dpi`	Dots per inch (figure resolution). Defaults to `500`.
`...`	Other arguments (currently not used).

Value

Return a `ggplot` object.

plot.dpi.curve *[S3 method] Plot DPI curve analysis results.*

Description

[S3 method] Plot DPI curve analysis results.

Usage

```
## S3 method for class 'dpi.curve'
plot(x, file = NULL, width = 6, height = 4, dpi = 500, ...)
```

Arguments

x	A data.frame (of new class dpi.curve) returned from DPI_curve.
file	File name of saved plot (".png" or ".pdf").
width, height	Width and height (in inches) of saved plot. Defaults to 6 and 4.
dpi	Dots per inch (figure resolution). Defaults to 500.
...	Other arguments (currently not used).

Value

Return a ggplot object.

print.dpi *[S3 method] Print DPI summary and plot.*

Description

[S3 method] Print DPI summary and plot.

Usage

```
## S3 method for class 'dpi'
print(x, digits = 3, ...)
```

Arguments

x	A data.frame (of new class dpi) returned from DPI.
digits	Number of decimal places. Defaults to 3.
...	Other arguments (currently not used).

Value

No return value.

print.summary.dpi *[S3 method] Print DPI summary.*

Description

[S3 method] Print DPI summary.

Usage

```
## S3 method for class 'summary.dpi'
print(x, digits = 3, ...)
```

Arguments

x	A data.frame (of new class dpi) returned from DPI.
digits	Number of decimal places. Defaults to 3.
...	Other arguments (currently not used).

Value

No return value.

summary.dpi *[S3 method] Summarize DPI results.*

Description

[S3 method] Summarize DPI results.

Usage

```
## S3 method for class 'dpi'
summary(object, ...)
```

Arguments

object	A data.frame (of new class dpi) returned from DPI.
...	Other arguments (currently not used).

Value

Return a list with new class summary.dpi of summarized results and raw DPI data.frame.

Index