Package ‘ei’

October 13, 2022

Version 1.3-3
Date 2016-11-04
Title Ecological Inference
Author Gary King <king@harvard.edu>, Molly Roberts <molly.e.roberts@gmail.com>
Maintainer James Honaker <zelig.zee@gmail.com>
Depends R (>= 2.5.0), eiPack
Imports mvtnorm, msm, tmvtnorm, ellipse, plotrix, MASS, ucminf, cubature, mnormt, foreach, sp
Suggests rgl
License GPL (>= 2)
URL http://gking.harvard.edu/eiR,
http://gking.harvard.edu/eicamera/kinroot.html
NeedsCompilation no
Repository CRAN
Date/Publication 2016-11-05 00:49:10

R topics documented:

  bounds1 .............................................. 2
  census1910 .......................................... 3
  ei ......................................................... 3
  eisim .................................................... 5
  eiread .................................................... 6
  eiRxCsample ......................................... 7
  fultongen ............................................. 8
  lavoteall .............................................. 8
  matproii ............................................... 9
  nj ....................................................... 10
  plot.ei ................................................. 10
bounds1

Computes Analytical Bounds from Accounting Identity

Description

Returns analytical bounds from accounting identity on unknown table relationships beta_b, beta_w, from known, observed, table marginals, x, t (and sample size n).

Usage

bounds1(x, t, n)

Arguments

x vector of characteristics, e.g. percentage of blacks in each district
t vector of characteristics, e.g. percentage of people that voted in each district
n size of each observation, e.g. number of voters in each district

Author(s)

Gary King «email: king@harvard.edu» and Molly Roberts «email: molly.e.roberts@gmail.com»

References


Examples

data(census1910)
output<-bounds1(x=census1910$x, t=census1910$t, n=census1910$n)
Black Literacy in 1910

Description
A dataset of aggregate literacy rates (t) and fraction of the population that is black (x), from the 1910 US Census. Each observation represents one county.

Usage
census1910

Format
A data frame containing 1030 observations.

Source

References


ei

Ecological Inference Estimation

Description
ei is the main command in the package EI. It gives observation-level estimates (and various related statistics) of $\beta^b_i$ and $\beta^w_i$ given variables $T_i$ and $X_i$ ($i = 1, ..., n$) in this accounting identity: $T_i = \beta^b_i * X_i + \beta^w_i * (1 - X_i)$. Results are stored in an ei object, that can be read with summary() or eiread() and graphed in plot().

Usage
ei(formula, total = NULL, Zb = 1, Zw = 1, id = NA, data =NA, erho = 0.5, esigma = 0.5, ebeta = 0.5, ealphab = NA, ealphaw = NA, truth = NA, simulate = TRUE, covariate = NULL, lambda1 = 4, lambda2 = 2, covariate.prior.list = NULL, tune.list = NULL, start.list = NULL, sample = 1000, thin = 1, burnin = 1000, verbose = 0, ret.beta = "r", ret.mcmc = TRUE, usrfun = NULL)
Arguments

formula A formula of the form \( t \times x \) in the \( 2 \times 2 \) case and \( \text{cbind} \left( \text{col}1, \text{col}2, \ldots \right) \) \( \text{cbind} \left( \text{row}1, \text{row}2, \ldots \right) \) in the \( R \times C \) case.

total ‘total’ is the name of the variable in the dataset that contains the number of individuals in each unit

\( Z_b \) \( p \times k^b \) matrix of covariates or the name of covariates in the dataset

\( Z_w \) \( p \times k^w \) matrix of covariates or the name of covariates in the dataset

id ‘id’ is the nae of the variable in the dataset that identifies the precinct. Used for ‘movie’ and ‘movieD’ plot functions.

data data frame that contains the variables that correspond to formula. If using covariates and data is specified, data should also contain \( Z_b \) and \( Z_w \).

\( \text{erho} \) The standard deviation of the normal prior on \( \phi_5 \) for the correlation. Default = 0.5.

\( \text{esigma} \) The standard deviation of an underlying normal distribution, from which a half normal is constructed as a prior for both \( \bar{\sigma}_b \) and \( \bar{\sigma}_w \). Default = 0.5.

\( \text{ebeta} \) Standard deviation of the "flat normal" prior on \( \bar{B}_b \) and \( \bar{B}_w \). The flat normal prior is uniform within the unit square and dropping outside the square according to the normal distribution. Set to zero for no prior. Setting to positive values probabilistically keeps the estimated mode within the unit square. Default = 0.5.

\( \text{ealphab} \) \( \text{cols}(Z_b) \times 2 \) matrix of means (in the first column) and standard deviations (in the second) of an independent normal prior distribution on elements of \( \alpha^b \). If you specify \( Z_b \), you should probably specify a prior, at least with mean zero and some variance (default is no prior). (See Equation 9.2, page 170, to interpret \( \alpha^b \).

\( \text{ealphaw} \) \( \text{cols}(Z_w) \times 2 \) matrix of means (in the first column) and standard deviations (in the second) of an independent normal prior distribution on elements of \( \alpha^w \). If you specify \( Z_w \), you should probably specify a prior, at least with mean zero and some variance (default is no prior). (See Equation 9.2, page 170, to interpret \( \alpha^w \).

truth A length(t) x 2 matrix of the true values of the quantities of interest.

simulate default = TRUE: see documentation in eiPack for options for RxC ei.

covariate see documentation in eiPack for options for RxC ei.

lambda1 default = 4: see documentation in eiPack for options for RxC ei.

lambda2 default = 2: see documentation in eiPack for options for RxC ei.

covariate.prior.list see documentation in eiPack for options for RxC ei.

tune.list see documentation in eiPack for options for RxC ei.

start.list see documentation in eiPack for options for RxC ei.

sample default = 1000

thin default = 1

burnin default = 1000
ei.sim

Simulate EI Solution via Importance Sampling

description
Simulate EI solution via importance sampling

Usage
ei.sim(ei.object)

Arguments
ei.object ei object

Author(s)
Gary King «email: king@harvard.edu» and Molly Roberts «email: molly.e.roberts@gmail.com»

References
**eiread**

*Quantities of Interest from Ecological Inference Estimation*

**Description**

*eiread* is the command that pulls quantities of interest from the *ei* object. The command returns a list of quantities of interest requested by the user.

**Usage**

```r
eiread(ei.object, ...)
```

**Arguments**

- `ei.object`: An *ei* object from the function *ei*.
- `...`: A list of quantities of interest for *eiread*() to return. See values below.

**Value**

- `betab`: $p \times 1$ point estimate of $\beta_i^b$ based on its mean posterior. See section 8.2
- `betaw`: $p \times 1$ point estimate of $\beta_i^w$ based on its mean posterior. See section 8.2
- `sbetab`: $p \times 1$ standard error for the estimate of $\beta_i^b$, based on the standard deviation of its posterior. See section 8.2
- `sbetaw`: $p \times 1$ standard error for the estimate of $\beta_i^w$, based on the standard deviation of its posterior. See section 8.2
- `phi`: Maximum posterior estimates of the CML
- `psisims`: Matrix of random simulations of $\psi$. See section 8.2
- `bounds`: $p \times 4$: bounds on $\beta_i^b$ and $\beta_i^w$, lowerB ~ upperB ~ lowerW ~ upperW. See Chapter 5.
- `abounds`: $2 \times 2$: aggregate bounds rows:lower, upper; columns: betab, betaw. See Chapter 5.
- `aggs`: Simulations of district-level quantities of interest $\hat{B}^b$ and $\hat{B}^w$. See Section 8.3.
- `maggs`: Point estimate of 2 district-level parameters, $\hat{B}^b$ and $\hat{B}^w$ based on the mean of aggs. See Section 8.3.
- `VCaggs`: Variance matrix of 2 district-level parameters, $\hat{B}^b$ and $\hat{B}^w$. See Section 8.3.
- `CI80b`: $p \times 2$: lower-upper 80% confidence intervals for $\beta_i^b$. See section 8.2.
- `CI80w`: $p \times 2$: lower-upper 80% confidence intervals for $\beta_i^w$. See section 8.2.
- `eaggbias`: Regressions of estimated $\hat{\beta}_i^b$ and $\hat{\beta}_i^w$ on a constant term and $X_i$.
- `goodman`: Goodman’s Regression. See Section 3.1

**Author(s)**

Gary King «email: king@harvard.edu» and Molly Roberts «email: molly.e.roberts@gmail.com»
References


Examples

```r
data(sample)
formula = t ~ x
dbuf <- ei(formula=formula, total="n", data=sample)
eiread(dbuf, "phi")
eiread(dbuf, "betab", "betaw")
```

Description

A description for this dataset

Usage

```r
eiRxCsample
```

Format

A data frame containing 93 observations.

Source

Source

References

**fultongen**  
*Voter Transitions*

**Description**
Aggregated data from 289 precincts in Fulton County, Georgia. The variable \( t \) represents the fraction voting in 1994 and \( x \) the fraction in 1992. Beta_b is then the fraction who vote in both elections, and Beta_w the fraction of nonvoters in 1992 who vote in the midterm election of 1994.

**Usage**
fultongen

**Format**
A data frame containing 289 observations.

**Source**

**References**

---

**lavoteall**  
*Turnout by Race in Louisiana*

**Description**
The fraction of blacks registered voters (x) and fraction of voter turnout (t) in each Louisiana precinct, along with the true fraction of black turnout (tb) and non-black turnout (tw).

**Usage**
lavoteall

**Format**
A data frame containing 3262 observations.
Source


References


matproii  Voter Registration by Race in Southern States

Description

Aggregate voter registration and fraction black, in counties in Florida, Louisiana, North Carolina and South Carolina

Usage

matproii

Format

A data frame containing 268 observations.

Source


References

nj  

Nonminority Turnout in New Jersey

Description

A description for this dataset

Usage

nj

Format

A data frame containing 493 observations.

Source


References


plot.ei  

Plotting Ecological Inference Estimates

Description

'plot' method for the class 'ei'.

Usage

## S3 method for class 'ei'
plot(x, ...)

Arguments

x  An ei object from the function ei.

...  A list of options to return in graphs. See values below.
plot.ei

Details

Returns any of a set of possible graphical objects, mirroring those in the examples in King (1997). Graphical option lci is a logical value specifying the use of the Law of Conservation of Ink, where the implicit information in the data is represented through color gradients, i.e. the color of the line is a function of the length of the tomography line. This can be passed as an argument and is used for “tomogD” and “tomog” plots.

Value

tomogD  Tomography plot with the data only. See Figure 5.1, page 81.
tomog  Tomography plot with ML contours. See Figure 10.2, page 204.
tomogCI  Tomography plot with 80\% confidence intervals. Confidence intervals appear on the screen in red with the remainder of the tomography line in yellow. The confidence interval portion is also printed thicker than the rest of the line. See Figure 9.5, page 179.
tomogCI95  Tomography plot with 95\% confidence intervals. Confidence intervals appear on the screen in red with the remainder of the tomography line in yellow. The confidence interval portion is also printed thicker than the rest of the line. See Figure 9.5, page 179.
tomogE  Tomography plot with estimated mean posterior $\beta_b^i$ and $\beta_w^i$ points.
tomogP  Tomography plot with mean posterior contours.
betab  Density estimate (i.e., a smooth version of a histogram) of point estimates of $\beta_b^i$'s with whiskers.
betaw  Density estimate (i.e., a smooth version of a histogram) of point estimates of $\beta_w^i$'s with whiskers.
xt  Basic $X_i$ by $T_i$ scatterplot.
xtc  Basic $X_i$ by $T_i$ scatterplot with circles sized proportional to $N_i$.
xtfit  $X_i$ by $T_i$ plot with estimated $E(T_i|X_i)$ and conditional 80\% confidence intervals. See Figure 10.3, page 206.
xtfitg  xtfit with Goodman’s regression line superimposed.
estsim  All the simulated $\beta_b^i$’s by all the simulated $\beta_w^i$’s. The simulations should take roughly the same shape of the mean posterior contours, except for those sampled from outlier tomography lines.
boundXb  $X_i$ by the bounds on $\beta_b^i$ (each precinct appears as one vertical line), see the lines in the left graph in Figure 13.2, page 238.
boundXw  $X_i$ by the bounds on $\beta_w^i$ (each precinct appears as one vertical line), see the lines in the right graph in Figure 13.2, page 238.
truth  Compares truth to estimates at the district and precinct-level. Requires truth in the ei object. See Figures 10.4 (page 208) and 10.5 (page 210).
movieD  For each observation, one tomography plot appears with the line for the particular observation darkened. After the graph for each observation appears, the user can choose to view the next observation (hit return), jump to a specific observation number (type in the number and hit return), or stop (hit ”s” and return).
For each observation, one page of graphics appears with the posterior distribution of $\beta^b_i$ and $\beta^w_i$ and a plot of the simulated values of $\beta^b_i$ and $\beta^w_i$ from the tomography line. The user can choose to view the next observation (hit return), jump to a specific observation number (type in the number and hit return), or stop (hit “s” and return).

Author(s)

Gary King «email: king@harvard.edu» and Molly Roberts «email: molly.e.roberts@gmail.com»

References


Examples

data(sample)
formula = t ~ x
dbuf <- ei(formula=formula, total="n",data=sample)
plot(dbuf, "tomog")
plot(dbuf, "tomog", "betab", "betaw", "xfit")

RxCdata  Sample Dataset

Description

A description for this dataset

Usage

RxCdata

Format

A data frame containing 60 observations.

Source

Source

References

Sample Data for Black Votes

Description
A description for this dataset

Usage
sample

Format
A vector containing 141 observations.

Source
Source

References

summary.ei

Summarize Ecological Inference Estimates

Description
‘summary’ method for the class ‘ei’.

Usage
## S3 method for class 'ei'
summary(object, ...)

Arguments

object      An ei object from the function ei.
...         A list of options to return in graphs. See values below.

Author(s)
Gary King «email: king@harvard.edu» and Molly Roberts «email: molly.e.roberts@gmail.com»
References


Examples

```r
data(sample)
formula = t ~ x
ddbuf <- ei(formula=formula, total="n",data=sample)
print(summary(ddbuf))
```

tomogRxC

Plotting Ecological Inference Estimates with eiRxC information

Description

A tomography plot for an estimated Ecological Inference model in RxC data.

Usage

```
tomogRxC(formula, data, total=NULL, refine=100)
```

Arguments

- `formula` A formula of the form `cbind(col1, col2,...)~cbind(row1,row2,...)`
- `data` data that contains the data that corresponds to the formula
- `total` 'total' is the name of the variable in the dataset that contains the number of individuals in each unit
- `refine` specifies the amount of refinement for the image. Higher numbers mean better resolution.

Author(s)

Gary King «email: king@harvard.edu» and Molly Roberts «email: molly.e.roberts@gmail.com»

References


Examples

```r
data(RxCdata)
formula = cbind(turnout, noturnout) ~ cbind(white, black,hisp)
tomogRxC(formula, data=RxCdata)
```
tomogRx3d

Plotting 2x3 Ecological Inference Estimates in 3 dimensions

Description

A tomography plot in 3 dimensions for RxC Ecological Inference data and an estimated Ecological Inference model in RxC data.

Usage

tomogRx3d(formula, data, total=NULL, lci=TRUE, estimates=FALSE, ci=FALSE, level=.95, seed=1234, color=hcl(h=30,c=100,l=60), transparency=.75, light=FALSE, rotate=TRUE)

Arguments

formula A formula of the form cbind(col1, col2,...)~cbind(row1,row2,...)
data data that contains the data that corresponds to the formula
total 'total' is the name of the variable in the dataset that contains the number of individuals in each unit
lci logical value specifying the use of the Law of Conservation of Ink, where the implicit information in the data is represented through color gradients, i.e. the color of the plane is a function of the area of the tomography plane.
estimates logical value specifying whether the point estimates of β’s are included for each observation on the tomography plot.
ci logical value specifying whether the estimated confidence ellipse is included on the tomography plot.
level numeric value from 0 to 1 specifying the significance level of the confidence ellipse; eg. .95 refers to 95% confidence ellipse.
seed seed value for model estimation.
color color of tomography planes if lci=F.
transparency numeric value from 0 to 1 specifying transparency of tomography planes; 0 is entirely transparent.
light logical value specifying whether lights should be included in the rgl interface. The inclusion of lights will create shadows in the plot that may distort colors.
rotate logical value specifying whether the plot will rotate for 20 seconds.

Details

Requires rgl package and rgl viewer.

Author(s)

Gary King «email: king@harvard.edu»; Molly Roberts «email: molly.e.roberts@gmail.com»; Soledad Prillaman «email: soledadartiz@fas.harvard.edu».
References


Examples

data(RxCdata)
formula <- cbind(turnout, noturnout) ~ cbind(white, black, hisp)
tomogRxC3d(formula, RxCdata, total=NULL, lci=TRUE, estimates=TRUE, ci=TRUE, transparency=.5, light=FALSE, rotate=FALSE)
Index

* datasets
  census1910, 3
eiRxCsample, 7
fultongen, 8
lavoteall, 8
matproii, 9
nj, 10
RxCdata, 12
sample, 13

bounds1, 2
census1910, 3

EI (ei), 3
ei, 3
ei.sim, 5
eiread, 6
eiRxCsample, 7

fultongen, 8
lavoteall, 8
matproii, 9
nj, 10
plot.ei, 10
RxCdata, 12
sample, 13
summary.ei, 13

tomogRxC, 14
tomogRxC3d, 15