Package ‘eco’

July 18, 2009

Version 3.1-4

Date 2009-07-12

Title R Package for Ecological Inference in 2x2 Tables

Author Kosuke Imai <kimai@princeton.edu>, Ying Lu <yl46@nyu.edu>, Aaron Strauss <abstraus@Princeton.EDU>.

Maintainer Ying Lu <yl46@nyu.edu>

Depends R (>= 2.0), MASS

Description eco is a publicly available R package that implements the Bayesian and likelihood methods proposed in Imai, Lu, and Strauss (2008) for ecological inference in $2 \{ }times 2$ tables as well as the method of bounds introduced by Duncan and Davis (1953). The package fits both parametric and nonparametric models using either the Expectation-Maximization algorithms (for likelihood models) or the Markov chain Monte Carlo algorithms (for Bayesian models). For all models, the individual-level data can be directly incorporated into the estimation whenever such data are available. Along with in-sample and out-of-sample predictions, the package also provides a functionality which allows one to quantify the effect of data aggregation on parameter estimation and hypothesis testing under the parametric likelihood models.

LazyLoad yes

LazyData yes

URL http://imai.princeton.edu/software/eco.html

Repository CRAN

Date/Publication 2009-07-18 15:41:13

R topics documented:

`census` *Black Illiteracy Rates in 1910 US Census*

Description

This data set contains the proportion of the residents who are black, the proportion of those who can read, the total population as well as the actual black literacy rate and white literacy rate for 1040 counties in the US. The dataset was originally analyzed by Robinson (1950) at the state level. King (1997) recoded the 1910 census at county level. The data set only includes those who are older than 10 years of age.

Usage

```
data(census)
```

Format

A data frame containing 5 variables and 1040 observations

X	numeric	the proportion of Black residents in each county
Y	numeric	the overall literacy rates in each county
N	numeric	the total number of residents in each county
W1	numeric	the actual Black literacy rate
W2	numeric	the actual White literacy rate

References

Robinson, W.S. (1950). "Ecological Correlations and the Behavior of Individuals." *American Sociological Review*, vol. 15, pp.351-357.

King, G. (1997). "A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data". Princeton University Press, Princeton, NJ.

`eco` *Fitting the Parametric Bayesian Model of Ecological Inference in 2x2 Tables*

Description

`eco` is used to fit the parametric Bayesian model (based on a Normal/Inverse-Wishart prior) for ecological inference in 2×2 tables via Markov chain Monte Carlo. It gives the in-sample predictions as well as the estimates of the model parameters. The model and algorithm are described in Imai, Lu and Strauss (2008, Forthcoming).

Usage

```
eco(formula, data = parent.frame(), N = NULL, supplement = NULL,
    context = FALSE, mu0 = 0, tau0 = 2, nu0 = 4, S0 = 10,
    mu.start = 0, Sigma.start = 10, parameter = TRUE,
    grid = FALSE, n.draws = 5000, burnin = 0, thin = 0,
    verbose = FALSE)
```

Arguments

`formula` A symbolic description of the model to be fit, specifying the column and row margins of 2×2 ecological tables. `Y ~ X` specifies `Y` as the column margin (e.g., turnout) and `X` as the row margin (e.g., percent African-American). Details and specific examples are given below.

`data` An optional data frame in which to interpret the variables in `formula`. The default is the environment in which `eco` is called.

`N` An optional variable representing the size of the unit; e.g., the total number of voters. `N` needs to be a vector of same length as `Y` and `X` or a scalar.

`supplement` An optional matrix of supplemental data. The matrix has two columns, which contain additional individual-level data such as survey data for W_1 and W_2, respectively. If `NULL`, no additional individual-level data are included in the model. The default is `NULL`.

`context` Logical. If `TRUE`, the contextual effect is also modeled, that is to assume the row margin X and the unknown W_1 and W_2 are correlated. See Imai, Lu and Strauss (2008, Forthcoming) for details. The default is `FALSE`.

`mu0`	A scalar or a numeric vector that specifies the prior mean for the mean parameter μ for (W_1, W_2) (or for (W_1, W_2, X) if `context=TRUE`). When the input of `mu0` is a scalar, its value will be repeated to yield a vector of the length of μ, otherwise, it needs to be a vector of same length as μ. When `context=TRUE`, the length of μ is 3, otherwise it is 2. The default is `0`.
`tau0`	A positive integer representing the scale parameter of the Normal-Inverse Wishart prior for the mean and variance parameter (μ, Σ). The default is `2`.
`nu0`	A positive integer representing the prior degrees of freedom of the Normal-Inverse Wishart prior for the mean and variance parameter (μ, Σ). The default is `4`.
`S0`	A positive scalar or a positive definite matrix that specifies the prior scale matrix of the Normal-Inverse Wishart prior for the mean and variance parameter (μ, Σ) . If it is a scalar, then the prior scale matrix will be a diagonal matrix with the same dimensions as Σ and the diagonal elements all take value of `S0`, otherwise `S0` needs to have same dimensions as Σ. When `context=TRUE`, Σ is a 3×3 matrix, otherwise, it is 2×2. The default is `10`.
`mu.start`	A scalar or a numeric vector that specifies the starting values of the mean parameter μ. If it is a scalar, then its value will be repeated to yield a vector of the length of μ, otherwise, it needs to be a vector of same length as μ. When `context=FALSE`, the length of μ is 2, otherwise it is 3. The default is `0`.
`Sigma.start`	A scalar or a positive definite matrix that specified the starting value of the variance matrix Σ. If it is a scalar, then the prior scale matrix will be a diagonal matrix with the same dimensions as Σ and the diagonal elements all take value of `S0`, otherwise `S0` needs to have same dimensions as Σ. When `context=TRUE`, Σ is a 3×3 matrix, otherwise, it is 2×2. The default is `10`.
`parameter`	Logical. If `TRUE`, the Gibbs draws of the population parameters, μ and Σ, are returned in addition to the in-sample predictions of the missing internal cells, W. The default is `TRUE`.
`grid`	Logical. If `TRUE`, the grid method is used to sample W in the Gibbs sampler. If `FALSE`, the Metropolis algorithm is used where candidate draws are sampled from the uniform distribution on the tomography line for each unit. Note that the grid method is significantly slower than the Metropolis algorithm. The default is `FALSE`.
`n.draws`	A positive integer. The number of MCMC draws. The default is `5000`.
`burnin`	A positive integer. The burnin interval for the Markov chain; i.e. the number of initial draws that should not be stored. The default is `0`.
`thin`	A positive integer. The thinning interval for the Markov chain; i.e. the number of Gibbs draws between the recorded values that are skipped. The default is `0`.
`verbose`	Logical. If `TRUE`, the progress of the Gibbs sampler is printed to the screen. The default is `FALSE`.

Details

An example of 2×2 ecological table for racial voting is given below:

black voters white voters

vote	W_{1i}	W_{2i}	Y_i
not vote	$1 - W_{1i}$	$1 - W_{2i}$	$1 - Y_i$
	X_i	$1 - X_i$	

where Y_i and X_i represent the observed margins, and W_1 and W_2 are unknown variables. In this exmaple, Y_i is the turnout rate in the ith precint, X_i is the proprotion of African American in the ith precinct. The unknowns W_{1i} an dW_{2i} are the black and white turnout, respectively. All variables are proportions and hence bounded between 0 and 1. For each i, the following deterministic relationship holds, $Y_i = X_i W_{1i} + (1 - X_i) W_{2i}$.

Value

An object of class `eco` containing the following elements:

`call`	The matched call.
`X`	The row margin, X.
`Y`	The column margin, Y.
`N`	The size of each table, N.
`burnin`	The number of initial burnin draws.
`thin`	The thinning interval.
`nu0`	The prior degrees of freedom.
`tau0`	The prior scale parameter.
`mu0`	The prior mean.
`S0`	The prior scale matrix.
`W`	A three dimensional array storing the posterior in-sample predictions of W. The first dimension indexes the Monte Carlo draws, the second dimension indexes the columns of the table, and the third dimension represents the observations.
`Wmin`	A numeric matrix storing the lower bounds of W.
`Wmax`	A numeric matrix storing the upper bounds of W.
`mu`	The posterior draws of the population mean parameter, μ.
`Sigma`	The posterior draws of the population variance matrix, Σ.

Author(s)

Kosuke Imai, Department of Politics, Princeton University, ⟨kimai@Princeton.Edu⟩, http://imai.princeton.edu; Ying Lu, Department of Humanities and Social Sciences in the Professions, Steinhardt School of Culture, Education and Human Development, New York University, ⟨yl46@Nyu.Edu⟩

References

Imai, Kosuke, Ying Lu and Aaron Strauss. (Forthcoming). "eco: R Package for Ecological Inference in 2x2 Tables" Journal of Statistical Software, available at http://imai.princeton.edu/research/eco.html

Imai, Kosuke, Ying Lu and Aaron Strauss. (2008). "Bayesian and Likelihood Inference for 2 x 2 Ecological Tables: An Incomplete Data Approach" Political Analysis, Vol. 16, No. 1 (Winter), pp. 41-69. available at http://imai.princeton.edu/research/eiall.html

See Also

ecoML, ecoNP, predict.eco, summary.eco

Examples

```
## load the registration data
data(reg)

## NOTE: convergence has not been properly assessed for the following
## examples. See Imai, Lu and Strauss (2008, Forthcoming) for more
## complete analyses.

## fit the parametric model with the default prior specification
res <- eco(Y ~ X, data = reg, verbose = TRUE)
## summarize the results
summary(res)

## obtain out-of-sample prediction
out <- predict(res, verbose = TRUE)
## summarize the results
summary(out)

## load the Robinson's census data
data(census)

## fit the parametric model with contextual effects and N
## using the default prior specification
res1 <- eco(Y ~ X, N = N, context = TRUE, data = census, verbose = TRUE)
## summarize the results
summary(res1)

## obtain out-of-sample prediction
out1 <- predict(res1, verbose = TRUE)
## summarize the results
summary(out1)
```

---

ecoBD *Calculating the Bounds for Ecological Inference in RxC Tables*

Description

ecoBD is used to calculate the bounds for missing internal cells of $R \times C$ ecological table. The data can be entered either in the form of counts or proportions.

Usage

```
ecoBD(formula, data = parent.frame(), N = NULL)
```

Arguments

`formula`	A symbolic description of ecological table to be used, specifying the column and row margins of $R \times C$ ecological tables. Details and specific examples are given below.
`data`	An optional data frame in which to interpret the variables in `formula`. The default is the environment in which `ecoBD` is called.
`N`	An optional variable representing the size of the unit; e.g., the total number of voters. If `formula` is entered as counts and the last row and/or column is omitted, this input is necessary.

Details

The data may be entered either in the form of counts or proportions. If proportions are used, `formula` may omit the last row and/or column of tables, which can be calculated from the remaining margins. For example, `Y ~ X` specifies `Y` as the first column margin and `X` as the first row margin in 2×2 tables. If counts are used, `formula` may omit the last row and/or column margin of the table only if `N` is supplied. In this example, the columns will be labeled as `X` and `not X`, and the rows will be labeled as `Y` and `not Y`.

For larger tables, one can use `cbind()` and `+`. For example, `cbind(Y1, Y2, Y3) ~ X1 + X2 + X3 + X4)` specifies 3×4 tables.

An $R \times C$ ecological table in the form of counts:

$$\begin{array}{ccccc} n_{i11} & n_{i12} & \ldots & n_{i1C} & n_{i1.} \\ n_{i21} & n_{i22} & \ldots & n_{i2C} & n_{i2.} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ n_{iR1} & n_{iR2} & \ldots & n_{iRC} & n_{iR.} \\ n_{i.1} & n_{i.2} & \ldots & n_{i.C} & N_i \end{array}$$

where $n_{nr.}$ and $n_{i.c}$ represent the observed margins, N_i represents the size of the table, and n_{irc} are unknown variables. Note that for each i, the following deterministic relationships hold; $n_{ir.} = \sum_{c=1}^{C} n_{irc}$ for $r = 1, \ldots, R$, and $n_{i.c} = \sum_{r=1}^{R} n_{irc}$ for $c = 1, \ldots, C$. Then, each of the unknown inner cells can be bounded in the following manner,

$$\max(0, n_{ir.} + n_{i.c} - N_i) \le n_{irc} \le \min(n_{ir.}, n_{i.c}).$$

If the size of tables, `N`, is provided,

An $R \times C$ ecological table in the form of proportions:

$$\begin{array}{ccccc} W_{i11} & W_{i12} & \ldots & W_{i1C} & Y_{i1} \\ W_{i21} & W_{i22} & \ldots & W_{i2C} & Y_{i2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ W_{iR1} & W_{iR2} & \ldots & W_{iRC} & Y_{iR} \\ X_{i1} & X_{i2} & \ldots & X_{iC} & \end{array}$$

where Y_{ir} and X_{ic} represent the observed margins, and W_{irc} are unknown variables. Note that for each i, the following deterministic relationships hold; $Y_{ir} = \sum_{c=1}^{C} X_{ic} W_{irc}$ for $r = 1, \ldots, R$, and

$\sum_{r=1}^{R} W_{irc} = 1$ for $c = 1, \ldots, C$. Then, each of the inner cells of the table can be bounded in the following manner,

$$\max(0, (X_{ic} + Y_{ir} - 1)/X_{ic}) \leq W_{irc} \leq \min(1, Y_{ir}/X_{ir}).$$

Value

An object of class `ecoBD` containing the following elements (When three dimensional arrays are used, the first dimension indexes the observations, the second dimension indexes the row numbers, and the third dimension indexes the column numbers):

`call`	The matched call.
`X`	A matrix of the observed row margin, X.
`Y`	A matrix of the observed column margin, Y.
`N`	A vector of the size of ecological tables, N.
`aggWmin`	A three dimensional array of aggregate lower bounds for proportions.
`aggWmax`	A three dimensional array of aggregate upper bounds for proportions.
`Wmin`	A three dimensional array of lower bounds for proportions.
`Wmax`	A three dimensional array of upper bounds for proportions.
`Nmin`	A three dimensional array of lower bounds for counts.
`Nmax`	A three dimensional array of upper bounds for counts.

The object can be printed through `print.ecoBD`.

Author(s)

Kosuke Imai, Department of Politics, Princeton University ⟨kimai@Princeton.Edu⟩, http://imai.princeton.edu/; Ying Lu, Institute for Quantitative Social Sciences, Harvard University ⟨ylu@Latte.Harvard.Edu⟩

References

Imai, Kosuke, Ying Lu and Aaron Strauss. (Forthcoming) "eco: R Package for Ecological Inference in 2x2 Tables" Journal of Statistical Software, available at http://imai.princeton.edu/research/eco.html

Imai, Kosuke, Ying Lu and Aaron Strauss. (Forthcoming) "Bayesian and Likelihood Inference for 2 x 2 Ecological Tables: An Incomplete Data Approach" Political Analysis, available at http://imai.princeton.edu/research/eiall.html

See Also

`eco`, `ecoNP`

Examples

```
## load the registration data
data(reg)

## calculate the bounds
res <- ecoBD(Y ~ X, N = N, data = reg)
## print the results
print(res)
```

---

ecoML *Fitting Parametric Models and Quantifying Missing Information for Ecological Inference in 2x2 Tables*

Description

ecoML is used to fit parametric models for ecological inference in 2×2 tables via Expectation Maximization (EM) algorithms. The data is specified in proportions. At it's most basic setting, the algorithm assumes that the individual-level proportions (i.e., W_1 and W_2) and distributed bivariate normally (after logit transformations). The function calculates point estimates of the parameters for models based on different assumptions. The standard errors of the point estimates are also computed via Supplemented EM algorithms. Moreover, ecoML quantifies the amount of missing information associated with each parameter and allows researcher to examine the impact of missing information on parameter estimation in ecological inference. The models and algorithms are described in Imai, Lu and Strauss (Forthcoming).

Usage

```
ecoML(formula, data = parent.frame(), N = NULL, supplement = NULL,
      theta.start = c(0,0,1,1,0), fix.rho = FALSE,
      context = FALSE, sem = TRUE, epsilon = 10^(-10),
  maxit = 1000, loglik = TRUE, hyptest = FALSE, verbose = FALSE)
```

Arguments

formula — A symbolic description of the model to be fit, specifying the column and row margins of 2×2 ecological tables. Y ~ X specifies Y as the column margin (e.g., turnout) and X (e.g., percent African-American) as the row margin. Details and specific examples are given below.

data — An optional data frame in which to interpret the variables in formula. The default is the environment in which ecoML is called.

N — An optional variable representing the size of the unit; e.g., the total number of voters. N needs to be a vector of same length as Y and X or a scalar.

supplement — An optional matrix of supplemental data. The matrix has two columns, which contain additional individual-level data such as survey data for W_1 and W_2, respectively. If NULL, no additional individual-level data are included in the model. The default is NULL.

`fix.rho`	Logical. If `TRUE`, the correlation (when `context=TRUE`) or the partial correlation (when `context=FALSE`) between W_1 and W_2 is fixed through the estimation. For details, see Imai, Lu and Strauss(2006). The default is `FALSE`.
`context`	Logical. If `TRUE`, the contextual effect is also modeled. In this case, the row margin (i.e., X) and the individual-level rates (i.e., W_1 and W_2) are assumed to be distributed tri-variate normally (after logit transformations). See Imai, Lu and Strauss (2006) for details. The default is `FALSE`.
`sem`	Logical. If `TRUE`, the standard errors of parameter estimates are estimated via SEM algorithm, as well as the fraction of missing data. The default is `TRUE`.
`theta.start`	A numeric vector that specifies the starting values for the mean, variance, and covariance. When `context = FALSE`, the elements of `theta.start` correspond to $(E(W_1), E(W_2), var(W_1), var(W_2), cor(W_1, W_2))$. When `context = TRUE`, the elements of `theta.start` correspond to $(E(W_1), E(W_2), var(W_1), var(W_2), corr(W_1, X), corr(W_2, X), corr(W_1, W_2))$. Moreover, when `fix.rho=TRUE`, $corr(W_1, W_2)$ is set to be the correlation between W_1 and W_2 when `context = FALSE`, and the partial correlation between W_1 and W_2 given X when `context = FALSE`. The default is `c(0,0,1,1,0)`.
`epsilon`	A positive number that specifies the convergence criterion for EM algorithm. The square root of `epsilon` is the convergence criterion for SEM algorithm. The default is `10^(-10)`.
`maxit`	A positive integer specifies the maximum number of iterations before the convergence criterion is met. The default is `1000`.
`loglik`	Logical. If `TRUE`, the value of the log-likelihood function at each iteration of EM is saved. The default is `TRUE`.
`hyptest`	Logical. If `TRUE`, model is estimated under the null hypothesis that means of $W1$ and $W2$ are the same. The default is `FALSE`.
`verbose`	Logical. If `TRUE`, the progress of the EM and SEM algorithms is printed to the screen. The default is `FALSE`.

Details

When `SEM` is `TRUE`, `ecoML` computes the observed-data information matrix for the parameters of interest based on Supplemented-EM algorithm. The inverse of the observed-data information matrix can be used to estimate the variance-covariance matrix for the parameters estimated from EM algorithms. In addition, it also computes the expected complete-data information matrix. Based on these two measures, one can further calculate the fraction of missing information associated with each parameter. See Imai, Lu and Strauss (2006) for more details about fraction of missing information.

Moreover, when `hytest=TRUE`, `ecoML` allows to estimate the parametric model under the null hypothesis that `mu_1=mu_2`. One can then construct the likelihood ratio test to assess the hypothesis of equal means. The associated fraction of missing information for the test statistic can be also calculated. For details, see Imai, Lu and Strauss (2006) for details.

Value

An object of class `ecoML` containing the following elements:

- `call` The matched call.
- `X` The row margin, X.
- `Y` The column margin, Y.
- `N` The size of each table, N.
- `context` The assumption under which model is estimated. If `context = FALSE`, CAR assumption is adopted and no contextual effect is modeled. If `context = TRUE`, NCAR assumption is adopted, and contextual effect is modeled.
- `sem` Whether SEM algorithm is used to estimate the standard errors and observed information matrix for the parameter estimates.
- `fix.rho` Whether the correlation or the partial correlation between W_1 an W_2 is fixed in the estimation.
- `r12` If `fix.rho = TRUE`, the value that $corr(W_1, W_2)$ is fixed to.
- `epsilon` The precision criterion for EM convergence. $\sqrt{\epsilon}$ is the precision criterion for SEM convergence.
- `theta.sem` The ML estimates of $E(W_1)$,$E(W_2)$, $var(W_1)$,$var(W_2)$, and $cov(W_1, W_2)$. If `context = TRUE`, $E(X)$,$cov(W_1, X)$, $cov(W_2, X)$ are also reported.
- `W` In-sample estimation of W_1 and W_2.
- `suff.stat` The sufficient statistics for `theta.em`.
- `iters.em` Number of EM iterations before convergence is achieved.
- `iters.sem` Number of SEM iterations before convergence is achieved.
- `loglik` The log-likelihood of the model when convergence is achieved.
- `loglik.log.em` A vector saving the value of the log-likelihood function at each iteration of the EM algorithm.
- `mu.log.em` A matrix saving the unweighted mean estimation of the logit-transformed individual-level proportions (i.e., W_1 and W_2) at each iteration of the EM process.
- `Sigma.log.em` A matrix saving the log of the variance estimation of the logit-transformed individual-level proportions (i.e., W_1 and W_2) at each iteration of EM process. Note, non-transformed variances are displayed on the screen (when `verbose = TRUE`).
- `rho.fisher.em` A matrix saving the fisher transformation of the estimation of the correlations between the logit-transformed individual-level proportions (i.e., W_1 and W_2) at each iteration of EM process. Note, non-transformed correlations are displayed on the screen (when `verbose = TRUE`).
- `DM` The matrix characterizing the rates of convergence of the EM algorithms. Such information is also used to calculate the observed-data information matrix
- `Icom` The (expected) complete data information matrix estimated via SEM algorithm. When `context=FALSE, fix.rho=TRUE`, `Icom` is 4 by 4. When `context=FALSE, fix.rho=FALSE`, `Icom` is 5 by 5. When `context=TRUE`, `Icom` is 9 by 9.
- `Iobs` The observed information matrix. The dimension of `Iobs` is same as `Icom`.

`Imiss`	The difference between `Icom` and `Iobs`. The dimension of `Imiss` is same as `miss`.
`Vobs`	The (symmetrized) variance-covariance matrix of the ML parameter estimates. The dimension of `Vobs` is same as `Icom`.
`Iobs`	The (expected) complete-data variance-covariance matrix. The dimension of `Iobs` is same as `Icom`.
`Vobs.original`	The estimated variance-covariance matrix of the ML parameter estimates. The dimension of `Vobs` is same as `Icom`.
`Fmis`	The fraction of missing information associated with each parameter estimation.
`VFmis`	The proportion of increased variance associated with each parameter estimation due to observed data.
`Ieigen`	The largest eigen value of `Imiss`.
`Icom.trans`	The complete data information matrix for the fisher transformed parameters.
`Iobs.trans`	The observed data information matrix for the fisher transformed parameters.
`Fmis.trans`	The fractions of missing information associated with the fisher transformed parameters.

Author(s)

Kosuke Imai, Department of Politics, Princeton University, ⟨kimai@Princeton.Edu⟩, http://imai.princeton.edu; Ying Lu, Department of Humanities and Social Sciences in the Professions, Steinhardt School of Culture, Education and Human Development, New York University, ⟨yl46@Nyu.Edu⟩; Aaron Strauss, Department of Politics, Princeton University, ⟨abstraus@Princeton.Edu⟩.

References

Imai, Kosuke, Ying Lu and Aaron Strauss. (Forthcoming). "eco: R Package for Ecological Inference in 2x2 Tables" Journal of Statistical Software, available at http://imai.princeton.edu/research/eco.html

Imai, Kosuke, Ying Lu and Aaron Strauss. (Forthcoming). "Bayesian and Likelihood Inference for 2 x 2 Ecological Tables: An Incomplete Data Approach" Political Analysis, available at http://imai.princeton.edu/research/eiall.html

See Also

`eco`, `ecoNP`, `summary.ecoML`

Examples

```
## load the census data
data(census)

## NOTE: convergence has not been properly assessed for the following
## examples. See Imai, Lu and Strauss (2006) for more complete analyses.
## In the first example below, in the interest of time, only part of the
```

```
## data set is analyzed and the convergence requirement is less stringent
## than the default setting.

## In the second example, the program is arbitrarily halted 100 iterations
## into the simulation, before convergence.

## load the Robinson's census data
data(census)

## fit the parametric model with the default model specifications
## Not run: res <- ecoML(Y ~ X, data = census[1:100,],N=census[1:100,3],epsilon=10^(-6), ver
## summarize the results
## Not run: summary(res)

## obtain out-of-sample prediction
## Not run: out <- predict(res, verbose = TRUE)
## summarize the results
## Not run: summary(out)

## fit the parametric model with some individual
## level data using the default prior specification
surv <- 1:600
## Not run:
res1 <- ecoML(Y ~ X, context = TRUE, data = census[-surv,],
                   supplement = census[surv,c(4:5,1)], maxit=100, verbose = TRUE)
## End(Not run)
## summarize the results
## Not run: summary(res1)
```

`ecoNP` *Fitting the Nonparametric Bayesian Models of Ecological Inference in 2x2 Tables*

Description

`ecoNP` is used to fit the nonparametric Bayesian model (based on a Dirichlet process prior) for ecological inference in 2×2 tables via Markov chain Monte Carlo. It gives the in-sample predictions as well as out-of-sample predictions for population inference. The models and algorithms are described in Imai, Lu and Strauss (2008, Forthcoming).

Usage

```
ecoNP(formula, data = parent.frame(), N = NULL, supplement = NULL,
      context = FALSE, mu0 = 0, tau0 = 2, nu0 = 4, S0 = 10,
      alpha = NULL, a0 = 1, b0 = 0.1, parameter = FALSE,
      grid = FALSE, n.draws = 5000, burnin = 0, thin = 0,
      verbose = FALSE)
```

Arguments

`formula` A symbolic description of the model to be fit, specifying the column and row margins of 2×2 ecological tables. `Y ~ X` specifies `Y` as the column margin (e.g., turnout) and `X` as the row margin (e.g., percent African-American). Details and specific examples are given below.

`data` An optional data frame in which to interpret the variables in `formula`. The default is the environment in which `ecoNP` is called.

`N` An optional variable representing the size of the unit; e.g., the total number of voters. `N` needs to be a vector of same length as `Y` and `X` or a scalar.

`supplement` An optional matrix of supplemental data. The matrix has two columns, which contain additional individual-level data such as survey data for W_1 and W_2, respectively. If `NULL`, no additional individual-level data are included in the model. The default is `NULL`.

`context` Logical. If `TRUE`, the contextual effect is also modeled, that is to assume the row margin X and the unknown W_1 and W_2 are correlated. See Imai, Lu and Strauss (2008, Forthcoming) for details. The default is `FALSE`.

`mu0` A scalar or a numeric vector that specifies the prior mean for the mean parameter μ of the base prior distribution G_0 (see Imai, Lu and Strauss (2008, Forthcoming) for detailed descriptions of Dirichlete prior and the normal base prior distribution) . If it is a scalar, then its value will be repeated to yield a vector of the length of μ, otherwise, it needs to be a vector of same length as μ. When `context=TRUE` , the length of μ is 3, otherwise it is 2. The default is `0`.

`tau0` A positive integer representing the scale parameter of the Normal-Inverse Wishart prior for the mean and variance parameter (μ_i, Σ_i) of each observation. The default is `2`.

`nu0` A positive integer representing the prior degrees of freedom of the variance matrix Σ_i. the default is `4`.

`S0` A positive scalar or a positive definite matrix that specifies the prior scale matrix for the variance matrix Σ_i. If it is a scalar, then the prior scale matrix will be a diagonal matrix with the same dimensions as Σ_i and the diagonal elements all take value of `S0`, otherwise `S0` needs to have same dimensions as Σ_i. When `context=TRUE`, Σ is a 3×3 matrix, otherwise, it is 2×2. The default is `10`.

`alpha` A positive scalar representing a user-specified fixed value of the concentration parameter, α. If `NULL`, α will be updated at each Gibbs draw, and its prior parameters `a0` and `b0` need to be specified. The default is `NULL`.

`a0` A positive integer representing the value of shape parameter of the gamma prior distribution for α. The default is `1`.

`b0` A positive integer representing the value of the scale parameter of the gamma prior distribution for α. The default is `0.1`.

`parameter` Logical. If `TRUE`, the Gibbs draws of the population parameters, μ and Σ, are returned in addition to the in-sample predictions of the missing internal cells, W. The default is `FALSE`. This needs to be set to `TRUE` if one wishes to make population inferences through `predict.eco`. See an example below.

grid	Logical. If TRUE, the grid method is used to sample W in the Gibbs sampler. If FALSE, the Metropolis algorithm is used where candidate draws are sampled from the uniform distribution on the tomography line for each unit. Note that the grid method is significantly slower than the Metropolis algorithm.
n.draws	A positive integer. The number of MCMC draws. The default is 5000.
burnin	A positive integer. The burnin interval for the Markov chain; i.e. the number of initial draws that should not be stored. The default is 0.
thin	A positive integer. The thinning interval for the Markov chain; i.e. the number of Gibbs draws between the recorded values that are skipped. The default is 0.
verbose	Logical. If TRUE, the progress of the Gibbs sampler is printed to the screen. The default is FALSE.

Value

An object of class ecoNP containing the following elements:

call	The matched call.
X	The row margin, X.
Y	The column margin, Y.
burnin	The number of initial burnin draws.
thin	The thinning interval.
nu0	The prior degrees of freedom.
tau0	The prior scale parameter.
mu0	The prior mean.
S0	The prior scale matrix.
a0	The prior shape parameter.
b0	The prior scale parameter.
W	A three dimensional array storing the posterior in-sample predictions of W. The first dimension indexes the Monte Carlo draws, the second dimension indexes the columns of the table, and the third dimension represents the observations.
Wmin	A numeric matrix storing the lower bounds of W.
Wmax	A numeric matrix storing the upper bounds of W.
mu	A three dimensional array storing the posterior draws of the population mean parameter, μ. The first dimension indexes the Monte Carlo draws, the second dimension indexes the columns of the table, and the third dimension represents the observations.
Sigma	A three dimensional array storing the posterior draws of the population variance matrix, Σ. The first dimension indexes the Monte Carlo draws, the second dimension indexes the parameters, and the third dimension represents the observations.
alpha	The posterior draws of α.
nstar	The number of clusters at each Gibbs draw.

Author(s)

Kosuke Imai, Department of Politics, Princeton University, ⟨kimai@Princeton.Edu⟩, http://imai.princeton.edu; Ying Lu, Department of Humanities and Social Sciences in the Professions, Steinhardt School of Culture, Education and Human Development, New York University, ⟨yl46@Nyu.Edu⟩

References

Imai, Kosuke, Ying Lu and Aaron Strauss. (Forthcoming). "eco: R Package for Ecological Inference in 2x2 Tables" Journal of Statistical Software, available at http://imai.princeton.edu/research/eco.html

Imai, Kosuke, Ying Lu and Aaron Strauss. (2008). "Bayesian and Likelihood Inference for 2 x 2 Ecological Tables: An Incomplete Data Approach" Political Analysis, Vol. 16, No. 1 (Winter), pp. 41-69. available at http://imai.princeton.edu/research/eiall.html

See Also

`eco`, `ecoML`, `predict.eco`, `summary.ecoNP`

Examples

```
## load the registration data
data(reg)

## NOTE: We set the number of MCMC draws to be a very small number in
## the following examples; i.e., convergence has not been properly
## assessed. See Imai, Lu and Strauss (2006) for more complete examples.

## fit the nonparametric model to give in-sample predictions
## store the parameters to make population inference later
res <- ecoNP(Y ~ X, data = reg, n.draws = 50, param = TRUE, verbose = TRUE)

##summarize the results
summary(res)

## obtain out-of-sample prediction
out <- predict(res, verbose = TRUE)

## summarize the results
summary(out)

## density plots of the out-of-sample predictions
par(mfrow=c(2,1))
plot(density(out[,1]), main = "W1")
plot(density(out[,2]), main = "W2")

## load the Robinson's census data
data(census)

## fit the parametric model with contextual effects and N
```

```
## using the default prior specification

res1 <- ecoNP(Y ~ X, N = N, context = TRUE, param = TRUE, data = census,
n.draws = 25, verbose = TRUE)

## summarize the results
summary(res1)

## out-of sample prediction
pres1 <- predict(res1)
summary(pres1)
```

`forgnlit30` *Foreign-born literacy in 1930*

Description

This data set contains, on a state level, the proportion of white residents ten years and older who are foreign born, and the proportion of those residents who are literate. Data come from the 1930 census and were first analyzed by Robinson (1950).

Usage

```
data(forgnlit30)
```

Format

A data frame containing 5 variables and 48 observations

X	numeric	proportion of the white population at least 10 years of age that is foreign born
Y	numeric	proportion of the white population at least 10 years of age that is illiterate
W1	numeric	proportion of the foreign-born white population at least 10 years of age that is illiterate
W2	numeric	proportion of the native-born white population at least 10 years of age that is illiterate
ICPSR	numeric	the ICPSR state code

References

Robinson, W.S. (1950). “Ecological Correlations and the Behavior of Individuals.” *American Sociological Review*, vol. 15, pp.351-357.

`forgnlit30c` *Foreign-born literacy in 1930, County Level*

Description

This data set contains, on a county level, the proportion of white residents ten years and older who are foreign born, and the proportion of those residents who are literate. Data come from the 1930 census and were first analyzed by Robinson (1950). Counties with fewer than 100 foreign born residents are dropped.

Usage

```
data(forgnlit30c)
```

Format

A data frame containing 6 variables and 1976 observations

X	numeric	proportion of the white population at least 10 years of age that is foreign born
Y	numeric	proportion of the white population at least 10 years of age that is illiterate
W1	numeric	proportion of the foreign-born white population at least 10 years of age that is illiterate
W2	numeric	proportion of the native-born white population at least 10 years of age that is illiterate
state	numeric	the ICPSR state code
county	numeric	the ICPSR (within state) county code

References

Robinson, W.S. (1950). "Ecological Correlations and the Behavior of Individuals." *American Sociological Review*, vol. 15, pp.351-357.

`housep88` *Electoral Results for the House and Presidential Races in 1988*

Description

This data set contains, on a House district level, the percentage of the vote for the Democratic House candidate, the percentage of the vote for the Democratic presidential candidate (Dukakis), the number of voters who voted for a major party candidate in the presidential race, and the ratio of voters in the House race versus the number who cast a ballot for President. Eleven (11) uncontested races are not included. Dataset compiled and analyzed by Burden and Kimball (1988). Complete dataset and documentation available at ICSPR study number 1140.

Usage

```
data(housep88)
```

Format

A data frame containing 5 variables and 424 observations

X	numeric	proportion voting for the Democrat in the presidential race
Y	numeric	proportion voting for the Democrat in the House race
N	numeric	number of major party voters in the presidential contest
HPCT	numeric	House election turnout divided by presidential election turnout (set to 1 if House turnout exceeds president
DIST	numeric	4-digit ICPSR state and district code: first 2 digits for the state code, last two digits for the district number

References

Burden, Barry C. and David C. Kimball (1988). “A New Approach To Ticket- Splitting.” The American Political Science Review. vol 92., no. 3, pp. 553-544.

`predict.eco` *Out-of-Sample Posterior Prediction under the Parametric Bayesian Model for Ecological Inference in 2x2 Tables*

Description

Obtains out-of-sample posterior predictions under the fitted parametric Bayesian model for ecological inference. `predict` method for class `eco` and `ecoX`.

Usage

```
## S3 method for class 'eco':
predict(object, newdraw = NULL, subset = NULL,
                 verbose = FALSE, ...)
## S3 method for class 'ecoX':
predict(object, newdraw = NULL, subset = NULL,
                 newdata = NULL, cond = FALSE, verbose = FALSE, ...)
```

Arguments

`object`	An output object from `eco` or `ecoNP`.
`newdraw`	An optional list containing two matrices (or three dimensional arrays for the nonparametric model) of MCMC draws of μ and Σ. Those elements should be named as `mu` and `Sigma`, respectively. The default is the original MCMC draws stored in `object`.
`newdata`	An optional data frame containing a new data set for which posterior predictions will be made. The new data set must have the same variable names as those in the original data.
`subset`	A scalar or numerical vector specifying the row number(s) of `mu` and `Sigma` in the output object from `eco`. If specified, the posterior draws of parameters for those rows are used for posterior prediction. The default is `NULL` where all the posterior draws are used.
`cond`	logical. If `TRUE`, then the conditional prediction will made for the parametric model with contextual effects. The default is `FALSE`.
`verbose`	logical. If `TRUE`, helpful messages along with a progress report on the Monte Carlo sampling from the posterior predictive distributions are printed on the screen. The default is `FALSE`.
`...`	further arguments passed to or from other methods.

Details

The posterior predictive values are computed using the Monte Carlo sample stored in the `eco` output (or other sample if `newdraw` is specified). Given each Monte Carlo sample of the parameters, we sample the vector-valued latent variable from the appropriate multivariate Normal distribution. Then, we apply the inverse logit transformation to obtain the predictive values of proportions, W. The computation may be slow (especially for the nonparametric model) if a large Monte Carlo sample of the model parameters is used. In either case, setting `verbose = TRUE` may be helpful in monitoring the progress of the code.

Value

`predict.eco` yields a matrix of class `predict.eco` containing the Monte Carlo sample from the posterior predictive distribution of inner cells of ecological tables. `summary.predict.eco` will summarize the output, and `print.summary.predict.eco` will print the summary.

Author(s)

Kosuke Imai, Department of Politics, Princeton University, ⟨kimai@Princeton.Edu⟩, http://imai.princeton.edu; Ying Lu, Department of Sociology, University of Colorado at Boulder, ⟨ying.lu@Colorado.Edu⟩

See Also

`eco`, `predict.ecoNP`

`predict.ecoNP` *Out-of-Sample Posterior Prediction under the Nonparametric Bayesian Model for Ecological Inference in 2x2 Tables*

Description

Obtains out-of-sample posterior predictions under the fitted nonparametric Bayesian model for ecological inference. `predict` method for class `ecoNP` and `ecoNPX`.

Usage

```
## S3 method for class 'ecoNP':
predict(object, newdraw = NULL, subset = NULL, obs = NULL,
                 verbose = FALSE, ...)
## S3 method for class 'ecoNPX':
predict(object, newdraw = NULL, subset = NULL, obs = NULL,
                 cond = FALSE, verbose = FALSE, ...)
```

Arguments

object	An output object from ecoNP.
newdraw	An optional list containing two matrices (or three dimensional arrays for the nonparametric model) of MCMC draws of μ and Σ. Those elements should be named as mu and Sigma, respectively. The default is the original MCMC draws stored in object.
subset	A scalar or numerical vector specifying the row number(s) of mu and Sigma in the output object from eco. If specified, the posterior draws of parameters for those rows are used for posterior prediction. The default is NULL where all the posterior draws are used.
obs	An integer or vector of integers specifying the observation number(s) whose posterior draws will be used for predictions. The default is NULL where all the observations in the data set are selected.
cond	logical. If TRUE, then the conditional prediction will made for the parametric model with contextual effects. The default is FALSE.
verbose	logical. If TRUE, helpful messages along with a progress report on the Monte Carlo sampling from the posterior predictive distributions are printed on the screen. The default is FALSE.
...	further arguments passed to or from other methods.

Details

The posterior predictive values are computed using the Monte Carlo sample stored in the eco or ecoNP output (or other sample if newdraw is specified). Given each Monte Carlo sample of the parameters, we sample the vector-valued latent variable from the appropriate multivariate Normal distribution. Then, we apply the inverse logit transformation to obtain the predictive values of proportions, W. The computation may be slow (especially for the nonparametric model) if a large Monte Carlo sample of the model parameters is used. In either case, setting verbose = TRUE may be helpful in monitoring the progress of the code.

Value

predict.eco yields a matrix of class predict.eco containing the Monte Carlo sample from the posterior predictive distribution of inner cells of ecological tables. summary.predict.eco will summarize the output, and print.summary.predict.eco will print the summary.

Author(s)

Kosuke Imai, Department of Politics, Princeton University, ⟨kimai@Princeton.Edu⟩, http://imai.princeton.edu; Ying Lu, Department of Sociology, University of Colorado at Boulder, ⟨ying.lu@Colorado.Edu⟩

See Also

eco, ecoNP, summary.eco, summary.ecoNP

Qfun *Fitting the Parametric Bayesian Model of Ecological Inference in 2x2 Tables*

Description

Qfun returns the complete log-likelihood that is used to calculate the fraction of missing information.

Usage

```
Qfun(theta, suff.stat, n)
```

Arguments

theta	A vector that contains the MLE $E(W_1)$,$E(W_2)$, $var(W_1)$,$var(W_2)$, and $cov(W_1, W_2)$. Typically it is the element theta.em of an object of class ecoML.
suff.stat	A vector of sufficient statistics of $E(W_1)$, $E(W_2)$, $var(W_1)$,$var(W_2)$, and $cov(W_1, W_2)$.
n	A integer representing the sample size.

Author(s)

Kosuke Imai, Department of Politics, Princeton University, ⟨kimai@Princeton.Edu⟩, http://imai.princeton.edu; Ying Lu, Department of Humanities and Social Sciences in the Professions, Steinhardt School of Culture, Education and Human Development, New York University, ⟨yl46@Nyu.Edu⟩ Aaron Strauss, Department of Politics, Princeton University, ⟨abstraus@Princeton.Edu⟩.

References

Imai, Kosuke, Ying Lu and Aaron Strauss. (Forthcoming). "eco: R Package for Ecological Inference in 2x2 Tables" Journal of Statistical Software, available at http://imai.princeton.edu/research/eco.html

Imai, Kosuke, Ying Lu and Aaron Strauss. (2008). "Bayesian and Likelihood Inference for 2 x 2 Ecological Tables: An Incomplete Data Approach" Political Analysis, Vol. 16, No. 1 (Winter), pp. 41-69. available at http://imai.princeton.edu/research/eiall.html

See Also

ecoML

reg — *Voter Registration in US Southern States*

Description

This data set contains the racial composition, the registration rate, the number of eligible voters as well as the actual observed racial registration rates for every county in four US southern states: Florida, Louisiana, North Carolina, and South Carolina.

Usage

```
data(reg)
```

Format

A data frame containing 5 variables and 275 observations

X	numeric	the fraction of Black voters
Y	numeric	the fraction of voters who registered themselves
N	numeric	the total number of voters in each county
W1	numeric	the actual fraction of Black voters who registered themselves
W2	numeric	the actual fraction of White voters who registered themselves

References

King, G. (1997). "A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data". Princeton University Press, Princeton, NJ.

summary.eco — *Summarizing the Results for the Bayesian Parametric Model for Ecological Inference in 2x2 Tables*

Description

`summary` method for class `eco`.

Usage

```
## S3 method for class 'eco':
summary(object, CI = c(2.5, 97.5), param = TRUE,
        units = FALSE, subset = NULL, ...)

## S3 method for class 'summary.eco':
print(x, digits = max(3, getOption("digits") - 3), ...)
```

Arguments

`object`	An output object from `eco`.
`CI`	A vector of lower and upper bounds for the Bayesian credible intervals used to summarize the results. The default is the equal tail 95 percent credible interval.
`x`	An object of class `summary.eco`.
`digits`	the number of significant digits to use when printing.
`param`	Logical. If `TRUE`, the posterior estimates of the population parameters will be provided. The default value is `TRUE`.
`units`	Logical. If `TRUE`, the in-sample predictions for each unit or for a subset of units will be provided. The default value is `FALSE`.
`subset`	A numeric vector indicating the subset of the units whose in-sample predications to be provided when `units` is `TRUE`. The default value is `NULL` where the in-sample predictions for each unit will be provided.
`...`	further arguments passed to or from other methods.

Value

`summary.eco` yields an object of class `summary.eco` containing the following elements:

`call`	The call from `eco`.
`n.obs`	The number of units.
`n.draws`	The number of Monte Carlo samples.
`agg.table`	Aggregate posterior estimates of the marginal means of W_1 and W_2 using X and N as weights.
`param.table`	Posterior estimates of model parameters: population mean estimates of W_1 and W_2 and their logit transformations.
`W1.table`	Unit-level posterior estimates for W_1.
`W2.table`	Unit-level posterior estimates for W_2.

This object can be printed by `print.summary.eco`

Author(s)

Kosuke Imai, Department of Politics, Princeton University, ⟨kimai@Princeton.Edu⟩, `http://imai.princeton.edu`; Ying Lu, Department of Humanities and Social Sciences in the Professions, Steinhardt School of Culture, Education and Human Development, New York University, ⟨yl46@Nyu.Edu⟩

See Also

`eco`, `predict.eco`

`summary.ecoML` *Summarizing the Results for the Maximum Likelihood Parametric Model for Ecological Inference in 2x2 Tables*

Description

`summary` method for class `eco`.

Usage

```
## S3 method for class 'ecoML':
summary(object, CI = c(2.5, 97.5),  param = TRUE, units = FALSE,
   subset = NULL, ...)

## S3 method for class 'summary.ecoML':
print(x, digits = max(3, getOption("digits") - 3), ...)
```

Arguments

`object`	An output object from `eco`.
`CI`	A vector of lower and upper bounds for the Bayesian credible intervals used to summarize the results. The default is the equal tail 95 percent credible interval.
`param`	Ignored.
`subset`	A numeric vector indicating the subset of the units whose in-sample predications to be provided when `units` is `TRUE`. The default value is `NULL` where the in-sample predictions for each unit will be provided.
`units`	Logical. If `TRUE`, the in-sample predictions for each unit or for a subset of units will be provided. The default value is `FALSE`.
`x`	An object of class `summary.ecoML`.
`digits`	the number of significant digits to use when printing.
`...`	further arguments passed to or from other methods.

Value

`summary.eco` yields an object of class `summary.eco` containing the following elements:

`call`	The call from `eco`.
`sem`	Whether the SEM algorithm was executed, as specified by the user upon calling `ecoML`.
`fix.rho`	Whether the correlation parameter was fixed or allowed to vary, as specified by the user upon calling `ecoML`.
`epsilon`	The convergence threshold specified by the user upon calling `ecoML`.
`n.obs`	The number of units.

iters.em	The number iterations the EM algorithm cycled through before convergence or reaching the maximum number of iterations allowed.
iters.sem	The number iterations the SEM algorithm cycled through before convergence or reaching the maximum number of iterations allowed.
loglik	The final observed log-likelihood.
rho	A matrix of `iters.em` rows specifying the correlation parameters at each iteration of the EM algorithm. The number of columns depends on how many correlation parameters exist in the model. Column order is the same as the order of the parameters in `param.table`.
param.table	Final estimates of the parameter values for the model. Excludes parameters fixed by the user upon calling `ecoML`. See `ecoML` documentation for order of parameters.
agg.table	Aggregate estimates of the marginal means of W_1 and W_2
agg.wtable	Aggregate estimates of the marginal means of W_1 and W_2 using X and N as weights.
W.table	Unit-level estimates for W_1 and W_2.

This object can be printed by `print.summary.eco`

Author(s)

Kosuke Imai, Department of Politics, Princeton University, ⟨kimai@Princeton.Edu⟩, http://imai.princeton.edu; Ying Lu, Department of Humanities and Social Sciences in the Professions, Steinhardt School of Culture, Education and Human Development, New York University, ⟨yl46@Nyu.Edu⟩; Aaron Strauss, Department of Politics, Princeton University, ⟨abstraus@Princeton.Edu⟩

See Also

`ecoML`

---

summary.ecoNP — *Summarizing the Results for the Bayesian Nonparametric Model for Ecological Inference in 2x2 Tables*

Description

`summary` method for class `ecoNP`.

Usage

```
## S3 method for class 'ecoNP':
summary(object, CI = c(2.5, 97.5), param = FALSE,
        units = FALSE, subset = NULL, ...)

## S3 method for class 'summary.ecoNP':
print(x, digits = max(3, getOption("digits") - 3), ...)
```

Arguments

object	An output object from ecoNP.
CI	A vector of lower and upper bounds for the Bayesian credible intervals used to summarize the results. The default is the equal tail 95 percent credible interval.
x	An object of class summary.ecoNP.
digits	the number of significant digits to use when printing.
param	Logical. If TRUE, the posterior estimates of the population parameters will be provided. The default value is FALSE.
units	Logical. If TRUE, the in-sample predictions for each unit or for a subset of units will be provided. The default value is FALSE.
subset	A numeric vector indicating the subset of the units whose in-sample predications to be provided when units is TRUE. The default value is NULL where the in-sample predictions for each unit will be provided.
...	further arguments passed to or from other methods.

Value

summary.ecoNP yields an object of class summary.ecoNP containing the following elements:

call	The call from ecoNP.
n.obs	The number of units.
n.draws	The number of Monte Carlo samples.
agg.table	Aggregate posterior estimates of the marginal means of W_1 and W_2 using X and N as weights.
param.table	Posterior estimates of model parameters: population mean estimates of W_1 and W_2. If subset is specified, only a subset of the population parameters are included.
W1.table	Unit-level posterior estimates for W_1.
W2.table	Unit-level posterior estimates for W_2.

This object can be printed by print.summary.ecoNP

Author(s)

Kosuke Imai, Department of Politics, Princeton University, ⟨kimai@Princeton.Edu⟩, http://imai.princeton.edu; Ying Lu, Department of Humanities and Social Sciences in the Professions, Steinhardt School of Culture, Education and Human Development, New York University, ⟨yl46@Nyu.Edu⟩

See Also

ecoNP, predict.eco

`wallace` *Black voting rates for Wallace for President, 1968*

Description

This data set contains, on a county level, the proportion of county residents who are Black and the proportion of presidential votes cast for Wallace. Demographic data is based on the 1960 census. Presidential returns are from ICPSR study 13. County data from 10 southern states (Alabama, Arkansas, Georgia, Florida, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas) are included. (Virginia is excluded due to the difficulty of matching counties between the datasets.) This data is analyzed in Wallace and Segal (1973).

Usage

```
data(wallace)
```

Format

A data frame containing 3 variables and 1009 observations

X	numeric	proportion of the population that is Black
Y	numeric	proportion presidential votes cast for Wallace
FIPS	numeric	the FIPS county code

References

Wasserman, Ira M. and David R. Segal (1973). “Aggregation Effects in the Ecological Study of Presidential Voting.” American Journal of Political Science. vol. 17, pp. 177-81.

Index