Package 'reweight'

February 20, 2015

Type Package

Title Adjustment of Survey Respondent Weights

Version 1.2.1

Date 2006-06-30

Author Feiming Chen

Maintainer Feiming Chen <cfm271-stat@yahoo.com>

Description Adjusts the weights of survey respondents so that the marginal distributions of certain variables fit more closely to those from a more precise source (e.g. Census Bureau's data).

License GPL (>= 2)

Repository CRAN

Date/Publication 2012-04-12 11:14:54

NeedsCompilation no

R topics documented:

---

pumswgt   *Household Distribution of Tenure and Household Size From PUMS*

Description

This data gives the joint and marginal distributions of Tenure (2 levels) and Household Size (5 levels) in Florida from US Census data.

Usage

```
data(pumswgt)
```

Format

A list with 4 components:

ori   A data frame with 10 observations on the following 2 variables.

    tenure   Factor *Tenure* with two levels: 1=Owner, 2=Renter

    hhsize   Factor *Household Size* with five levels: 1=1 Person, 2=2 Person,3=3 Person,4=4 Person,5=5+ Person

mar   A vector of marginal counts with the following 7 values.

    ten1   Counts of owned households (Tenure)

    ten2   Counts of rented households (Tenure)

    np1   Counts of 1 person households (Household Size)

    np2   Counts of 2 person households (Household Size)

    np3   Counts of 3 person households (Household Size)

    np4   Counts of 4 person households (Household Size)

    np5   Counts of 5 person households (Household Size)

raw   Raw counts of households in each factor level combination.

wgt   Original weights of households in each factor level combination

Details

The ori, raw, and wgt components are from US Census ACS (American Community Survey) PUMS (Public Use Micro Sample) 2004 data set containing two demographic factors: Tenure (*ten*) and Household Size (*np*), along with a weight variable *wgtp*, for the state Florida. They are further collapsed using the R function aggregate so that each factor combination in ori is unique.

The mar component gives the marginal distribution of Tenure (2 levels) and Household Size (5 levels) from US Census SF1 (Summary File 1) 2000 data containing table H4 (Tenure) and H13 (Household Size) for the state Florida.

Source

The data is downloaded from two data sources in Census website http://dataferrett.census.gov:

- ACS (American Community Survey) PUMS (Public Use Micro Sample) 2004.
- SF1 (Summary File 1) 2000.

References

Feiming Chen (2006) *A Heuristic Method for Weighting Survey Respondents*. JSM 2006 Proceedings.

See Also

reweight

Examples

```
data(pumswgt)
```

---

reweight   *Adjust survey respondent weights*

Description

Adjusts the weights of survey repondents so that the marginal distributions of certain variables fit more closely to those from a more precise source (e.g. Census Bureau's data).

Usage

```
reweight(ori, mar, raw=NA, wgt=NA, unique=T, bound=c(0, 100),
trace=F, tolerance=0.1, penalty=0, ...)

## S3 method for class 'reweight'
print(x, ...)
## S3 method for class 'reweight'
summary(object, ...)
## S3 method for class 'reweight'
plot(x, ...)
```

Arguments

ori
: a matrix containing the factor levels. The levels should start from 1 and count upwards as in $1, 2, \ldots$.

mar
: a vector giving the marginal distributions for each of the factors that are listed in the same order as in `ori`.

raw
: a vector of the raw counts of survey respondents corresponding to each line in `ori`. Default is all one's.

wgt
: a vector of the original weights corresponding to each line in `ori`. Default is the value of `raw`.

unique
: whether the factor level combination in `ori` matrix is already unique. If `FALSE`, it will be made so by aggregating the non-unique lines, along with `raw` and `wgt`. Default is `TRUE`.

bound
: a vector of two elements giving the lower and upper bounds for the final weight ratios. The extreme weight ratio is reset to either the lower or upper bound. Default is c(0,100). In application a more practical bound might be c(0.5,2).

trace
: if `TRUE`, show the path of the golden selection search of best regularization parameter r. Default is `FALSE`.

tolerance
: the percentage of the smallest eigenvalue that is to be used as the lower start point of the golden selection searhch. Default is 0.1.

penalty
: measures the strength of a penalty term (it puts penalty if the number of zero weighting ratios is large) in the GCV function. It is a multiplicative factor of the form $(1+q)^p$, where p is the `penalty` and q is the percentage of zero weighting ratios. Default is 0 (no penalty).

x	an object returned by calling `reweight`.
object	an object returned by calling `reweight`.
...	parameters to be passed to the generic fucntion.

Details

`ori`, `raw`, `wgt` typically come from survey data with categorical responses. The intent is to adjust the `wgt` so that the survey sample is more representative of the universe from where it comes. It is accomplished by fitting the marginal distributions of the sample to those of the universe, or those from a more precise source (e.g. census data). The method is based on the Tikhonov regularization.

The `print` method prints out the weight ratios, along with their corresponding factor level combinations. This data can then be matched back to the original sample data to adjust the original weights (by multipling each original weight with the weight ratio).

The `summary` method prints out various running statistics.

The `plot` method makes a panel of four diagnostic plots.

Value

None.

Author(s)

Feiming Chen

References

Feiming Chen (2006) *A Heuristic Method for Weighting Survey Respondents*. JSM 2006 Proceedings.

See Also

pumswgt

Examples

```
data(pumswgt)
r1 <- reweight(pumswgt$ori,pumswgt$mar,pumswgt$wgt)
plot(r1)
summary(r1)
r2 <- reweight(pumswgt$ori,pumswgt$mar,pumswgt$raw)
plot(r2)
summary(r2)
w <- print(r2)
```

Index

5