Package 'weightTAPSPACK'

September 25, 2015

Title Weight TAPS Data

Version 0.1

Author David G. Carlson, Michelle Torres, and Taeyong Park

Maintainer David G. Carlson <carlson.david@wustl.edu>

Description

The weightTAPSPACK subsets The American Panel Survey (TAPS) data by outcome and covariates, models the attrition rates, imputes data for attrited individuals, and finds weights for analysis.

Depends R (>= 3.0.0), methods, survey, plyr, HotDeckImputation, mice

License GPL (>= 2)

LazyDataCompression xz

LazyData yes

R topics documented:

---

attritTAPS   *Find attrition stats for TAPS waves*

Description

Find response attrition statistics for TAPS waves

Usage

```
attritTAPS(outcome)
```

Arguments

outcome   A character vector of the names of outcome variables of interest

Value

A list of attrition rates in population margins. The number in each category corresponds to the percentage of people within that category that left/attrited the panel

Author(s)

David G. Carlson <carlson.david@wustl.edu> Michelle Torres: <smtorres@wustl> Taeyong Park <t.park@wustl.edu>

See Also

weightTAPSPACK weightTAPS variablesTAPS weightTAPSoutput simpleWeight subsetTAPS multipleImp hotdeckImp wavesTAPS

---

dd   *TAPS imputed demographics*

Description

This is the imputed demographics for the TAPS survey.

Usage

```
dd
```

Format

This is a .rda file

See Also

weightTAPSPACK weightTAPS weightTAPSoutput simpleWeight attritTAPS multipleImp hotdeckImp
subsetTAPS wavesTAPS

---

hotdeckImp	*Impute data using hotdeck imputation method*

Description

Impute missing data in a dataframe using hotdeck imputation method

Usage

```
hotdeckImp(data, outcome, covars)
```

Arguments

data	A data frame with missing values to be imputed
outcome	Character strings indicating outcome variable(s) of the user's interest
covars	Character strings indicating covariates of the user's interest

Value

A data frame with imputed values

Author(s)

David G. Carlson <carlson.david@wustl.edu> Michelle Torres: <smtorres@wustl> Taeyong
Park <t.park@wustl.edu>

See Also

weightTAPSPACK weightTAPS variablesTAPS weightTAPSoutput simpleWeight subsetTAPS
attritTAPS hotdeckImp wavesTAPS

---

multipleImp	*Impute data using multiple imputation method*

Description

Impute missing data in a dataframe using multiple imputation method

Usage

```
multipleImp(data, outcome, covars, m = 5)
```

Arguments

data	A data frame with missing values to be imputed
outcome	Character strings indicating outcome variable(s) of the user's interest
covars	Character strings indicating covariates of the user's interest
m	A nunmeric object indicating number of data sets to create - default is 5

Value

A data frame with imputed values

Author(s)

David G. Carlson <carlson.david@wustl.edu> Michelle Torres: <smtorres@wustl> Taeyong Park <t.park@wustl.edu>

See Also

weightTAPSPACK weightTAPS variablesTAPS weightTAPSoutput simpleWeight subsetTAPS attritTAPS hotdeckImp wavesTAPS

---

pop.margins	*Population margins*

Description

Population margins for the U.S.

Usage

```
pop.margins
```

Format

This is a .rda file

See Also

weightTAPSPACK weightTAPS weightTAPSoutput simpleWeight attritTAPS multipleImp hotdeckImp subsetTAPS wavesTAPS

---

simpleWeight   *Find weights for survey data*

Description

Find weights for survey from a dataframe of observations and population margins. This function performs a raking process based on the population margins specified by the user.

Usage

```
simpleWeight(data, pop.margins, names = c(~agegend, ~ethm, ~educat, ~regmetro,
  ~incomcat, ~ppnet), trunc_at = 5)
```

Arguments

data	A dataframe of observations and values for population categories
pop.margins	A list of known population parameters
names	The names of the population parameters. The dataframe MUST include variables with the same names as the population parameters specified.
trunc_at	A numeric object specifying where to truncate the weights (what should the max weight be?) - default is 5

Value

A vector of weights

Author(s)

David G. Carlson <carlson.david@wustl.edu> Michelle Torres: <smtorres@wustl> Taeyong Park <t.park@wustl.edu>

See Also

weightTAPSPACK weightTAPS variablesTAPS weightTAPSoutput subsetTAPS attritTAPS multipleImp hotdeckImp wavesTAPS

---

subsetTAPS   *Subset TAPS data*

Description

Subset TAPS data by outcome and covariates specified.

Usage

```
subsetTAPS(outcome, covars = NULL, weight = FALSE, refusedasNA = TRUE)
```

Arguments

outcome	A character vector of the names of outcome variables of interest
covars	A character vector of the names of the covariate variables of interest
weight	A logical argument specifying whether to use TAPS base weights or not - default is FALSE
refusedasNA	A logical argument specifying whether to consider the response refused as an NA - default and suggested value is TRUE

Value

A dataframe of subsetted TAPS data

Author(s)

David G. Carlson <carlson.david@wustl.edu> Michelle Torres: <smtorres@wustl> Taeyong Park <t.park@wustl.edu>

See Also

[weightTAPSPACK weightTAPS variablesTAPS weightTAPSoutput simpleWeight attritTAPS multipleImp hotdeckImp wavesTAPS](#)

TAPScum   *TAPS data*

Description

This is the data for the TAPS survey. The codebook can be found at http://taps.wustl.edu/data-archive

Usage

```
TAPScum
```

Format

This is a .rda file

Source

taps.wustl.edu

See Also

[weightTAPSPACK](weightTAPS) [weightTAPS](weightTAPSoutput) [simpleWeight](attritTAPS) [multipleImp](hotdeckImp) [subsetTAPS](wavesTAPS)

TAPSimputeddemographics
TAPS demographics data

Description

This is the data for the demographics of the TAPS survey.

Usage

dd

Format

This is a .rda file and tab delineated

Source

taps.wustl.edu

See Also

[weightTAPSPACK](weightTAPS) [weightTAPS](weightTAPSoutput) [simpleWeight](attritTAPS) [multipleImp](hotdeckImp) [subsetTAPS](wavesTAPS)

---

variablesTAPS	*Lists names of variables by wave*

Description

Lists the names of the variables in the TAPS dataset for the wave(s) specified

Usage

```
variablesTAPS(month, year)
```

Arguments

month
: A character vector specifying the first three letters of the month(s) of the wave(s) of interest

year
: A numeric vector specifying the year(s) of the wave(s) desired. This vector MUST BE the same length of the month vector (one year per month specified)

Value

A list with the names of the variables per wave specified

Author(s)

David G. Carlson <carlson.david@wustl.edu> Michelle Torres: <smtorres@wustl>

See Also

weightTAPS weightTAPSPACK subsetTAPS weightTAPSoutput simpleWeight attritTAPS multipleImp hotdeckImp wavesTAPS

Examples

```
variablesTAPS(month=c("Feb","Mar"), year=c(2012,2012))
```

---

wavesTAPS	*Find population statistics for different waves*

Description

Find the population statistics for each wave in TAPS data to be analyzed. Useful for attrition information.

Usage

```
wavesTAPS(outcome)
```

Arguments

outcome	A character vector of the names of outcome variables of interest

Value

A list of population stats

Author(s)

David G. Carlson <carlson.david@wustl.edu> Michelle Torres: <smtorres@wustl> Taeyong Park <t.park@wustl.edu>

See Also

[weightTAPSPACK](#) [weightTAPS](#) [variablesTAPS](#) [weightTAPSoutput](#) [simpleWeight](#) [attritTAPS](#) [multipleImp](#) [hotdeckImp](#) [subsetTAPS](#)

---

weightTAPS	*Subset TAPS data and find weights*

Description

Subset TAPS data by outcome and covariates, model the attrition rates, impute data for attrited individuals, and find weights for analysis

Usage

```
weightTAPS(interact = TRUE, outcome = NULL, covars = NULL,
  weight = FALSE, refusedasNA = TRUE, method = "multi",
  na.delete = TRUE, m = 5, pop.base = 1, trunc_at = 5,
  stringsAsFactors = TRUE)
```

Arguments

interact	A logical vector indicating if the function is to be run interactively. If TRUE arguments are not needed - default is TRUE
outcome	A character vector of the names of outcome variables of interest. It is highly suggested that the outcome variables be entered starting with the earliest wave.
covars	A character vector of the names of the covariate variables of interest
weight	A logical argument specifying whether to use TAPS base weights or not - default is FALSE
refusedasNA	A logical argument specifying whether to conisder the response 'Refused' as a missing value - default and suggested value is TRUE
method	A character object indicating type of imputation to be used. hotdeck for hotdeck imputation, multi (default) for mulitple imputation, none for no imputation

na.delete   A logical argument specifying whether to eliminate rows with NAs before cal-
                   culating weights if method chosen is none - default is TRUE. Only set to FALSE
                   if planning to use NA observations.

m   A numeric argument specifying number of imputed data sets to produce if using
                   multiple imputation - default is 5

pop.base   A numeric object specifying which CPS data to use as a baseline. 1 is Dec.
                   2011, 2 is Dec. 2012, 3 is Dec. 2013. Default is 1.

trunc_at   A numeric object specifying where to truncate the weights (what should the max
                   weight be?) - default is 5

stringsAsFactors

                   A logical vector indicating whether non-numeric variables should be factors
                   rather than strings - default is TRUE

Details

This package is meant to subset The American Panel Survey (TAPS) data by outcome and by
covariate variables of interest through the function weightTAPS. The subsetting process accounts
for respondents attriting from at least one of the waves under analysis, as well as for outcome
non-response. The variables of interest must be entered exactly as named in the TAPS dataframe.
See http://taps.wustl.edu/data-archive or use the variablesTAPS function to explore the
names of the variables by wave. It is important to revise the particular features of each of the
variables of interest.

It is strongly suggested that the outcome variables be entered starting with the earliest wave for eas-
ier interpretation of the attrition rates. Other arguments are listed in the help file of the weightTAPS()
function, and must be considered based on the user's needs. The function can be run in interactive
mode by simply running weightTAPS(). The user must answer the questions based on her needs.

The function weightTAPS() should be assigned to an object in order to conduct the analysis of
TAPS. weightTAPS() returns a subset of the complete TAPS dataset that includes only the outcome
variable and covariates specified by the user, a set of standard demographics and a new variable with
the corresponding weight for each respondent.

It also retains the respondents that gave an answer to the outcome variable of interest through the
waves specified by the user. Respondents that attritted or did not provide an answer to the outcome
variable for any of the waves under analysis are removed from the subset data. Missing values in
sociodemographic variables are imputed for the respondents that remain in the sample, in order
to compute their proper weight. The missing values observed in the covariates of the remaining
respondents are imputed through the method selected by the user. Once the TAPS data is subset,
the function calculates weights based on the demographic group membership of the respondents in
the final subset. These weights will be appended to the end of the data frame(s) with the column
name new.weights.

The output (see weightTAPSoutput) is of class weightTAPSoutput. This class implies the existence
of certain slots that save useful information. These slots are df, attrit and stats.

The slot df contains the dataframe(s) that represent the final subset data. The final subset data keeps
only the outcome and covariates of interest specified as well as a set of demographics and the new
dynamic weights. It also accounts for non-response in the outcome variable and attrition across
waves through the waves specified. Respondents with missing values in the outcome variable for
any of the waves desired are removed from the final dataframe.

The missing covariate data can be left as is, by specifying method="none". If imputation is desired, the argument method can be set to 'multi' for multiple imputation, or 'hotdeck' for hotdeck imputation. If multiple imputation is done, the argument m should be set to the number of imputed dataframes to be created. Depending on the imputation method selected, df can be a list of m elements or a list containing a single element. Each element of df stores a dataframe. If method="multi" is specified, df contains a list of m dataframes.

To access the dataframes, use getdf(objectname). Objectname corresponds to the object where the value of weightTAPS() was originally stored. If hotdeck or no imputation was used, the final dataset is the first element of the df list, and can be accessed with getdf(objectname)[[1]].

The slot attrit is a list of attrition rates from the first wave specified in the outcome argument. Each quantity represents the percentage of people (by demographic group) that attritted TAPS through the waves specified. It compares the initial composition of each demographic group (from the oldest wave specified) to the composition of the same demographic group in the final subset data delivered by weightTAPS(). Large values, particularly large values relative to other values in the same sociodemographic category, indicate high rates of attrition. It is important to highlight that high rates of attrition may cause problems in data analysis. The slot stats lists each sociodemographic group's share of the overall population as represented in the final sample for each outcome.

The information contained in both the attrit and stats slots can be graphically illustrated using the plot(objectname) function. Two different types of plots are displayed after running the plot function: a dot chart and a set of trend plots. The dot chart shows the differences between the sociodemographic composition of the sample in the first wave specified and the final subset dataframe. This information is disaggregated by the following sociodemographic groups: Age and Gender, Ethnicity, Education, Income, Region and Metropolitan status, and Internet use. The trend plots presented illustrate the changing composition of the sample by demographic group across the waves specified. The lines shown in each plot correspond to the different categories within each of the groups mentioned. The lines show the percentage of the final subset data belonging to each category by wave. The plots aim to show the variation in the composition of the sociodemographic groups through the waves specified.

Value

An object of class weightTAPSoutput with the following slots:

- df List of dataframes (or single dataframe) of subsetted TAPS data with imputed data for missing values in covariates, with column for weights
- attrit A list of vectors of attrition rates for populations
- stats A list of population statistics from wave to wave

Author(s)

David Carlson <carlson.david@wustl.edu>, Michelle Torres <smtorres@wustl>,and Taeyong Park <t.park@wustl.edu>

See Also

weightTAPSPACK variablesTAPS subsetTAPS weightTAPSoutput simpleWeight attritTAPS multipleImp hotdeckImp wavesTAPS

Examples

```
myOutcome <- c("APPRCONGS2","APPRCONGS6")
myCovars <- c("POLKNOW3S2","POLKNOW6S2")
test<-weightTAPS(interact=FALSE,outcome=myOutcome,
    covars=myCovars,weight=FALSE,refusedasNA=TRUE,
    method="hotdeck",m=5,pop.base=1,trunc_at=5,stringsAsFactors=TRUE)
```

---

weightTAPSoutput-class

An object outputted by weightTAPS function

Description

Objects of class `weightTAPSoutput` are created by the `weightTAPS` function

Details

An object of the class 'weightTAPSoutput' has the following slots:

- `df` List of dataframes of subsetted TAPS data with imputed data for non-response in covariates, with column for weights - if method none chosen, of class data.frame
- `attrit` A list of vectors of attrition rates by sociodemographic groups
- `stats` A list of population statistics from wave to wave

The slot `df` contains the dataframe(s) that represent the final subset data. The final subset data keeps only the outcome and covariates of interest specified. It also accounts for non-response in the outcome variable and attrition. Respondents with missing values in the outcome variable for any of the waves desired are removed from the final dataframe. Missing values in sociodemographic variables are imputed for the respondents that remain in the sample, in order to compute their proper weight. The missing covariate data can be left as is, by specifying `method="none"`. If imputation is desired, the argument `method` can be set to 'multi' for multiple imputation, or 'hotdeck' for hotdeck imputation. If multiple imputation is done, the argument `m` should be set to the number of imputed dataframes to be created.

The slot `attrit` is a list of attrition rates from the first wave specified in the outcome argument. Each quantity represents the percentage of people (by demographic group) that attritted TAPS through the waves specified. It compares the initial composition of each demographic group (from the oldest wave specified) to the composition of the same demographic group in the final subset data delivered by `weightTAPS()`. Large values, particularly large values relative to other values in the same sociodemographic category, indicate high rates of attrition.

The slot `stats` lists each sociodemographic group's share of the overall population as represented in the final sample for each outcome. The information contained in both the `attrit` and `stats` slots can be graphically illustrated using the `plot(objectname)` function. Two different types of plots are displayed after running the plot function: a dot chart and a set of trend plots. The dot chart shows the differences between the sociodemographic composition of the sample in the first wave specified and the final subset dataframe. This information is disaggregated by the following

sociodemographic groups: Age and Gender, Ethnicity, Education, Income, Region and Metropolitan status, and Internet use. The trend plots presented illustrate the changing composition of the sample by demographic group across the waves specified. The lines shown in each plot correspond to the different categories within each of the groups mentioned. The lines show the percentage of the final subset data belonging to each category by wave. The plots aim to show the variation in the composition of the sociodemographic groups through the waves specified.

`print(objectname)` will show a summary of the attrition rates. `show(objectname)` will run the `print` function.

Author(s)

David Carlson <carlson.david@wustl.edu>, Michelle Torres <smtorres@wustl>, and Taeyong Park <t.park@wustl.edu>

See Also

weightTAPSPACK weightTAPS variablesTAPS subsetTAPS wavesTAPS simpleWeight attritTAPS multipleImp hotdeckImp wavesTAPS

---

weightTAPSPACK	*weightTAPSPACK*

Description

The weightTAPSPACK subsets The American Panel Survey (TAPS) data by outcome and covariates, models the attrition rates, imputes data for attrited individuals, and finds weights for analysis.

Details

This package is meant to subset The American Panel Survey (TAPS) data by outcome and by covariate variables of interest through the function weightTAPS. The subsetting process accounts for respondents attriting from at least one of the waves under analysis, as well as for outcome non-response. The variables of interest must be entered exactly as named in the TAPS dataframe. See http://taps.wustl.edu/data-archive or use the variablesTAPS function to explore the names of the variables by wave. It is important to revise the particular features of each of the variables of interest.

It is strongly suggested that the outcome variables be entered starting with the earliest wave for easier interpretation of the attrition rates. Other arguments are listed in the help file of the weightTAPS() function, and must be considered based on the user's needs. The function can be run in interactive mode by simply running weightTAPS(). The user must answer the questions based on her needs.

The function weightTAPS() should be assigned to an object in order to conduct the analysis of TAPS. weightTAPS() returns a subset of the complete TAPS dataset that includes only the outcome variable and covariates specified by the user, a set of standard demographics and a new variable with the corresponding weight for each respondent.

It also retains the respondents that gave an answer to the outcome variable of interest through the waves specified by the user. Respondents that attrited or did not provide an answer to the outcome

variable for any of the waves under analysis are removed from the subset data. Missing values in sociodemographic variables are imputed for the respondents that remain in the sample, in order to compute their proper weight. The missing values observed in the covariates of the remaining respondents are imputed through the method selected by the user. Once the TAPS data is subset, the function calculates weights based on the demographic group membership of the respondents in the final subset. These weights will be appended to the end of the data frame(s) with the column name `new.weights`.

The output (see `weightTAPSoutput`) is of class weightTAPSoutput. This class implies the existence of certain slots that save useful information. These slots are `df`, `attrit` and `stats`.

The slot `df` contains the dataframe(s) that represent the final subset data. The final subset data keeps only the outcome and covariates of interest specified as well as a set of demographics and the new dynamic weights. It also accounts for non-response in the outcome variable and attrition across waves through the waves specified. Respondents with missing values in the outcome variable for any of the waves desired are removed from the final dataframe.

The missing covariate data can be left as is, by specifying `method="none"`. If imputation is desired, the argument `method` can be set to 'multi' for multiple imputation, or 'hotdeck' for hotdeck imputation. If multiple imputation is done, the argument `m` should be set to the number of imputed dataframes to be created. Depending on the imputation method selected, `df` can be a list of `m` elements or a list containing a single element. Each element of `df` stores a dataframe. If `method="multi"` is specified, `df` contains a list of `m` dataframes.

To access the dataframes, use `getdf(objectname)`. Objectname corresponds to the object where the value of `weightTAPS()` was originally stored. If hotdeck or no imputation was used, the final dataset is the first element of the `df` list, and can be accessed with `getdf(objectname)[[1]]`.

The slot `attrit` is a list of attrition rates from the first wave specified in the outcome argument. Each quantity represents the percentage of people (by demographic group) that attritted TAPS through the waves specified. It compares the initial composition of each demographic group (from the oldest wave specified) to the composition of the same demographic group in the final subset data delivered by `weightTAPS()`. Large values, particularly large values relative to other values in the same sociodemographic category, indicate high rates of attrition. It is important to highlight that high rates of attrition may cause problems in data analysis. The slot `stats` lists each sociodemographic group's share of the overall population as represented in the final sample for each outcome.

The information contained in both the `attrit` and `stats` slots can be graphically illustrated using the `plot(objectname)` function. Two different types of plots are displayed after running the plot function: a dot chart and a set of trend plots. The dot chart shows the differences between the sociodemographic composition of the sample in the first wave specified and the final subset dataframe. This information is disaggregated by the following sociodemographic groups: Age and Gender, Ethnicity, Education, Income, Region and Metropolitan status, and Internet use. The trend plots presented illustrate the changing composition of the sample by demographic group across the waves specified. The lines shown in each plot correspond to the different categories within each of the groups mentioned. The lines show the percentage of the final subset data belonging to each category by wave. The plots aim to show the variation in the composition of the sociodemographic groups through the waves specified.

Author(s)

David Carlson <carlson.david@wustl.edu>, Michelle Torres <smtorres@wustl>, and Taeyong Park <t.park@wustl.edu>

See Also

weightTAPS variablesTAPS subsetTAPS weightTAPSoutput simpleWeight attritTAPS multipleImp hotdeckImp wavesTAPS

Index