R documentation

of all in
'/home/dorer/projects/census/PovertyAssessmentToolkit/source/PAT/man'

April 28, 2024

R topics documented:

`Builtin.datasets` *PUMA (2012 & 2022) to Tract, Blockgroup, Block (2020) Relationship Files*

Description

PUMA to Trace/Blockgroup/Block Correlation/Relationship Correspondence Datasets

Usage

```
data(PUMA.20.Tract)
data(PUMA.20.BlockGroup)
data(Tract.25.2020.Block.2020)
data(PUMA.12.Tract)
data(PUMA.12.BlockGroup)
```

Format

245590 observations (2012 PUMAs)
or 240942 observations (2022 PUMAs)

Value

A data.frame

Variable names

1. **State** State FIPS code.
2. **Puma** PUMA FIPS code.
3. **Place** Place FIPS code.
4. **County** County FIPS code.
5. **CSD** CSD (cosub) FIPS code.
6. **Tract** Tract FIPS code.
7. **BlockGroup** Blockgroup FIPS code.
8. **Block** Block FIPS code.
9. **StateAb** State Postal Abbrevation.
10. **CountyName** County name.
11. **PumaName** PUMA name.
12. **PlaceName** Place name.
13. **CSDName** CSD name.
14. **AllocCSD** CSD Geocorr Allocation Factor.
15. **AllocCSD2** CSD Geocorr Allocation Factor-2.
16. **AllocPlace** Place Geocorr Allocation Factor.
17. **AllocPlace2** Place Geocorr Allocation Factor-2.
18. **AllocBlock** Block Geocorr Allocation Factor.
19. **AllocBlock2** Block Geocorr Allocation Factor-2.
20. **TotalPop** Total Population.

References

Missouri Census Data Center Geocorr Engine.
https://mcdc.missouri.edu/applications/geocorr2022.html

`Builtin.marginals` *Builtin Marginal Tables for Marginal Adjustment*

Value

Marginal table function (ends in .m) to be used in the marginal.tables component of a model.

List of tables

AgeRaceSex.m	Age by Race by Sex Age given by Age.code Race by Race.code
AgeSex.m	Age by Sex Age given by Age.code Race by Race.code
Race.m	Number of categories given by PAT.race.code
HispanicAgeSex.m	Hispanic yes/no Age given by PAT.age.code
Hispanic.m	Works on Blockgroups
Sex.m	
Poverty2.m	Poverty 2 categories: Below/Above/Undefined
Poverty3.m	Poverty 3 categories: Below/100_299/300_and_over/Undefined
PovertyRaceSex.m	Poverty 2 categories: Below/Above/Undefined Race given by PAT.race.code()
PovertyAgeSex.m	Poverty 2 categories: Below/Above/Undefined Age age code: "9b"
Poverty4.m	Poverty 4 categories Under 100_200, 200_300, 300+, Undefined
Employed.m	Employed, Unemployed, Not_in Under_16
HouseType.m	Non_inst, Inst (Institutional), House (Household)
DisabilityRace.m	Disability Yes, No, Other, Race give by Race.code
DisabilityAgeSex.m	Disability Yes, No, Other, Race give by Race.code
Education.m	Under_18, Less_than_high_school, High_school Some_College, College_degree

Education4.m	Less_than_high_school, High_school, Some_college, College_degree
TenurePerson.m	Rent, Own, Group_quarters. Person table
TenureHouse.m	Rent, Own. Household table
TenureUnits.m	Rent, Own. Household table
TenureRooms.m	Rent, Own. by Number of Rooms (House table)
TenureBedrooms.m	Rent, Own. by Number of Bedrooms (House table)
TenureFuel.m	Rent/Own by Fuel type (House table)
HeatingFuel.m	Heating Fuel 9 categories.
YearBuilt.m	House table
MaritalStatus3.m	Married, Single Mother, Single Father.
MaritalStatus3a.m	Married, Single, Under_15_years
MaritalStatus7.m	Married, Widowed, Separated, Divorced Never_married, Other, Under_15_years.
HealthIns3.m	Insured, Uninsured, Military_inst (Military or Institutional Group Quarters
FamilyType3.m	Married_head, Female_head, Male_head, Other

Author(s)

David Dorer 11 Jan 2024

`Builtin.models` *Builtin Models*

Models

1. **BrooklineI.model**
   Marginal Tables: AgeRaceSex, MaritalStatus3, HispanicAgeSex, Education4, Employed, Tenure, DisabilityRace, Poverty2, HouseType, FamilyType3, HealthIns3
   Extra Variables: SPM3 (Supplemental Poverty Measure)
   Parameters: geotype="tract", model.type="person", nages="7a", nraces="5"

2. **BrooklineIII.model**

   Marginal Tables: AgeRaceSex, MaritalStatus7, DisabilityRace ,Poverty2, Tenure, Employed, HouseType, FamilyType3

   Extra Variables: SNAP, SPM3 (Supplemental Poverty Measure)

   Parameters: geotype="tract", nages="9", nraces="5", model.type="person"

3. **BrooklineIIIb.model**

   Marginal Tables: AgeSex.m, Race.m, Poverty2.m, Hispanic.m, HouseType.m, Employed.m, Tenure.m, MaritalStatus7.m

   Extra Variables: SNAP, SPM3 (Supplemental Poverty Measure)

   Parameters: geotype="blockgroup", nages="9", nraces="5", model.type="person"

4. **PennsylvaniaI.model**

   Marginal Tables: AgeRaceSex, Poverty4, Employed, Education4, HispanicAgeSex, MaritalStatus3, Tenure.

   Extra Variables: WIC, Age6a, EmployedHouse18.

   Parameters: nages="9", nraces="5", geotype="tract", model.type="person".

5. **NewYorkCityI.model**

   Variables: Age, Sex, Race, Education, Employed, Poverty2, Hispanic.

   Marginal Tables: AgeRaceSex, HispanicAgeSex, Education, PovertyRaceSex, Employed.

   Parameters: geotype="tract", nages="9", nraces="5".

6. **TestI.model**

   Extra Variables: SPM3

   Parameters: nages= "9", nraces= "5", geotype="person", model.type="person".

7. **LanguageI.model**

   Variables: Language, English, TenurePerson (own/rent), Age, Sex, Race, Education, Hispanic, Poverty2, Employed.

   Marginal Tables: AgeRaceSex, HispanicAgeSex, PovertyRaceSex, Education, TenurePerson

   Parameters: geotype="tract", language.codes=NULL, model.type="person", nages="9a", nraces="5".

   Optional Parameter: Language.codes(passed to Language.v) only language codes (See LANP in PUMS codebook) in listed in language codes will be used. Other codes will be set to "Missing".

Author(s)

David Dorer 28 Apr 2024 09:05

`Builtin.variables` *Builtin Derived PUMS variables*

Value

PUMS variable function (ends in .v) to be used in the variables component of a model.

List of variables

Note All household variables are also available for person level models due to matching the PUMS person dataset with the PUMS household data set using SERIALNO.

Age.v	Age arguments AGEP and nages
WIC.v	Does person live in a household that receives WIC benefits.
Race.v	Race with arguments RAC1P and nraces
SNAP.v	Does person live in a household that receives SNAP benefits.
Hispanic.v	Is the person of Hispanic Ethnicity ("Hispanic" or "No")
Employed.v	(PUMS ESR) Is the person employed ("Under_16","Employed","Unemployed","No
EmployedHouse18.v	Are all adults in household over 18 employed?
TenurePerson.v	"Group_quarters","Own","Rent".
HousingCost.v	"Low", "High", "Vacant".
Poverty2.v	(PUMS POVPIP) Poverty Threshold "Below", "Above", "Undefined".
Poverty4.v	Poverty Threshold "Below", "100_to_199","200_to_299", "Over_300", "Undefined".
HealthIns3.v	"Insured", "Uninsured", "Military_inst" (Military or Institutional Group Quarters).
Disability.v	"Yes", "No", "Other" (See table B18101)
HouseType.v	Non_inst, Inst (Institutional), House (Household)
Education4.v	(PUMS SCHL) "Less_than_high_school", "High_school", "Some_college", "College_degree"
FamilyType3.v	(PUMS HHT) "Married_head", "Female_head", "Male_head", "Other".
FamilyType.v	"Married_head", "Female_head", "Male_head", "Child_group", "18_years_and_over".
MaritalStatus3a.v	Married, Single, Under_15_years
MaritalStatus7.v	Married, Widowed, Separated, Divorced
Language.v	PUMS variable LANP languages.
English.v	Under_5_years, Very_well, Well, Not_well, Not_at_all
TravelWork.v	Not_working, Other, Bicycle, Walk, Work_home, Car_truck_van, Public
Relationship.v	see PUMS variabel RELSHIPP. Never_married, Other, Under_15_years.

House Level Variables	
TenureHouse.v	Own, Rent, GQ_vacant.
HeatingFuel.v	(PUMS HFL) "Gas","LP","Electric","Oil","Coal" ,"Wood","Solar","Other","None"
Units.v	Units in Structure (PUMS BLD) "One_attached","One_detached" "Two","3_or_4","5_to_9","10_or_more","Other"

Author(s)

David Dorer 28 Apr 2024 11:33

`GettingStarted` *Getting Started - Installation and Setup*

Description

Downloading Installing and Initializing PAT

Downloading

1a. (Linux/Ubuntu) Download tar.gz file for current version of PAT e.g. PAT_0.1.tar.gz

1b. (Windows) Download zip file for current version of PAT e.g. PAT_0.1.zip

Setup outside of R session.

1. Decide on and create the folder where you want to store files.

2. Create sub-folders/directories:

2a. logs
2b. data
2c. output

Installation Within R

In an R session run (one time)

Linux: `install.packages(pkgs="path_to_PAT_<vers>.tar.gz",repos=NULL);`

or
Windows: `install.packages(pkgs="path_to_PAT_<vers>.zip>",repos=NULL);`

Setup within R session

1. every time you start R run

```
require(PAT)
```

2. Every time set your census key PAT.census.key("your_key_here")
You can get your census key at
Census key request page https://api.census.gov/data/key_signup.html

3. Every time set your root folder/directory
`PAT.root("path_to_base_folder");`
The default for the root folder is your working directory: `getwd();`
which is where you were when you started R (probably wrong choice)

4. Optionally set nraces (number of race categories) for PUMS
derived variables and marginal Detail/Subject/DataProfile tables.

4a. To see current value use: `PAT.race.code()`

4b. To change: `PAT.race.code(<new_value>);`

4c. To see choices: `help(Pat.race.code);`

5. Optionally set nages (number of ages categories) for PUMS
derived variables and Detail/Subject/DataProfile tables.

5a. To see current value use: `PAT.age.code();`

5b. To change: `PAT.age.code("new_value");`

5c. To see choices: `help(Pat.age.code);`

6a. Set/check PUMS vintage (every time)
To see current value: `PAT.pums.vintage();`
To set: `PAT.pums.vintage(<new_value>);`

6b. Set/check PUMS period (every time)
To see current value: `PAT.pums.period();`
To set: `PAT.pums.period("new_value");` (valid values 1 and 5)

7a. Set/check vintage for marginal tables, usually the same as PUMS vintage (every time).
To see current value: `PAT.vintage()`
To set: `PAT.vintage(<new_value>);`

7b. Set/check period for marginal tables, usually 5 (every time). To see current value: `PAT.period()`
To set: `PAT.period(<new_value>)`

NOTE: There are many fewer ACS tables for the 1 year period.
Many of your marginal tables may fail with a 1 year period.

8. Check/Set other options:

```
PAT.vintage()
PAT.period()
PAT.pums.vintage()
PAT.pums.period()
PAT.verbose()
```

9. `help(package="PAT")`

Author(s)

David Dorer 16 Dec 2023 10:09

`loglinf3` *Loglinear Model fit with specified marginal tables*

Description

Wrapper R function for compiled C IPF subroutine.

Usage

```
loglin3f(seed.table, target.list, target.data, niter=5, maxdev=0.001, debug=0)
```

Arguments

`seed`	Starting seed table for fit.
`target.list`	list with index specifing variable is seed table. Note the marginal.tables function computes this list from the model.
`target.data`	list containing the data for the marginal tables specified by target.list
`niter`	Maximum number of iterations. Same as iter in the PAT.synth.data function.
`maxdev`	Maximum relative change in difference between tarage marginals and current fit margian between successive iteration.
`debug`	Verbose/debug level for messages. Default PAT.verbose()

Details

The arguments are the same as the Ipf function in mipfp package.

Value

`seed`	Argument passed to function.
`p.hat`	Model fit scaled to sum to 1.
`nlast`	Last iteration.
`ifault`	Did the fit converge 0: yes 1: no
`maxdev`	Maximum deviation at last iteration.

Author(s)

David Dorer

References

https://cran.r-project.org/web/packages/mipfp/mipfp.pdf https://www.jstatsoft.org/article/view/v086c02

`PAT.acs.table` *Download an ACS Detail (B, C, S, or DP, P, PCT) Table*

Description

Download and ACS or Decennial table for a given geography.

Usage

```
PAT.acs.table()
```

Arguments

`group`	ACS table name or "group" (API term). For example the table B01001 has title "Sex by Age." Check out tables at data.census.gov.
`state`	State FIPS code. Required.
`county`	County FIPS code. Character variable - optional defaults to "".
`tract`	County FIPS code. Character variable (6 characters) - optional defaults to ""
`blockgroup`	Blockgroup FIPS code. Character variable (1 character) defaults to "" (ignored)
`puma`	PUMA FIPS code. 5 characters - optional defaults to ""
`place`	Place FIPS code. Defaults to "" which will be ignored.
`csd`	County Subdivision FIPS code - defaults to "" which will be ignored
`vintage`	Defaults to PAT.vintage().
`period`	Defaults to PAT.period().
`ddir`	character variable indicating where to store downloaded tables - default datadir().
`debug`	Level of messages to print - default PAT.verbose().
`cache.metadata`	Cache level for metadata 0: no caching, download from Census website for every table. 1: cache metadata in file 2: cache metadata in both a file and in the .GlobalEnv or computer memory. Default 2 speeds downloading and decoding many tables. If you seem to be having difficulties with stale cached data use 0. Default PAT.cache.metadata().
`cache.tables`	Cache level for ACS tables. 0: download table from census for each table. 1: cache tables in a file (best). 2: cache table in both a file and in computer memory. With a run of many tracts 2 will quickly exhaust memory. For testing where you are repeatedly downloading the same table you might temporarily use 2 to save time.

Details

Download B, C, S, DP (SF1) P (DCH), and PCT (DCH) tables attaches names for rows using metadata, table has 2 columns: "Est" and "MoE". Various geographies can be specified. The function sorts out the geography if you use too many geography values.

Value

list	list with 2 components "data" (matrix with table rownames and 2 columns "Est" and "MoE") and "par" various parameters such as the table "object" name, folder where the table was stored, name of the filename of where the table was stored, date time of download, etc.

Author(s)

David Dorer 28 Apr 2024 13:15

References

To find, examine and check data, see Census webpage data.census.gov

---

PAT.age.race *Set age and race categories*

Description

PAT.age.code(code) Set the PAT.age.code option.
PAT.race.code(code) Set the PAT.race code option.

Usage

```
PAT.age.code("7a")
PAT.race.code("5")
```

Arguments

code	Age code
code	Race code

With no argument return the Age or Race Code

Details

Recode age catgories for B01001 or B01001A etc

Value

The race or age code.
or
Sets the corresponding Global race or age code option.

Values for age code

"7a"	7 age categories 14 and under, 15 to 19, 20 to 34, 35 to 44, 45 to 54, 55 to 64, 65 and over.
"7b"	7 age categories under 5, 5 to 9, 15 to 17, 18 to 24, 25 to 64, 65 and over
"9"	xx
"6"	xx
"6a"	xx
...	use RecodeAges(doc=1) to see possible values. then RecodeAges(nages="9",doc=1) to see recoding.

Values for race code

"3"	3 race categories: White, Black, Other
"5"	5 race categories White, Black, Asian, Other, Two.
"7"	7 race categories white, black, native, asian, Hawaiian, Other, Two.

Author(s)

David Dorer 26 Dec 2023 13:42

PAT.cache.metadata *Set Cache Level For Table Metadata.*

Description

Set metadata cache level.

Usage

```
PAT.cache.metadata(2)
```

Arguments

value	Value for metadata cache level. No value given return level. Values 0, 1, 2.

Details

Sets global PAT option "PAT.cache.metadata"

Value

`level`	Current value of PAT.cache.metadata.
`0`	Download metadata for every ACS table. Do not save in a file or in .GlobalEnv
`1`	Save metadata in file for later use. Better.
`2`	Save metadata in file and in .GlobalEnv. Speeds up table rowname decoding. Best for many ACS tables.

Author(s)

David Dorer 19 Dec 2023 11:38

`PAT.cache.tables` *Set Table Cache Level*

Description

Set PAT.cache.tables global PAT option.

Description

Set ACS table cache level.

Usage

```
PAT.cache.tables(1). Set option PAT.cache.tables to 1. With no
argument return current value of option "PAT.cache.tables".
```

Arguments

`value`	Value for table cache level. No argument return table cache level.

Details

Sets global PAT option "PAT.cache.tables"

Value

`level`	Current value table cache level.
`0`	Download ACS table every time. Do not save in a file or in .GlobalEnv
`1`	Save ACS table in file for later use. Best.
`2`	Save ACS table in file and in .GlobalEnv. Quickly uses up memory if a large number of tables are downloaded.

Author(s)

David Dorer 19 Dec 2023 11:38

PAT.check.csd *Validate model for a County Subdivision*

Description

Aggregate Tract or Blockgroup synthetic data for a County Subdivision (City/Town) and check tabulations against ACS detail tables. Runs PAT.synth.data for CSD tracts, PAT.merge.synth.tracts, then PAT.check.synth.model.

Arguments

state	State FIPS code.
county	County FIPS code.
csd	CSD FIPS code.
outtag	character string/tag to use in output files (default "_test").
model	name of model (character without the .model)
odir	output directory/folder for synthetic data passed to PAT.synth.data default outdir()<model><outtag>
vintage	marginal table vintage. Passed to PAT.synth.data. Default 2022
period	marginal table period. Default 5
download	Download files ? Default PAT.download().
synth	0: skip PAT.synth.data step. 1: synthesize data.
merge	0: skip merge step PAT.merge.synth.tracts and go directly to PAT.test.model step
iter	number of IPF iterations. Passed to PAT.synth.data. Default 30.
add	Add small amount to model/seed table Passed to PAT.synth.data. Default 0.0001
minwt	Keep synthetic records with weight >minwt. Passed to PAT.synth.data. Default 0.001
pums.period	Default 1 (may be changed when 2022 5 year PUMS data comes out).
pums.vintage	Vintage for PUMS data. Default 2022
type	"tract" or "blockgroup" if type=="blockgroup" then run PAT.synth.blockgroups step (not implemented yet)

Value

return from last PAT.check.synth.model step.

Author(s)

David J Dorer

References

PAT.synth.data, PAT.merge.synth.tracts, PAT.check.synth.model.

PAT.check.synth.model *Check fit for synthetic data*

Arguments

state	State FIPS Code
county	County FIPS Code
csd	County Subdivision FIPS code (optional)
tract	character vector of tract FIPS codes (6 characters)
model	character string model name
puma	PUMA FIPS codes
place	Place FIPS code (optional)
vintage	Marginal tables vintage (default 2022)
period	Marginal tables period (default 5)
inptag	character string for input file reads paste0(datadir()".synth_data_tract_all_<state>,<model><inptag>.csv") (PAT.synth.data output file)
download	default PAT.download()
debug	Default PAT.verbose()
digits	number of rounding digits for output table diagnostics

Author(s)

David J Dorer

References

PAT.synth.data

PAT.convert.spm *Convert Supplemental Povert Measure sas7bdat file to .RData format and subset on state.*

Description

Convert downloaded SAS data file to R .RData format.
Subset National SPM for a state.

Usage

```
PAT.convert.spm(vintage=2022,ddir=datadir(),dowload=0)\\{}
PAT.make.spm.state(state,ddir=datdir(),vintage=PAT.spm.vintage());
```

Arguments

vintage	SPM vintage.
state	State FIPS for subset.
ddir	Data directory to read/store output.
download	Should SAS file be downloaded from Census website.

Details

PAT.convert.spm() Convert SPM SAS sas7bdat to R RData format.
PAT.make.spm.state() Subset SPM RData dataset by a state.
Dataset can be downloaded by hand and then converted.

Value

An R .RData file named SPM.<vintage>.RData with R object SPM.<vintage>.
An R .RData file named SPM.<vintage>.<state>.RData with R object SPM.<vintage>.<state>

Author(s)

David Dorer 28 Apr 2024 14:28

References

Datasets Census webpage.

PAT.merge.synth.blockgroups

Merge Synthetic Blockgroup Files

Description

Zip individual blockgroup files and merge into a single csv file.

Usage

```
PAT.merge.synth.blockgroups(state="25",ddir=datadir(),odir=outdir(),zip=TRUE,outtag="_test")
```

Arguments

state	State FIPS code.
ddir	Data directory for output zip and output csv file.
odir	Directory with input csv tract files.
zip	1: Zip tract files before merging. 0: Use existing zip file.
minwt	Only output records to the csv file with weights >= minwt. Total population of subset is scaled to agree with input data. Default 0.0
outtag	Tag to append to zip file and output file names.

Details

xxx

Value

zip file	Output zip file synth_data_<state><outtag>.zip contains individual tract csv files.
csv file	Merged csv data file synth_data_blockgroup_all_<state><outtag>.csv contains stacked individual tract csv files.

Author(s)

David Dorer 15 Dec 2023 14:54

PAT.merge.synth.tracts

Merge Synthetic Tract Files

Description

Zip individual tract files and merge into a single csv file.

Usage

```
PAT.merge.synth.tracts(state="25",ddir=datadir(),odir=outdir(),zip=TRUE,outtag="_test")
```

Arguments

state	State FIPS code.
ddir	Data directory for output zip and output csv file.
odir	Directory with input csv tract files.
zip	1: create zip file with tract data. 0: Merge files in existing zip file.
minwt	Only output records to the csv file with weights >= minwt. Total population of subset is scaled to agree with input data.
outtag	Tag to append to zip file name and merged csv file name.

Details

Merge tract synthetic data files in odir output: ddir/synth_data_tract_all_<state><outtag<.csv and zip them into synth_tract_<state><outtag>.zip

Value

zip file	Output zip file synth_data_<state><outtag>.zip contains individual tract csv files.
csv file	Merged csv data file synth_data_tract_all_<state><outtag>.csv contains stacked individual tract csv files.

Author(s)

David Dorer 22 Dec 2023 09:57

PAT.model *Statistical Models for Synthetic Data*

Description

A model with PUMS variables and marginal tables.

Usage

```
See BrooklineI.model and PennsylvaniaI.model for examples
```

List Elements

name	character variable with name for model, e.g. "BrooklineI"
variables	A named list with functions that define the variables.
marginal.tables	A named list with functions that define the marginal tables. See PAT.acs.table for required arguments.
parameters	A named list with model parameters:
model	name of model. Same as name component.
nages	the number of age categories for model (see PAT.nage.codes)
nraces	the number of race categories for model (see PAT.nrace.codes)
geotype	type of geography "tract" or "blockgroup".
comment	comment (any text).
...	optional user defined parameters that can be passed to the model variables and marginal.tables functions.

Author(s)

David Dorer

PAT.pums.data *Download PUMS Data*

Description

Download PUMS data for a PUMA and create drived variables.

Usage

```
PAT.pums.data(state="25",puma="03301",model="BrooklineIII")
```

Arguments

state	State FIPS code.
puma	PUMA FIPS code.
model	A model or the name of a model. Default NULL i.e. use variables to select variables
variables	Ignored in model is given. A character vector of variable names to download. May include defined derived variables. Ignored if model is given.
parameters	Named list of parameters for derived variables. Same as model component "parameters." Ignored if model is given.
vintage	PUMS vintage. Default PAT.pums.vintage().
period	PUMS period. Default PAT.pums.period().
key	Census key. Default PAT.census.key().
debug	Level of messages to print. Default PAT.verbose().

Details

If SPM variables are used then period is set to 1.

Value

A list with:

person	Person level data.frame. Note person data.frame includes person levels for house variables.
house	House level data.frame. Note house data.frame is person data frame subset on SPORDER==1.
par	parameters from model (if model is used) or parameters argument.

Author(s)

David Dorer 11 Dec 2023 13:46

References

https://www.census.gov/programs-surveys/acs/microdata/documentation.html https://www.census.gov/data/datasets/tim series/demo/supplemental-poverty-measure/acs-research-files.html

---

PAT.reptable *Tabulations for Replicate Weights Designs*

Description

Creates table with an Estimate (Est) and Margin of Error (MoE) last dimension.

Arguments

form	Formula that defines cross-tabulation
design	svrepdesign survey object. See survey package.

Value

multiway tabulation with with a last dimension stat with "Est" and "MoE".

Author(s)

David Dorer 28 Apr 2024 17:51

Examples

```
\code{my.design<-svrepdesign(id=~1,data,weights,repweights)}
\code{age.sex.table<-PAT.reptable(~Age+Sex,design=my.design)}
```

`PAT.root` *Set root directory*

Description

Set root folder/directory.

Usage

```
PAT.root("C:/Users/<my_user_name>/census/<project>/")
  or PAT.root() returns current value of PAT.root
```

Arguments

`root`	Value to set root directory. NA (default) return root directory.
`default`	Default Value to set root directory. Default for default getwd()

Details

Sets global PAT option "PAT.root"

Value

`root`	Current value of root directory/folder.

Author(s)

David Dorer 10 Dec 2023 14:18

PAT.select.geo *Select Geographies*

Description

Select geographies base on arguments.

Usage

```
PAT.select.geo(state="25",county="021",tract=c("400100","400500"))
```

Arguments

state	State FIPS code (character vector).
county	County FIPS code (character vector).
csd	County Subdivision (cosub or csd character vector).
place	Place FIPS code (character vector).
tract	Place FIPS code. (character vector)
blockgroup	Blockgroup FIPS code. (character vector)
block	Block FIPS code. (character vector)
type	input data.frame type: tract, blockgroup, block
update	update file name, geographies in updata will be dropped from return.
logfile	file to log messages.
debug	Vebose argument (default PAT.verbose()).

Details

The function sorts out the Summary Level based on the combination of geography arguments. In certain cases, the type argument is used. For tracts, places, counties, tracts, the Geography correlation files are build in. For blocks you must download a state GEOCORR file.

Value

data.frame Data.frame with columns:

State	State FIPS code.
Puma	PUMA FIPS code.
Place	Place FIPS code.
County	County FIPS code.
CSD	CSD (cosub) FIPS code.
Tract	Tract FIPS code.
BlockGroup	BlockGroup FIPS code.
Block	Block FIPS code.
PumaName	PUMA Name.
PlaceName	Place Name.
CSDName	CSD (cosub) Name.

Author(s)

David Dorer 28 Apr 2024 16:19

References

Missouri Census Data Center Geocorr Engine.
https://mcdc.missouri.edu/applications/geocorr2022.html

`PAT.synth.data` *Synthesize Data*

Description

Run a synthetic data model using a PUMA and set of marginal tables

Usage

```
PAT.synth.data(model="BrooklineIII",state="25",puma="03400",outtag="_test",iter=30)
```

Arguments

`model`	Model see vignette("CreatingModels")
`state`	State FIPS code
`puma`	PUMA FIPS code
`outtag`	A tag to include in some output file names
`iter`	Maximum number of IPF iterations
`vers`	Version of IPF program 1:Internal C code 2:mipfp package Ipfp function
`maxdev`	Termination criteria for relative change in marginal table deviation from one iteration to the next. Default 0.001
`add`	Add a small amount to every cell of seed/model table. This allows the data to take on non-zero weights even though the weight is zero in the PUMS data. Default 0.0
`minwt`	Records with weights < minwt will be dropped from the synthetic data. Default 0.001
`update`	Name of input checkpoint file. "" no input check point file Geographies in checkpoint file are skipped
`county`	Vector of county FIPS codes. Default character(0). For the default all the tracts in county will be run.
`tract`	Vector of tract FIPS codes Default character(0). For the default only tracts in tract will be run for county[1]
`bdir`	base directory for program run. Default PAT.root()
`odir`	Output directory. Tract synthetic files go here. Default PAT.root()output/
`ddir`	Data directory. Cached ACS table go here. Various data files are stored here. For example the cross walk files. Default PAT.root()data/
`logfile`	File to log messages. Default synth_<state>_checkpoint<outtag>.txt.
`vintage`	ACS vintage for marginal tables default: PAT.vintage()

period	ACS period for marginal tables default: PAT.period()
checkfile	Output check point file. As tract output files are written a record with the geography is appended to this file.
download	Should marginal tables be downloaded? Default PAT.download()
pums.vintage	PUMS/PUMA vintage default: PAT.pums.vintage()
pums.period	PUMS/PUMA period default: PAT.pums.period()
key	Census key. Default PAT.census.key().
debug	Verbose/debug level for printing messages. Default PAT.verbose()
dump.seed	Dump the PUMA seed table/dataset in a file. 0 no dump 1 dump. Default 0

Details

stuff

Value

A log of the run. The other effects are the output files.

synthetic data

The output file go in odir with names synth_data_<state_FIPS>_<puma_FIPS>_<county_FIPS>_<tra
The file has 3 initial comment lines with the output file name, last iteration, number of iterations (argument), convergence flag (1:yes 2:no) nages and nraces. The 4th row is a header record.
Each data row has header variables:

state	State FIPS
puma	PUMA FIPS
county	County FIPS
tract	Tract FIPS
blockgroup	Blockgroup FIPS, set to "".
model	Model name
type	person or house.
ages	age parameter e.g. "9", "7", "7a".
races	race parameter, "3", "5", "7".
date	date-time when record was written.
variables ...	columns with the value of the model variables.

Check point file

synth_<state>_checkpoint<outtag>.txt
The file has the state, puma,count,tract and additional information. By using the check point file previously completed tracts will be skipped. Useful when you are doing an entire state and your computer updates and reboots. The file is appended to so if you want to start over but still use the update feature delete the check point file before you start a clean run.

Tract files — The synthetic data for each tract is put in a separate file in folder odir.

log file — Folder logdir() file name. Default log<outtag>.txt

Author(s)

David Dorer 20 Dec 2023 11:52

References

vignette("CreateModels")

vignette("SynthesizeData")

`PAT.synth.example` *Run Synthetic Data Example/Test*

Description

Example/Test run of PAT.synth.data()

Usage

```
PAT.synth.example()
```

Arguments

No arguments.

Details

Runs the synthetic data generation step on an example. Used for testing setup.

Value

Log file, checkpoint file and synthetic tract data files.

Author(s)

David Dorer 16 Dec 2023 19:17

`PAT.synth.repweights` *Create synthetic data with replicate weights*

Description

Creates synthetic data for tracts, takes synthetic data for tracts and creates synthetic data for block-groups, takes synthetic data for blockgroups and creates synthetic data for blocks.

Usage

```
PAT.synth.repweights() Create synthetic data for tracts.
PAT.rep.blockgroups() Create synthetic data for blockgroups.
PAT.rep.blocks() Create synthetic data for blocks.
```

Arguments

modela	model that was used to create of tract level synthetic data
modelb	model to use to create blockgroup level synthetic data (for PAT.rep.blockgroups and PAT.rep.blocks only)
modelc	model to use to create block level synthetic data (for PAT.rep.blocks only)
state	state FIPS code (character vector)
county	county FIPS code (character vector)
tract	tract FIPS code (character vector can be "")
blockgroup	tract FIPS code (character vector can be "" for PAT.rep.blocks and PAT.rep.blocks only)
block	block FIPS code (character vector can be "" for PAT.rep.blocks only)
add	amount to add to seed table so that there will not be structural zeros
niter	maximum number of itertions in Iterative Proportional Fit (IPF) algorithm (default 25)
addm	small amount to add to marginal table cells to avoid structural zeros (default 0.1)
maxdev	stopping critera for IPF when the incremental change in the fit is less than maxdev iterations stop
eps	stopping criteria in replicate weight generation (default 1.0)
maxit	maximum number of iterations in replicate weight generation (default 50)
period	period for ACS marginal tables (default 5)
pums.period	period for ACS PUMS data (default 1)
vintage	vintage for ACS marginal tables (default 2022)
pums.vintage	vintage for ACS PUMS data (default 2022)
update	update file (checkpoint file from previous run) geographies in this file will be skipped. Useful for a restart.
odir	output directory/folder for synthetic data. (default outdir())
ddir	data directory/folder for cached downloaded ACS marginal tables (default datadir())
minwt	minimum weights for output file records with weight<minwt will be dropped from output file. Remaing weight are rescaled to total population for that replicate weight. (default 0.1)
cut	used in selection of minimum weight for outfile
outtag	tag for output log file. (default "_test")
load.data	load previously cached PUMS? data. (default FALSE)
inptag	tag for input files. (default "_test"
checkfile	name for output checkpoint file. As output files are written a record with state,county,tract etc will be written in this file. Can be used for later restart. (default logdir()/synth_<state>_checkpoint
debug	verboseness of output messages (default PAT.verbose())
download	Download files using Census API? (default PAT.download())
key	Census key. (default PAT.census.key())

Details

Output of `PAT.repweights()` is used as input to `PAT.reptable()`

Value

Output data file with synthetic data. Logfile. Cached ACS marginal tables.

Author(s)

David Dorer 28 Apr 2024 17:42

Examples

```
example
```

`PAT.test.model` *Test PUMS model*

Description

Test/Run a PUMS model checking for errors.

Usage

```
PAT.test.model(model="BrooklineIII")
```

Arguments

`model`	model or model name (with or without quotes). Default: "TestI"
`state`	state FIPS
`puma`	PUMA FIPS
`key`	census key (default PAT.census.key())
`vintage`	marginal tables vintage (default PAT.vintage())
`period`	marginal tables period (default PAT.period())
`download`	download data 1:download 0:used cached data (default PAT.download())
`debug`	debug/message level - higher more messages 0:no messages 1+: more messages (default PAT.verbose())

Details

Tests model to be used for synthetizing data.

Value

list with components:

`marginals`	marginal tables/targets ("B","S","DP" PUMA geography table)
`variables`	one-way PUMS/PUMA frequency (weighted) of PUMA model table for variables used in marginal tables
`model.vars`	one-way PUMS/PUMA frequency tables for variables not used in marginal tables ("carry along" variables)
`total.pop`	total population for marginal tables.

puma.population	total population for PUMA PUMA data (sum of PUMS weights)
diag	Fit diagnostics for marginal fit
diag$marginal	Observed marginal.
diag$fit	"Expected" fitted marginal based on the PUMS table/model.
diag$std.moe	Standardized residual based on marginal table MoE.
diag$std.res	Standardized residual based on sqrt((Observed-fit)^2)/fit) with sign
diag$devres	Deviance residual (with sign)
par	various paramenters state FIPS, PUMA FIPS, detail tables vintage & period, PUMS vintage & period.

Author(s)

David Dorer 10 Dec 2023 14:34

References

help(PAT.models)
help(Builtin.variables)
help(Builtin.marginals)

Examples

```
PAT.test.model("TestI",state="25",puma="03400")
PAT.test.model("NewYorkCityI",state="36",puma="03701")
```

PAT.vintage *Set Period & Vintage*

Description

Set Marginal Tables Vintage/Period and PUMS Variable Vintage/Period

Usage

```
PAT.vintage(2021)
PAT.vintage()
PAT.period(5)
PAT.pums.vintage(2021)
PAT.pums.period(1)
With no argument return value.
```

Arguments

period	Period PAT.acs.table or period for PUMS data (1 or 5).

Details

Without an argument return current value.

Value

Current vintage or period.

Author(s)

David Dorer

`RecodeAges` *Recode Age Categories*

Description

Recode age categories of tables derived from B01001 and B01001A-I.

Usage

```
RecodeAges(tab,nages="9")
```

Arguments

`tab`	table with 14 or 23 rows, age rownames, and colnames c("Est","MoE")
`nages`	"number" of age categories, e.g. 9, "7a" etc.
`doc`	Flag TRUE/1 print documentation/return age recode list for nages. Default 0.
`debug`	debug/message level. Default 0. Higher values of debug print more messages.

Details

B01001 has 2 levels Sex (Male/Female) and age categories. For example the "Male" rows need to be pulled out before the recode.

Value

`table`	Table with Est and MoE columns and recode age rownames/values.

Author(s)

David Dorer 13 Dec 2023 11:08

SplitString *Split Table Rownames*

Description

Split the rownames of a table into components.

Usage

```
SplitString("Male : Under 5 years",s=" : ")
```

Arguments

x	character vector.
s	String to use to split x into components. Default " : ".

Details

Separates the "levels" of a table rowname by splitting it at a " : " string. Used to separate components of an ACS table rownames/labels.

Value

Character matrix with as many rows as the length of x and as many columns as the element of x with the largest number of substrings. The rows of the matrix are "padded out" by "" character string as needed.

Author(s)

David Dorer 12 Dec 2023 19:56

SPM.2022.25 *Supplemental Poverty Measure dataset for Massachusetts (FIPS 25).*

Description

Supplemental Poverty Measure dataset subset for Massachusetts (FIPS 25).

Usage

```
data(SPM.2022.25)
```

Value

A list with 2 components:

data	A data.frame with 67926 observations and 46 variables.
param	a named character vector with (state, object, vintage, date, ddir)

References

SPM datasets https://www.census.gov/topics/income-poverty/supplemental-poverty-measure/data/datasets.html
Data Dictionary https://www2.census.gov/programs-surveys/supplemental-poverty-measure/datasets/spm/spm-asc-2020.pdf

---

`WeightedSum` *Operations on Multidimensional Tables with Est and MoE Dimensions*

Description

Compute weighted sums, ratios, product and percents of table with an "Est" and "MoE" dimension.

Usage

```
WeightedSum(table,weights=c(1,-1))

      ProductEstMoE(table1,table2);

      RatioEstMoE(numerator.table,denominator.table)

      PercentEstMoE(numerator.table,denominator.table)
```

Arguments

`x`	Table with rows and columns. The dimnames for the last dimension are c("Est","MoE")
`weights`	Used with WeightedSum. Numeric vector of weights with length 1 or nrows(x). If weights has length 1 it is repeated down the rows. Default 1.
`y`	For RatioEstMoE and PercentEstMoE the denominator table.

Details

Computes the weighted sum, ratio and product of the "Est" column and the associated MoE.

For WeightedSum the default weights (1) the Est result is the sum of the table rows and MoE is the associated MoE. To take the difference of 2 rows use c(1,-1) as the weights.

RatioEstMoe computes the ratio of 2 table/matricies with Est and MoE columns.

ProductEstMoE Computes the product of w tables with Est and MoE columns.

PercentEstMoE works on multidimensional tables. If the denominator table is a matrix, then the denominator table is repeated to match the dimensions of first argument x.

The MoE is computed using the approximate formulas for standard error and variance.

Value

`table`	A table with 1 row and 2 columns "Est" and "MoE". For PercentEstMoE value is a multiway table with the same dimension as x. The last dimension of x must be c("Est","MoE")

Author(s)

David Dorer 28 Dec 2023 20:35

References

Understanding and Using the American Community Survey. Chapter 8 https://www.census.gov/content/dam/Census/libra

Index