R documentation

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

December 9, 2023

R topics documented:

`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
Poverty2.m	Poverty 2 categories: Below/Above/Undefined
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
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
Tenure.m	Rent, Own, Group_quarters.
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

`Builtin.models` *Builtin Models*

Value

`BrooklineI.model`

marginal tables: AgeRaceSex, MaritalStatus3, HispanicAgeSex, Education4, Employed, Tenure, DisabilityRace, Poverty2, HouseType, FamilyType3, HealthIns3

extra variables: SPM3 (Supplemental Poverty Measure)

`PennsylvaniaI.model`
: marginal.tables: AgeRaceSex, Poverty4, Employed, Education4, HispanicAge-Sex, MaritalStatus3, Tenure.
extra variables: WIC, Age6a, EmployedHouse18.

Author(s)

David Dorer

`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

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	Is the person employed ("Under_16","Employed","Unemployed","Not_in")
Tenure.v	"Group_quarters","Own","Rent".
HousingCost.v	"Low", "High", "Vacant".
Poverty2.v	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.m	Non_inst, Inst (Institutional), House (Household)
Education4.m	"Less_than_high_school", "High_school", "Some_college", "College_degree"
FamilyType3.v	"Married_head", "Female_head", "Male_head", "Other".
FamilyType.v	"Married_head", "Female_head", "Male_head", "Child_group", "18_years_and_over".
MaritalStatus3a.m	Married, Single, Under_15_years

MaritalStatus7.m	Married, Widowed, Separated, Divorced Never_married, Other, Under_15_years.

Author(s)

David Dorer

`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, S, or DP) Table*

Description

Download and ACS 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.acs.vintage().
`period`	Defaults to PAT.acs.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, S, DP (SF1) 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

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.set.ages("7a")
PAT.set.races("5")
```

Arguments

code	Age code
code	Race code

With no argument return the Age or Race Code

Details

Without an argument return current setting for Age or Race code.

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

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

References

xxx

`PAT.convert.spm` *Convert Supplementel Poverty Measure US SAS Data*

Description

Converts sas7bdat binary file downloaded from Census website.

Usage

```
PAT.convert.spm(vintage="2021")
```

Arguments

`vintage`	Vintage for SPM SAS file.
`ddir`	Output data folder/directory. Default datadir().
`download`	Should sas7bdat file be downloaded. Default PAT.download().

Details

The data file is very large, typically 900 MB and download and conversion can take 5-10 minutes depending the the speed of your connection and your processor speed. Requires package require("sas7bdat") which may need to be installed install.packages("sas7bdat") On Linux sas7bdat often throws and error and the conversion step fails. The Windows package doesn't seem to have this problem.

Value

Output U.S. file spm_pu_<vintage>.RData.

For download url=

https://www2.census.gov/programs-surveys/supplemental-poverty-measure/datasets/spm/spm_pu_<vintage>.sas7bdat

Author(s)

David Dorer

PAT.merge.synth.data *Merge Synthetic Tract Files*

Description

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

Usage

```
PAT.merge.synth.data(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	Zip tract files before merging.
outtag	Tag to append to 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_<state><outtag>.csv contains stacked individual tract csv files.

Author(s)

David Dorer

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.root *Set root directory*

Description

Set root folder/directory.

Usage

```
junk
```

Arguments

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

Details

xxx

Value

root	current value of root directory/folder

Author(s)

David Dorer

PAT.synth.data *Synthetize 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.acs.vintage()
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()
download	download files. Default: PAT.download().
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

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 had 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 stage 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 log<outtag>.txt

Author(s)

David Dorer 8 Dec 2023 14:41

References

vignette("CreateModels") vignette("SynthesizeData")

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

Usage

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

Arguments

`model`	model or model name (with or without quotes)
`state`	state FIPS
`puma`	PUMA FIPS
`key`	census key (default PAT.census.key())
`vintage`	marginal tables vintage (default PAT.acs.vintage())
`period`	marginal tables period (default PAT.acs.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)
`par`	various paramenters state FIPS, PUMA FIPS, detail tables vintage & period, PUMS vintage & period.

Author(s)

David Dorer

References

PAT.model

`PUMA.2012.Tract.2020` *PUMA (2012) to Tract (2020) Relationship File*

Description

PUMA to Tract Correlation/Relationship Correspondence Dataset

Usage

```
data(PUMA.2020.Tract.2020)  data(PUMA.2012.Tract.2020)
```

Format

88865 observations (2012 PUMAs)
or 85395 observations (2020 PUMAs)

Value

A data.frame PUMA.2012.Tract.2020 or PUMA.2020.Tract.2020

Variable names

State	State FIPS code.
Puma	PUMA FIPS code.
County	County FIPS code.
Tract	Tract FIPS code.
StateAb	State Postal Abbrevation.
PumaName	County name.
PumaName	PUMA name.
Allocation	Fraction of Tract intersecting PUMA.

References

Missouri Census Data Center Geocorr Engine.
https://mcdc.missouri.edu/applications/geocorr2022.html
Census Public Use Microdata Areas (PUMAs) webpage. 2020_Census_Tract_to_2020_PUMA.txt

SPM_2021_25_data *Supplemental Poverty Measure Data for Massachusetts*

Description

Census 2021 SPM data with records for Massachusetts FIPS 25

Usage

```
data(SPM.2021.25.data
```

Format

A data.frame with 67951 observations and 46 variabless.

Value

An object SPM.2020.25.data

References

Codebook

Index