Package 'sparseregMRP'

May 7, 2019

Title MRP Estimates from sparsereg Regressions

Version 0.0.1

Description Functions to facilitate MRP estimates from sparsereg multilevel regressions. The package is a set of custom functions, but adapted for general use in the future.

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

R topics documented:

`count_cellsize` *Given a individual level population dataset and variables of interest, count the number of population and the fraction of that subpopulationin the geography of interest*

Description

Given a individual level population dataset and variables of interest, count the number of population and the fraction of that subpopulationin the geography of interest

Usage

```
count_cellsize(size, popvar, geovar, ...)
```

Arguments

size	A population, or census data frame
popvar	The variable that indicates the population
geovar	The variable that indicates the final grouping variable of interest
...	variables that are in the regression

Examples

```
size <- read_csv("https://www.shirokuriwaki.com/datasets/popsize08.csv")
size_cell <- count_cellsize(size, popvar = pop2008, geovar = stt, eth, age, sex, inc, edu)
```

---

dev_logit *Deviance from predicted values of a logit model*

Description

Deviance from predicted values of a logit model

Usage

```
dev_logit(y, xb = NULL, phat = NULL, w = 1)
```

Arguments

y	vector of the outcome that was regressed on, or the true values to predict
xb	estimates from the logit model, before the link function is applied.
phat	estimates from the logit model, after link function is applied. Only either xb or phat is necessary.
w	survey weights if appplicable

Details

The deviance without weights, $dev(\hat{p})$ for target y is computed as

$$-2 \times \sum_{i=1}^{n} y_i \log(\hat{p_i}) + (1 - y_i) \log(1 - \hat{p_i})$$

where

$$\hat{p} = \frac{\exp X\beta}{\exp X\beta + 1}$$

.

With weights, this simply becomes

$$-2 \times \sum_{i=1}^{n} w_i y_i \log(\hat{p_i}) + w_i(1 - y_i) \log(1 - \hat{p_i})$$

eval_mrp *function to evaluate RMSE and correlation*

Description

function to evaluate RMSE and correlation

Usage

```
eval_mrp(tbl, truth, est_suffix)
```

Arguments

truth	variable name of the truth to predict (no quotes)
est_suffix	regex for variable names of estimates to consider

Value

A dataframe with three columns, `model`, `rmse`, and `corr`. One row for each model specified

expit *Inverse logit function (xb to p)*

Description

Inverse logit function (xb to p)

Usage

```
expit(x)
```

logit *Logit function (p to xb)*

Description

log(p / (1 - p))

Usage

```
logit(x)
```

mrp_estimate *wrapper function for MRP estimates*

Description

wrapper function for MRP estimates

Usage

```
mrp_estimate(model, est_name, cells = size_cell, geovar = stt,
  fracvar = frac_of_state, return_vec = FALSE)
```

Arguments

model	a lm, glm, glmer, or sparsereg object
est_name	pick a name for your estimate
cells	cell counts produced from count_cellsize
geovar	variable for geography of interest
fracvar	variabel for the cell fraction
return_vec	whether to return the results as a data frame or vector

predict_on_cells *Predict a model on a dataset. Intended to predict on to a cell count*

Description

Predict a model on a dataset. Intended to predict on to a cell count

Usage

```
predict_on_cells(model, data, draws = 1000)
```

Arguments

model	a lm, glm, glmer, stan_glmer, or sparsereg object
data	a data frame with the appropriate predictors, output from count_cellsize
draws	number of posterior draws in stan

Value

A vector of predicted values

scatter_mrp *Diagnostic graphic*

Description

Scatterplot with x as truth and y as estiamte

Usage

```
scatter_mrp(x, y, ggtemplate = gg0, data = mrp_df,
  null = global_mean)
```

Arguments

x	the xaxis variable (no quotes)
y	the yaxis variable (no quotes)
ggtemplate	an empty ggplot object with formatting layout
data	a dataframe which includes columns x and y
null	the global mean a null benchmark

sum_to_geo *Sum up the number of observations based on a geography*

Description

Sum up the number of observations based on a geography

Usage

```
sum_to_geo(data, geovar, predvar, fracvar, name)
```

Arguments

data	the dataset that has all the variables
geovar	the variable indicating the geography to sum up to
predvar	the variable for the predicted value
fracvar	the variable that indicates the relative size of the cell, a string
name	name for the new var, a string

synth_cellsize *Synthetic expansion*

Description

Creates a synthetic post-stratification dataset by simple multiplication, i.e. distribution of variables in new cells (entered as marginals) are orthogonal to original counts

Usage

```
synth_cellsize(start, marginal)
```

Arguments

start	starting cell counts. These will have been estimated by census data.
marginal	new information we want to incorporate by synthetic approach, usually only as a marginal distribution. Must be a dataframe with three variables: the geography variable (same name for original count data), one variable that indicates the category of the new variable (e.g. "voted" and "not voted"), and a variable called "frac" which indicates the proportion of people of that category within that geography – i.e. all values within a geography must sum to 1.

unit_rmse *RMSE between two vectors*

Description

RMSE between two vectors

Usage

```
unit_rmse(mu = truth, muhat)
```

Arguments

mu	A vector, conventionally the estimand
muhat	same length vectorm conventionally the estimate

Index