Package ‘SynDI’

October 12, 2022

Type Package

Title Synthetic Data Integration

Version 0.1.0


License GPL-2

URL https://github.com/umich-biostatistics/SynDI

BugReports https://github.com/umich-biostatistics/SynDI/issues

Depends R (>= 3.6.0)

Imports mice, magrittr, dplyr, StackImpute, arm, boot, broom, mvtnorm, randomForest, MASS, knitr

Suggests markdown

VignetteBuilder knitr

Encoding UTF-8

LazyData true

RoxygenNote 7.2.0

NeedsCompilation no

Author Tian Gu [aut], Jeremy M.G. Taylor [aut], Bhramar Mukherjee [aut], Michael Kleinsasser [cre]

Maintainer Michael Kleinsasser <mkleinsa@umich.edu>

Repository CRAN

Date/Publication 2022-05-25 07:50:05 UTC
Create.Synthetic

Description

Creates a synthetic data set from internal data and external models.

Usage

Create.Synthetic(
  datan,  
nrep,  
Y,  
XB,  
Ytype = "binary",  
parametric,  
betaHatExt_list,  
sigmaHatExt_list = NULL
)

Arguments

datan             internal data only
nrep              number of replication when creating the synthetic data
Y                 outcome name, e.g. Y='Y'
XB                all covariate names for both X and B in the target model, e.g. XB=c('X1','X2','X3','X4','B1','B2')
Ytype             the type of outcome Y, either 'binary' or 'continuous'.
parametric        choice of "Yes" or "No" for each external model. Specify whether the external model is parametric or not, e.g. parametric=c('Yes','No')
betaHatExt_list   a list of parameter estimates of the external models. The order needs to be the same as listed in XB, and variable name is required. See example for details.
sigmaHatExt_list  a list of sigma^2 for continuous outcome fitted from linear regression. If not available or the outcome type is binary, set sigmaHatExt_list=NULL
create_synthetic_example

Value

a data.frame. The combined dataset of the internal data (of size n) and the synthetic data for the given external model (of size n * nrep). This combined dataset contains a total of n*(1+nrep) rows, one intercept column (Int), one outcome column (Y), one indicator column (S), and all the predictors in the internal data. S is the indicator variable, where the internal data is indicated as S=0, and the synthetic data is indicated as S=1. The internal data part is a complete dataset without any missingness. The synthetic data part may contain missingness for certain predictors that were not used in the external model.

References


Examples

data(create_synthetic_example)
nrep = create_synthetic_example$nrep
datan = create_synthetic_example$datan
betaHatExt_list = create_synthetic_example$betaHatExt_list

data.combined = Create.Synthetic(nrep = nrep, datan = datan, Y = 'Y',
XB = c('X1', 'X2', 'X3', 'X4', 'B1', 'B2'), Ytype = 'binary',
parametric = c('Yes', 'No'), betaHatExt_list = betaHatExt_list,
sigmaHatExt_list = NULL)
expit

**Expit function**

**Description**

Expit function

**Usage**

expit(x)

**Arguments**

- **x**: vector to expit

**Value**

numeric vector with the value of the expit function \( y = \text{expit}(x) = \exp(x)/(1+\exp(x)). \)

Expit helper function.

---

**Initial.estimates**

**Internal estimation**

**Description**

Calculate the initial estimates for external populations.

**Usage**

Initial.estimates(datan, gamma.I, X, B, beta, Btype)

**Arguments**

- **datan**: internal data only
- **gamma.I**: regression estimates using internal data only (datan)
- **X**: a vector of predictor that were used in the external study, e.g. \( X = c(\text{"X1"}, \text{"X2"}, \text{"X3"}) \)
- **B**: a vector of covariates that were not used in the external study, e.g. \( B = c(\text{"X4"}, \text{"B1"}, \text{"B2"}) \)
- **beta**: a vector of external model estimates, the vector order should be the same as listed in X, e.g. names(beta) = c("int", "X1", "X2", "X3")
- **Btype**: a vector of type of B, either continuous or binary. If "continuous", linear regression will be used; if "binary", logistic regression will be used. More types can be implemented manually.
Value

a numeric vector of estimated coefficients of the target model for the given external population. Assume the internal data contains p predictors. The vector is of dimension (p+1), including the estimates of the intercept.

References


Examples

```r
# data(initial_estimates_example)

datan = initial_estimates_example$datan
gamma.I = initial_estimates_example$gamma.I
beta = initial_estimates_example$beta

# calculate the initial gamma for population S=1
gamma.S1.origin = Initial.estimates(datan = datan, gamma.I = gamma.I,
X = c('X1', 'X2', 'X3'), B = c('X4', 'B1', 'B2'),
beta = beta, Btype = c('continuous', 'continuous', 'binary'))
```

initial_estimates_example

*Example data for Initial.estimates()*

Description

Example data set for Initial.estimates()

Format

a list with

• datan simulated internal data set
• gamma.I internal gamma coefficients
• beta beta estimates from external model 1
Resample.gamma.binaryY

Resample for bootstrap variance for binary Y

Description
Resampling function to get bootstrap variance for binary Y. Note that readers need to modify the existing function Resample.gamma.binaryY() to match their own Steps 1-5. It was only included in the package for the purpose of providing an example.

Usage
Resample.gamma.binaryY(data, indices)

Arguments
- data: synthetic data
- indices: row indices to replicate

Value
numeric vector of regression coefficients

References

Resample.gamma.continuousY

Resample for bootstrap variance continuous Y

Description
Resampling function to get bootstrap variance for continuous Y. Note that readers need to modify the existing function Resample.gamma.continuousY() to match their own Steps 1-5. It was only included in the package for the purpose of providing an example.

Usage
Resample.gamma.continuousY(data, indices)
Resample.gamma.continuousY

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>data</td>
<td>synthetic data</td>
</tr>
<tr>
<td>indices</td>
<td>row indices to replicate</td>
</tr>
</tbody>
</table>

Value

numeric vector of regression coefficients

References

Index

* data
  create_synthetic_example, 3
  initial_estimates_example, 5

Create.Synthetic, 2
create_synthetic_example, 3

expit, 4

Initial.estimates, 4
initial_estimates_example, 5

Resample.gamma.binaryY, 6
Resample.gamma.continuousY, 6