Package 'SurrogateRegression'

October 1, 2023

Title Surrogate Outcome Regression Analysis

Date 2023-10-01

Version 0.6.0.1

Description Performs estimation and inference on a partially missing target outcome (e.g. gene expression in an inaccessible tissue) while borrowing information from a correlated surrogate outcome (e.g. gene expression in an accessible tissue). Rather than regarding the surrogate outcome as a proxy for the target outcome, this package jointly models the target and surrogate outcomes within a bivariate regression framework. Unobserved values of either outcome are treated as missing data. In contrast to imputation-based inference, no assumptions are required regarding the relationship between the target and surrogate outcomes. Estimation in the presence of bilateral outcome missingness is performed via an expectation conditional maximization either algorithm. In the case of unilateral target missingness, estimation is performed using an accelerated least squares procedure. A flexible association test is provided for evaluating hypotheses about the target regression parameters. For additional details, see: McCaw ZR, Gaynor SM, Sun R, Lin X: ``Leveraging a surrogate outcome to improve inference on a partially missing target outcome" <doi:10.1111/biom.13629>.

Depends R (>= 3.4.0)

License GPL-3

Encoding UTF-8

LinkingTo Rcpp, RcppArmadillo

Imports methods, Rcpp, stats

RoxygenNote 7.2.3

Suggests testthat (>= 3.0.0), knitr, rmarkdown, withr

VignetteBuilder knitr

Config/testthat/edition 3

NeedsCompilation yes

Author Zachary McCaw [aut, cre] (<https://orcid.org/0000-0002-2006-9828>)

Maintainer Zachary McCaw <zmccaw@alumni.harvard.edu>

Repository CRAN

Date/Publication 2023-10-01 16:10:02 UTC

R topics documented:

bnr-class *Bivariate Regression Model*

Description

Bivariate Regression Model

Slots

`Covariance` Residual covariance matrix.

`Covariance.info` Information for covariance parameters.

`Covariance.tab` Table of covariance parameters.

`Method` Method used for estimation.

`Regression.info` Information for regression coefficients.

`Regression.tab` Table of regression coefficients.

`Residuals` Outcome residuals.

CheckInit *Check Initiation*

Description

Check Initiation

Usage

```
CheckInit(init)
```

Arguments

`init`	Optional list of initial parameters for fitting the null model.

CheckTestSpec *Check Test Specification*

Description

Check Test Specification

Usage

```
CheckTestSpec(is_zero, p)
```

Arguments

is_zero	Logical vector, with as many entires as columns in the target model matrix, indicating which columns have coefficient zero under the null.
p	Number of columns for the target model matrix.

coef.bnr *Extract Coefficients from Bivariate Regression Model*

Description

Extract Coefficients from Bivariate Regression Model

Usage

```
## S3 method for class 'bnr'
coef(object, ..., type = NULL)
```

Arguments

object	bnr object.
...	Unused.
type	Either Target or Surrogate.

CovInfo *Covariance Information Matrix*

Description

Covariance Information Matrix

Usage

```
CovInfo(data_part, sigma)
```

Arguments

data_part	List of partitioned data. See PartitionData.
sigma	Target-surrogate covariance matrix.

Value

3x3 Numeric information matrix for the target variance, target-surrogate covariance, and surrogate variance.

CovTab *Tabulate Covariance Parameters*

Description

Tabulate Covariance Parameters

Usage

```
CovTab(point, info, sig = 0.05)
```

Arguments

point	Point estimates.
info	Information matrix.
sig	Significance level.

Value

Data.table containing the point estimate, standard error, and confidence interval.

CovUpdate *Covariate Update*

Description

Covariate Update

Usage

```
CovUpdate(data_part, b0, a0, b1, a1, sigma0)
```

Arguments

data_part	List of partitioned data. See PartitionData.
b0	Previous target regression coefficient.
a0	Previous surrogate regression coefficient.
b1	Current target regression coefficient.
a1	Current surrogate regression coefficient.
sigma0	Initial target-surrogate covariance matrix.

Value

ECM update of the target-surrogate covariance matrix.

FitBNEM *Fit Bivariate Normal Regression Model via Expectation Maximization.*

Description

Estimation procedure for bivariate normal regression models in which the target and surrogate outcomes are both subject to missingness.

Usage

```
FitBNEM(
  t,
  s,
  X,
  Z,
  sig = 0.05,
  b0 = NULL,
  a0 = NULL,
  sigma0 = NULL,
  maxit = 100,
  eps = 1e-06,
  report = TRUE
)
```

Arguments

`t`	Target outcome vector.
`s`	Surrogate outcome vector.
`X`	Target model matrix.
`Z`	Surrogate model matrix.
`sig`	Type I error level.
`b0`	Initial target regression coefficient.
`a0`	Initial surrogate regression coefficient.
`sigma0`	Initial covariance matrix.
`maxit`	Maximum number of parameter updates.
`eps`	Minimum acceptable improvement in log likelihood.
`report`	Report fitting progress?

Details

The target and surrogate model matrices are expected in numeric format. Include an intercept if required. Expand factors and interactions in advance. Initial values may be specified for any of the target coefficient `b0`, the surrogate coefficient `a0`, or the target-surrogate covariance matrix `sigma0`.

Value

An object of class 'bnr' with slots containing the estimated regression coefficients, the target-surrogate covariance matrix, the information matrices for the regression and covariance parameters, and the residuals.

`FitBNLS` *Fit Bivariate Normal Regression Model via Least Squares*

Description

Estimation procedure for bivariate normal regression models in which only the target outcome is subject to missingness.

Usage

```
FitBNLS(t, s, X, sig = 0.05)
```

Arguments

`t`	Target outcome vector.
`s`	Surrogate outcome vector.
`X`	Model matrix.
`sig`	Type I error level.

Details

The model matrix is expected in numeric format. Include an intercept if required. Expand factors and interactions in advance.

Value

An object of class 'bnr' with slots containing the estimated regression coefficients, the target-surrogate covariance matrix, the information matrices for the regression and covariance parameters, and the residuals.

FitBNR *Fit Bivariate Normal Regression Model*

Description

Estimation procedure for bivariate normal regression models. The EM algorithm is applied if `s` contains missing values, or if `X` differs from `Z`. Otherwise, an accelerated least squares procedure is applied.

Usage

```
FitBNR(t, s, X, Z = NULL, sig = 0.05, ...)
```

Arguments

`t`	Target outcome vector.
`s`	Surrogate outcome vector.
`X`	Target model matrix.
`Z`	Surrogate model matrix. Defaults to `X`.
`sig`	Significance level.
`...`	Additional arguments accepted if fitting via EM. See `FitBNEM`.

Details

The target and surrogate model matrices are expected in numeric format. Include an intercept if required. Expand factors and interactions in advance.

Value

An object of class 'mnr' with slots containing the estimated regression coefficients, the target-surrogate covariance matrix, the information matrices for regression parameters, and the residuals.

Examples

```
# Case 1: No surrogate missingness.
set.seed(100)
n <- 1e3
X <- stats::rnorm(n)
data <- rBNR(
  X = X,
  Z = X,
  b = 1,
  a = -1,
  t_miss = 0.1,
  s_miss = 0.0
)
t <- data[, 1]
s <- data[, 2]

# Model fit.
fit_bnls <- FitBNR(
  t = t,
  s = s,
  X = X
)

# Case 2: Target and surrogate missingness.
set.seed(100)
n <- 1e3
X <- stats::rnorm(n)
Z <- stats::rnorm(n)
data <- rBNR(
  X = X,
  Z = Z,
  b = 1,
  a = -1,
  t_miss = 0.1,
  s_miss = 0.1
)

# Log likelihood.
fit_bnem <- FitBNR(
  t = data[, 1],
  s = data[, 2],
  X = X,
  Z = Z
)
```

`fitOLS` *Ordinary Least Squares*

Description

Fits the standard OLS model.

Usage

```
fitOLS(y, X)
```

Arguments

y	Nx1 Numeric vector.
X	NxP Numeric matrix.

Value

List containing the following:

Beta	Regression coefficient.
V	Outcome variance.
Ibb	Information matrix for beta.
Resid	Outcome residuals.

FormatOutput *Format Output*

Description

Format Output

Usage

```
FormatOutput(data_part, method, b, a, sigma, sig)
```

Arguments

data_part	List of partitioned data. See PartitionData.
method	Estimation method.
b	Final target regression parameter.
a	Final surrogate regression parameter.
sigma	Final target-surrogate covariance matrix.
sig	Significance level.

Value

Object of class 'bnr'.

IterUpdate *Update Iteration*

Description

Update Iteration

Usage

```
IterUpdate(theta0, update, maxit, eps, report)
```

Arguments

theta0	List containing the initial parameter values.
update	Function to iterate. Should accept and return a list similar parameter values.
maxit	Maximum number of parameter updates.
eps	Minimum acceptable improvement in log likelihood.
report	Report fitting progress?

matDet *Matrix Determinant*

Description

Calculates the determinant of A.

Usage

```
matDet(A, logDet = FALSE)
```

Arguments

A	Numeric matrix.
logDet	Return the logarithm of the determinant?

Value

Scalar.

matInv *Matrix Inverse*

Description

Calcualtes A^{-1}.

Usage

```
matInv(A)
```

Arguments

A	Numeric matrix.

Value

Numeric matrix.

matIP *Matrix Inner Product*

Description

Calculates the product $A'B$.

Usage

```
matIP(A, B)
```

Arguments

A	Numeric matrix.
B	Numeric matrix.

Value

Numeric matrix.

matOP *Matrix Outer Product*

Description

Calculates the outer product AB'.

Usage

```
matOP(A, B)
```

Arguments

A	Numeric matrix.
B	Numeric matrix.

Value

Numeric matrix.

matQF *Quadratic Form*

Description

Calculates the quadratic form $X'AX$.

Usage

```
matQF(X, A)
```

Arguments

X	Numeric matrix.
A	Numeric matrix.

Value

Numeric matrix.

MMP *Matrix Matrix Product*

Description

Calculates the product AB.

Usage

```
MMP(A, B)
```

Arguments

A	Numeric matrix.
B	Numeric matrix.

Value

Numeric matrix.

ObsLogLik *Observed Data Log Likelihood*

Description

Observed Data Log Likelihood

Usage

```
ObsLogLik(data_part, b, a, sigma)
```

Arguments

data_part	List of partitioned data. See PartitionData.
b	Target regression coefficient.
a	Surrogate regression coefficient.
sigma	Target-surrogate covariance matrix.

Value

Observed data log likelihood.

ParamInit *Parameter Initialization*

Description

Parameter Initialization

Usage

```
ParamInit(data_part, b0, a0, sigma0)
```

Arguments

data_part	List of partitioned data. See PartitionData.
b0	Initial target regression coefficient.
a0	Initial surrogate regression coefficient.
sigma0	Initial covariance matrix.

Value

List containing initial values of beta, alpha, sigma.

PartitionData *Partition Data by Outcome Missingness Pattern.*

Description

Partition Data by Outcome Missingness Pattern.

Usage

```
PartitionData(t, s, X, Z = NULL)
```

Arguments

t	Target outcome vector.
s	Surrogate outcome vector.
X	Target model matrix.
Z	Surrogate model matrix.

Value

List containing these components:

- 'Orig' original data.
- 'Dims' dimensions and names.
- 'Complete', data for complete cases.
- 'TMiss', data for subjects with target missingness.
- 'SMiss', data for subjects with surrogate missingness.
- 'IPs', inner products.

Examples

```
# Generate data.
n <- 1e3
X <- rnorm(n)
Z <- rnorm(n)
data <- rBNR(X = X, Z = Z, b = 1, a = -1)
data_part <- PartitionData(
  t = data[, 1],
  s = data[, 2],
  X = X,
  Z = Z
)
```

---

print.bnr *Print for Bivariate Regression Model*

Description

Print for Bivariate Regression Model

Usage

```
## S3 method for class 'bnr'
print(x, ..., type = "Regression")
```

Arguments

x	bnr object.
...	Unused.
type	Either Regression or Covariance.

rBNR *Simulate Bivariate Normal Data with Missingness*

Description

Function to simulate from a bivariate normal regression model with outcomes missing completely at random.

Usage

```
rBNR(
  X,
  Z,
  b,
  a,
  t_miss = 0,
  s_miss = 0,
  sigma = NULL,
  include_residuals = TRUE
)
```

Arguments

X	Target design matrix.
Z	Surrogate design matrix.
b	Target regression coefficient.
a	Surrogate regression coefficient.
t_miss	Target missingness in [0,1].
s_miss	Surrogate missingness in [0,1].
sigma	2x2 target-surrogate covariance matrix.
include_residuals	Include the residual? Default: TRUE.

Value

Numeric Nx2 matrix. The first column contains the target outcome, the second contains the surrogate outcome.

Examples

```
set.seed(100)
# Observations.
n <- 1e3
# Target design.
X <- cbind(1, matrix(rnorm(3 * n), nrow = n))
# Surrogate design.
```

```
Z <- cbind(1, matrix(rnorm(3 * n), nrow = n))
# Target coefficient.
b <- c(-1, 0.1, -0.1, 0.1)
# Surrogate coefficient.
a <- c(1, -0.1, 0.1, -0.1)
# Covariance structure.
sigma <- matrix(c(1, 0.5, 0.5, 1), nrow = 2)
# Data generation, target and surrogate subject to 10% missingness.
y <- rBNR(X, Z, b, a, t_miss = 0.1, s_miss = 0.1, sigma = sigma)
```

`RegInfo` *Regression Information*

Description

Regression Information

Usage

```
RegInfo(data_part, sigma, as_matrix = FALSE)
```

Arguments

`data_part`	List of partitioned data. See `PartitionData`.
`sigma`	Target-surrogate covariance matrix.
`as_matrix`	Return as an information matrix? If FALSE, returns a list.

Value

List containing the information matrix for beta (Ibb), the information matrix for alpha (Iaa), and the cross information (Iba).

`RegTab` *Tabulate Regression Coefficients*

Description

Tabulate Regression Coefficients

Usage

```
RegTab(point, info, sig = 0.05)
```

Arguments

point	Point estimates.
info	Information matrix.
sig	Significance level.

Value

Data.table containing the point estimate, standard error, confidence interval, and Wald p-value.

RegUpdate *Regression Update*

Description

Regression Update

Usage

```
RegUpdate(data_part, sigma)
```

Arguments

data_part	List of partitioned data. See PartitionData.
sigma	Target-surrogate covariance matrix.

Value

List containing the generalized least squares estimates of beta and alpha.

---

residuals.bnr *Extract Residuals from Bivariate Regression Model*

Description

Extract Residuals from Bivariate Regression Model

Usage

```
## S3 method for class 'bnr'
residuals(object, ..., type = NULL)
```

Arguments

object	A bnr object.
...	Unused.
type	Either Target or Surrogate.

SchurC *Schur complement*

Description

Calculates the efficient information $I_{bb} - I_{ba} I_{aa}^{-1} I_{ab}$.

Usage

```
SchurC(Ibb, Iaa, Iba)
```

Arguments

Ibb	Information of target parameter
Iaa	Information of nuisance parameter
Iba	Cross information between target and nuisance parameters

Value

Numeric matrix.

ScoreBNEM *Score Test via Expectation Maximization.*

Description

Performs a Score test of the null hypothesis that a subset of the regression parameters for the target outcome are zero.

Usage

```
ScoreBNEM(
  t,
  s,
  X,
  Z,
  is_zero,
  init = NULL,
  maxit = 100,
  eps = 1e-08,
  report = FALSE
)
```

Arguments

t	Target outcome vector.
s	Surrogate outcome vector.
X	Target model matrix.
Z	Surrogate model matrix.
is_zero	Logical vector, with as many entires as columns in the target model matrix, indicating which columns have coefficient zero under the null.
init	Optional list of initial parameters for fitting the null model.
maxit	Maximum number of parameter updates.
eps	Minimum acceptable improvement in log likelihood.
report	Report model fitting progress? Default is FALSE.

Value

A numeric vector containing the score statistic, the degrees of freedom, and a p-value.

---

show,bnr-method *Show for Bivariate Regression Model*

Description

Show for Bivariate Regression Model

Usage

```
## S4 method for signature 'bnr'
show(object)
```

Arguments

object	bnr object.

SurrogateRegression *SurrogateRegression: Surrogate Outcome Regression Analysis*

Description

This package performs estimation and inference on a partially missing target outcome while borrowing information from a correlated surrogate outcome. Rather than regarding the surrogate outcome as a proxy for the target outcome, this package jointly models the target and surrogate outcomes within a bivariate regression framework. Unobserved values of either outcome are treated as missing data. In contrast to imputation-based inference, no assumptions are required regarding the relationship between the target and surrogate outcomes. However, in order for surrogate inference to improve power, the target and surrogate outcomes must be correlated, and the target outcome must be partially missing. The primary estimation function is FitBNR. In the case of bilateral missingness, i.e. missingness in both the target and surrogate outcomes, estimation is performed via an expectation conditional maximization either (ECME) algorithm. In the case of unilateral target missingness, estimation is performed using an accelerated least squares procedure. Inference on regression parameters for the target outcome is performed using TestBNR.

Author(s)

Zachary R. McCaw

TestBNR *Test Bivariate Normal Regression Model.*

Description

Performs a test of the null hypothesis that a subset of the regression parameters for the target outcome are zero in the bivariate normal regression model.

Usage

```
TestBNR(t, s, X, Z = NULL, is_zero, test = "Wald", ...)
```

Arguments

t	Target outcome vector.
s	Surrogate outcome vector.
X	Target model matrix.
Z	Surrogate model matrix.
is_zero	Logical vector, with as many entires as columns in the target model matrix, indicating which columns have coefficient zero under the null.
test	Either Score or Wald. Only Wald is available for LS.
...	Additional arguments accepted if fitting via EM. See FitBNEM.

Value

A numeric vector containing the test statistic, the degrees of freedom, and a p-value.

Examples

```
# Generate data.
set.seed(100)
n <- 1e3
X <- cbind(1, rnorm(n))
Z <- cbind(1, rnorm(n))
data <- rBNR(X = X, Z = Z, b = c(1, 0), a = c(-1, 0), t_miss = 0.1, s_miss = 0.1)

# Test 1st coefficient.
wald_test1 <- TestBNR(
  t = data[, 1],
  s = data[, 2],
  X = X,
  Z = Z,
  is_zero = c(TRUE, FALSE),
  test = "Wald"
)

score_test1 <- TestBNR(
  t = data[, 1],
  s = data[, 2],
  X = X,
  Z = Z,
  is_zero = c(TRUE, FALSE),
  test = "Score"
)

# Test 2nd coefficient.
wald_test2 <- TestBNR(
  t = data[, 1],
  s = data[, 2],
  X = X,
  Z = Z,
  is_zero = c(FALSE, TRUE),
  test = "Wald"
)

score_test2 <- TestBNR(
  t = data[, 1],
  s = data[, 2],
  X = X,
  Z = Z,
  is_zero = c(FALSE, TRUE),
  test = "Score"
)
```

tr *Matrix Trace*

Description

Calculates the trace of a matrix A.

Usage

```
tr(A)
```

Arguments

A	Numeric matrix.

Value

Scalar.

UpdateEM *EM Update*

Description

EM Update

Usage

```
UpdateEM(data_part, b0, a0, sigma0)
```

Arguments

data_part	List of partitioned data. See PartitionData.
b0	Initial target regression coefficient.
a0	Initial surrogate regression coefficient.
sigma0	Initial covariance matrix.

Value

List containing updated values for beta 'b', alpha 'a', 'sigma', the log likelihood 'loglik', and the change in log likelihood 'delta'.

vcov.bnr *Extract Covariance Matrix from Bivariate Normal Regression Model*

Description

Returns the either the estimated covariance matrix of the outcome, the information matrix for regression coefficients, or the information matrix for covariance parameters.

Usage

```
## S3 method for class 'bnr'
vcov(object, ..., type = "Regression", inv = FALSE)
```

Arguments

object	bnr object.
...	Unused.
type	Select "Covariance","Outcome",or "Regression". Default is "Regression".
inv	Invert the covariance matrix? Default is FALSE.

WaldBNEM *Wald Test via Expectation Maximization.*

Description

Performs a Wald test of the null hypothesis that a subset of the regression parameters for the target outcome are zero.

Usage

```
WaldBNEM(
  t,
  s,
  X,
  Z,
  is_zero,
  init = NULL,
  maxit = 100,
  eps = 1e-08,
  report = FALSE
)
```

Arguments

t	Target outcome vector.
s	Surrogate outcome vector.
X	Target model matrix.
Z	Surrogate model matrix.
is_zero	Logical vector, with as many entries as columns in the target model matrix, indicating which columns have coefficient zero under the null.
init	Optional list of initial parameters for fitting the null model, with one or more of the components: a0, b0, S0.
maxit	Maximum number of parameter updates.
eps	Minimum acceptable improvement in log likelihood.
report	Report model fitting progress? Default is FALSE.

Value

A numeric vector containing the Wald statistic, the degrees of freedom, and a p-value.

WaldBNLS *Wald Test via Least Squares.*

Description

Performs a Wald test of the null hypothesis that a subset of the regression parameters for the target outcome are zero.

Usage

```
WaldBNLS(t, s, X, is_zero)
```

Arguments

t	Target outcome vector.
s	Surrogate outcome vector.
X	Model matrix.
is_zero	Logical vector, with as many entires as columns in the target model matrix, indicating which columns have coefficient zero under the null.

Value

A numeric vector containing the Wald statistic, the degrees of freedom, and a p-value.

Index