Package: mccmeiv (via r-universe)

September 2, 2024

Type Package

Title Analysis of Matched Case Control Data with a Mismeasured Exposure that is Accompanied by Instrumental Variables

Version 2.1

Date 2019-02-07

Author Chris M Manuel, Samiran Sinha, and Suojin Wang

Maintainer Chris M Manuel <cmmanuel@stat.tamu.edu>

Description Applying the methodology from Manuel et al. to estimate parameters using a matched case control data with a mismeasured exposure variable that is accompanied by instrumental variables (Submitted).

License GPL-2

Imports numDeriv, survival, MASS

LazyData true

NeedsCompilation no

Date/Publication 2019-02-11 08:50:07 UTC

Repository https://chrismattstat.r-universe.dev

RemoteUrl https://github.com/cran/mccmeiv

RemoteRef HEAD

RemoteSha 7e34f7e24fbb34a30243da5e09ef34c6bf2314a3

Contents

`mccmeiv-package` *Analysis of Matched Case Control Data with a Mismeasured Exposure that is Accompanied by Instrumental Variables*

Description

This package implements the methodology found in Manuel et al. to estimate parameters of a logistic model using data from a matched case control design with a mismeasured exposure variable with the help of instrumental variables.

Author(s)

Chris M Manuel, Samiran Sinha, and Suojin Wang

Maintainer: Chris M Manuel <cmmanuel@stat.tamu.edu>

References

Manuel, CM, Sinha, S and Wang, S. Matched case-control data with a misclassified exposure: What can be done with instrumental variables? (Submitted).

`matcdata` *A sample dataset representing matched case control data.*

Description

An example data set with 1 binary response (y), 1 stratification variable (sv), 1 mismeasured binary exposure (w), 1 prognostic factor (z), and 1 instrumental variable (xs) for use with the meivm3 or meivm4 functions.

Usage

```
data(matcdata)
```

See Also

meivm3 meivm4 matcdatamult

Examples

```
data(matcdata)
out1=with(matcdata,meivm4(y=y,sv=sv,xs=xs,w=w,z=z,alpha=0.1))
```

matcdatamult — *A sample dataset representing matched case control data. Similar to matcdata except with multiple stratification/instrumental/miscellaneous variables.*

Description

An example data set with 1 binary response (y), 2 stratification variable (sv1 and sv2), 1 mismeasured binary exposure (w), 2 prognostic factors (z1 and z2), and 2 instruments (xs1 and xs2) for use with the meivm3 or meivm4 functions.

Usage

```
data(matcdatamult)
```

See Also

meivm3 meivm4 matcdata

Examples

```
data(matcdatamult)
out=with(matcdatamult,meivm3(y=y, sv=sv1,xs=xs1, w=w,z=cbind(z1,z2)))
```

meivm3 — *Two-step methodology for estimating parameters for a matched case control design with a mismeasured exposure using instrumental variables*

Description

Applies the two-step methodology from Manuel et al. to estimate parameters of a logistic model using matched case control data with a mismeasured exposure variable that is accompanied by a set of instrumental variables.

Usage

```
meivm3(y, sv, xs, w, z, sv.factor = NULL, xs.factor = NULL, z.factor = NULL,
alpha = 0.05, scale = TRUE, setalpha0.to.0 = FALSE, setalpha1.to.0 = FALSE)
```

Arguments

`y` A vector of the response variable, representing case (y=1) or control (y=0) for each observation. For a 1:M matched case-control dataset with n matched sets, y will be a vector of length N=nx(1+M), and it is y=rep(c(1, rep(0, M)), n).

`sv` A data frame or matrix of confounding variables used for matching each case with the control(s). This data frame should have N number of rows.

`xs` A data frame or matrix of instrumental variables used as proxies for the mismeasured variable. This data frame should have N number of rows.

`w` A vector of the mismeasured binary exposure variables for each observation. The length of this vector must be N.

`z` A data frame or matrix of prognostic factors used to study the association between the response and the mismeasured exposure. This data frame should have N number of rows.

`sv.factor` Specify whether any stratification variables are categorical by using the name of the column(s). For example if there is a factor variable in the user specified data frame sv, which is labeled as "Political Affiliation", then sv.factor="Political Affiliation". Alternatively if there are two factor variables in sv labeled as "Political Affiliation" and "SES", then sv.factor=c("Political Affiliation","SES"). Any stratification variable that is numeric binary does not need to be declared as a factor.

`xs.factor` Specify whether any instrumental variables are categorical by using the name of the column(s). For example if there is an instrument in the user specified xs, which is labeled as "Race of Mother", then xs.factor="Race of Mother". Alternatively if there are two factor variables in xs variable labeled as "Race of Mother" and "Race of Mother", then xs.factor=c("Race of Mother","Race of Father"). Any instrumental variable that is numeric binary does not need to be declared as a factor.

`z.factor` Specify whether any prognostic variables are categorical by using the name of the column(s). For example, if there is a factor variable in the user specified z labeled "Season", then z.factor="Season". Alternatively if there are two factor variables in z labeled "Season" and a factor variable labeled Ethnicity , then z.factor=c("Season","Ethnicity"). Any variable that is numeric binary does not need to be declared as a factor.

`alpha` Specify the level of significance for calculating the (1-alpha)100% Wald confidence intervals of the odds ratio parameter. For example, the default alpha=0.05 generates a 95% confidence interval of the odds ratio.

`scale` By default, all non factors/numeric variables (those that are not specified as either response, mismeasured covariate, or factors specified in xs.factor / sv.factor / z.factor) are automatically scaled and centered unless the user sets scale=FALSE. Moreover, no numeric binary variables are scaled.

`setalpha0.to.0` Sets the misclassification probability Pr(W=1|X=0,Y=0) = Pr(W=1|X=0,Y=1) to 0 so that it is not estimated. By default this set to FALSE. Note that this option and setalpha1.to.0 cannot be both set to TRUE simultaneously.

`setalpha1.to.0` Sets the misclassification probability Pr(W=0|X=1,Y=0) = Pr(W=0|X=1,Y=1) to 0 so that it is not estimated. By default this is set to FALSE. Note that this option and setalpha0.to.0 cannot be both set to TRUE simultaneously.

Details

Estimation of the parameters is done in two steps. In the first step the set of parameters gamma used to model the probability of true exposure status X, Pr(X=1|SV,XS,Y=0), and the parameters eta0 and eta1 used to model the misclassification probabilities alpha.0 = Pr(W=1|X=0,Y=0) = Pr(W=1|X=0,Y=1) and alpha.1 = Pr(W=0|X=1,Y=0) = Pr(W=0|X=1,Y=1) are estimated. This information is then used in the second step to calculate Pr(Y=1|SV,XS,Z). The solution is found via the optim function, using the "L-BFGS-B" method. For the first step, the starting values come from a logistic regression of W on the instruments, confounders, and prognostic factors. The starting values for eta 0 and eta 1 are set to 0. Finally, the starting values for the second step use the naive beta estimates.

Value

`two.step.results`
: Provides estimates for the beta parameters of the logistic model for the response y using the two step instrumental variable analysis. Standard errors, p-values, and the (1-alpha)100% Wald confidence intervals for exp(beta) are also included in the output.

`naive.results`
: Provides estimates for the beta parameters of the logistic model for the response y using the naive approach. Standard errors, p-values, and the (1-alpha)100% Wald confidence intervals for exp(beta) are also included in the output.

Author(s)

Chris M. Manuel, Samiran Sinha, and Suojin Wang

References

Manuel, CM, Sinha, S, and Wang, S. Matched case-control data with a misclassified exposure: What can be done with instrumental variables? (Submitted)

See Also

`meivm4` `matcdata` `matcdatamult`

Examples

```
data(matcdata)
out=with(matcdata,meivm3(y=y,sv=sv,xs=xs,w=w,z=z,alpha=0.05))
#For running data with multiple confounders/instruments/prognostic factors see 'matcdatamult'.
```

`meivm4` *Efficient procedure for estimating parameters for a matched case control design with a mismeasured exposure using instrumental variables*

Description

Apply the efficient procedure from Manuel et al. to estimate parameters of a logistic model for matched case control data with a mismeasured exposure that is accompanied by a set of instrumental variables.

Usage

```
meivm4(y, sv, xs, w, z, sv.factor = NULL, xs.factor = NULL, z.factor = NULL,
alpha = 0.05, scale = TRUE, setalpha0.to.0 = FALSE, setalpha1.to.0 = FALSE)
```

Arguments

`y`	A vector of the response variable, representing case (y=1) or control (y=0) for each observation. For a 1:M matched case-control dataset with n matched sets, y will be a vector of length N=nx(1+M), and it is y=rep(c(1, rep(0, M)), n).
`sv`	A data frame or matrix of confounding variables used for matching each case with the control(s). This data frame should have N number of rows.
`xs`	A data frame or matrix of instrumental variables used as proxies for the mismeasured variable. This data frame should have N number of rows.
`w`	A vector of the mismeasured binary exposure variables for each observation. The length of this vector must be N.
`z`	A data frame or matrix of prognostic factors used to study the association between the response and the mismeasured exposure. This data frame should have N number of rows.
`sv.factor`	Specify whether any stratification variables are categorical by using the name of the column(s). For example if there is a factor variable in the user specified data frame sv, which is labeled as "Political Affiliation", then sv.factor="Political Affiliation". Alternatively if there are two factor variables in sv labeled as "Political Affiliation" and "SES", then sv.factor=c("Political Affiliation","SES"). Any stratification variable that is numeric binary does not need to be declared as a factor.
`xs.factor`	Specify whether any instrumental variables are categorical by using the name of the column(s). For example if there is an instrument in the user specified xs, which is labeled as "Race of Mother", then xs.factor="Race of Mother". Alternatively if there are two factor variables in xs variable labeled as "Race of Mother" and "Race of Mother", then xs.factor=c("Race of Mother","Race of Father"). Any instrumental variable that is numeric binary does not need to be declared as a factor.

`z.factor` Specify whether any prognostic variables are categorical by using the name of the column(s). For example, if there is a factor variable in the user specified z labeled "Season", then z.factor="Season". Alternatively if there are two factor variables in z labeled "Season" and a factor variable labeled Ethnicity , then z.factor=c("Season","Ethnicity"). Any variable that is numeric binary does not need to be declared as a factor.

`alpha` Specify the level of significance for calculating the (1-alpha)100% Wald confidence intervals of the odds ratio parameter. For example, the default alpha=0.05 generates a 95% confidence interval of the odds ratio.

`scale` By default, all non factors/numeric variables (those that are not specified as either response, mismeasured covariate, or factors specified in xs.factor / sv.factor / z.factor) are automatically scaled and centered unless the user sets scale=FALSE. Moreover, no numeric binary variables are scaled.

`setalpha0.to.0` Sets the misclassification probability Pr(W=1|X=0,Y=0) = Pr(W=1|X=0,Y=1) to 0 so that it is not estimated. By default this set to FALSE. Note that this option and setalpha1.to.0 cannot be both set to TRUE simultaneously.

`setalpha1.to.0` Sets the misclassification probability Pr(W=0|X=1,Y=0) = Pr(W=0|X=1,Y=1) to 0 so that it is not estimated. By default this is set to FALSE. Note that this option and setalpha0.to.0 cannot be both set to TRUE simultaneously.

Details

In comparison to the methodology used in the function meivm3, the efficient estimation approach estimates the parameters in one step. These parameters include gamma used to model the probability of true exposure status X, Pr(X=1|SV,XS,Y=0), the parameters eta0 and eta1 used to model the misclassification probabilities alpha.0 = Pr(W=1|X=0,Y=0) = Pr(W=1|X=0,Y=1) and alpha.1 = Pr(W=0|X=1,Y=0) = Pr(W=0|X=1,Y=1), and the beta parameters used in modeling Pr(Y=1|SV,XS,Z). The solution is found via the optim function, using the "L-BFGS-B" method. The starting values for gamma come from a logistic regression of W on the instruments, confounders, and prognostic factors. The starting values for eta 0 and eta 1 are set to 0, and the starting values for beta are the naive estimates.

Value

`efficient.results`
Provides estimates for the beta parameters of the logistic model for the response y using the efficient estimator approach. Standard errors, p-values, and the (1-alpha)100% Wald confidence interval for exp(beta) are also included in the output.

`naive.results` Provides estimates for the beta parameters of the logistic model for the response y using the naive approach. Standard errors, p-values, and the (1-alpha)100% Wald confidence intervals for exp(beta) are also included in the output.

Author(s)

Chris M. Manuel, Samiran Sinha, and Suojin Wang

References

Manuel, CM, Sinha, S, and Wang, S. Matched case-control data with a misclassified exposure: What can be done with instrumental variables? (Submitted)

See Also

meivm3 matcdata matcdatamult

Examples

```
data(matcdata)
out2=with(matcdata,meivm4(y=y,sv=sv,xs=xs,w=w,z=z,alpha=0.01))
#For running data with multiple confounders/instruments/prognostic factors see 'matcdatamult'.
```

Index