Package ‘bucky’

October 12, 2022

Type Package

Title Bucky's Archive for Data Analysis in the Social Sciences

Version 1.0.7

Date 2022-03-25

Author Alexander Tahk [aut, cre] (<`https://orcid.org/0000-0001-7895-9420`>)

Maintainer Alexander Tahk <`tahk@alum.mit.edu`>

Depends R (>= 3.1.0)

Imports methods, sandwich, lmtest

Suggests MASS, parallel, Amelia, mice

Enhances texreg, stargazer

URL `https://github.com/atahk/bucky`

Description
Provides functions for various statistical techniques commonly used in the social sciences, including functions to compute clustered robust standard errors, combine results across multiply-imputed data sets, and simplify the addition of robust and clustered robust standard errors.

License GPL (>= 3)

NeedsCompilation no

Repository CRAN

Date/Publication 2022-03-25 23:10:02 UTC

R topics documented:

bucky-package 2
mi.eval 3
robust.summary 5
robustify 6
summary.mi.estimates 8
vcovCR 9

Index **11**

`bucky-package` *Bucky's Archive for Data Analysis in the Social Sciences*

Description

Provides functions for various statistical techniques commonly used in the social sciences, including functions to compute clustered robust standard errors, combine results across multiply-imputed data sets, and simplify the addition of robust and clustered robust standard errors.

Details

Package:	bucky
Type:	Package
Title:	Bucky's Archive for Data Analysis in the Social Sciences
Version:	1.0.7
Date:	2022-03-25
Authors@R:	c(person("Alexander", "Tahk", role = c("aut", "cre"), email = "tahk@alum.mit.edu", comment = c(ORCID = "
Author:	Alexander Tahk [aut, cre] (<https://orcid.org/0000-0001-7895-9420>)
Maintainer:	Alexander Tahk <tahk@alum.mit.edu>
Depends:	R (>= 3.1.0)
Imports:	methods, sandwich, lmtest
Suggests:	MASS, parallel, Amelia, mice
Enhances:	texreg, stargazer
URL:	https://github.com/atahk/bucky
Description:	Provides functions for various statistical techniques commonly used in the social sciences, including functions
License:	GPL (>= 3)

Index of help topics:

```
bucky-package           Bucky's Archive for Data Analysis in the Social
                        Sciences
mi.eval                 Multiple-imputation evaluation
robust.summary          Robust summary
robustify               Robustify a model
summary.mi.estimates    Summary for multiple imputation
vcovCR                  Clustered Robust Covariance Matrix Estimation
```

Author(s)

Alexander Tahk [aut, cre] (<https://orcid.org/0000-0001-7895-9420>)

Maintainer: Alexander Tahk <tahk@alum.mit.edu>

References

Cameron, A. Colin, and Douglas L. Miller. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50, no. 2 (Spring 2015): 317-372. doi: 10.3368/jhr.50.2.317

Little, Roderick J. A., and Donald B. Rubin. "The Analysis of Social Science Data with Missing Values." *Sociological Methods & Research* 18, no. 2-3 (November 1989): 292-326. doi: 10.1177/0049124189018002004

Rubin, Donald. B. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons, 2004.

See Also

See also `mice`, `amelia-package`, `MIcombine`, `sandwich` and `coeftest`.

`mi.eval` *Multiple-imputation evaluation*

Description

Evaluation of an expression across multiply imputed data sets.

Usage

```
mi.eval(EXPR, robust, cluster, coef., vcov., df.=NULL,
parallel=FALSE, lazy=NULL, ...)
```

Arguments

`EXPR`	An R expression to evaluate. This expression must contain a `data` argument that specifies a list containing the imputed data sets or a an object of class `amelia`, `mids`, or `imputationList`.
`robust`	Whether to use Huber-White robust standard errors. The default is `TRUE` if `cluster` is specified and `FALSE` otherwise.
`cluster`	A vector specifying clusters for the purpose of computing clustered robust standard errors. This can be a variable inside the imputed data set. If unspecified, standard errors are not clustered. If specified, `robust` cannot be `FALSE`.
`coef.`	The function used to get a numeric vector of coefficient estimates when evaluated on an object returned from evaluating `EXPR` for each data set. The default is to use `coef`.
`vcov.`	The function that returns a numeric matrix giving the variance-covariance matrix when evaluated on an object returned from evaluating `EXPR` for each data set. The default is to use `vcovCR` if `cluster` is specified, `vcovHC` if `robust=TRUE` and `cluster` is not specified, and `vcov` otherwise.
`df.`	Either the degrees of freedom for each model or a function that calculates degrees of freedom on an object returned from evaluating `EXPR` for each data set. The default value of `NULL` uses the minimum result of applying `df.residual` it returns a numeric value when applied to the object returned by `EXPR` and this object is not of class `glm` and `Inf` otherwise.

`parallel`	A logical indicating whether to evaluate `EXPR` across data sets in parallel using `mclapply`. Otherwise, evaluation is done serially using `lapply`. `NULL` means to use parallel evaluation if and only if the 'parallel' package can be loaded and `getOption("mc.cores", detectCores()-1L)` is greater than 1.
`lazy`	A logical indicating whether to use lazy evaluation to avoid copying all imputed data sets into memory. When the `data` argument to `EXPR` generates the multiply imputed data set, this is generally a bad idea because it means redoing the imputation multiple times. The default value of `NULL` means to use lazy evaluation if and only if the `data` argument to `EXPR` is a `name`.
`...`	Any additional arguments to be passed to `lapply` or `mclapply` when evaluating `EXPR` across data sets.

Details

This function evaluates a R command for each of several multiply imputed data sets and combines results across data sets into a single set of estimates. This is similar to the functionality provided by `with.mids` but also works with multiply-imputed data sets generated by other packages like 'Amelia' as well as those from 'mice'.

For generating formatted tables of regression coefficients, the outputted objects should be compatible with the 'texreg' package. When used with `lm`, `glm` or a few other types of models, these objects are also compatible with the 'stargazer' package.

Value

An object of class `mi.estimates` containing the coefficient estimates, variance-covariance matrix, and related information.

See Also

See Also `summary.mi.estimates`, `with.mids`, `amelia`, `mice`, `coef`, `vcov` and `df.residual`.

Examples

```
if (require("Amelia")) {
    ## Load data
    data(africa)
    africa$civlib <- factor(round(africa$civlib*6), ordered=TRUE)

    ## Estimate a linear model using imputed data sets
    model0 <- lm(trade ~ log(gdp_pc), data=africa, subset=year==1973)
    summary(model0)

    ## Impute using Amelia
    a.out <- amelia(x = africa, cs = "country", ts = "year",
                    logs = "gdp_pc", ord="civlib")

    ## Estimate a linear model using imputed data sets
    model1 <- mi.eval(lm(trade ~ log(gdp_pc), data=a.out, subset=year==1973))

    ## Show estimates
```

```
    model1
    coef(model1)

    ## Show summary information
    summary(model1)

    if (require("MASS")) {
        ## Estimate an ordered logit model
        model2 <- mi.eval(polr(civlib ~ log(gdp_pc) + log(population),
                               data=a.out))
        summary(model2)

        ## Also show thresholds by including thresholds with coefficients
        model3 <- mi.eval(polr(civlib ~ log(gdp_pc) + log(population),
                               data=a.out),
                          coef=function(x) c(x$coefficients, x$zeta))
        summary(model2)
    }
}
```

`robust.summary` *Robust summary*

Description

Output summary information using robust or clustered robust standard errors.

Usage

```
## S3 method for class 'robustified'
summary(object, ...)

robust.summary(x, cluster, type, omega, ...)
```

Arguments

`object`	An object of class `robustified`, usually generated by `robustify`.
`x`	A model of class `lm`, `glm`, or any other class which contains a `call` object and methods for `estfun` and `nobs`.
`cluster`	The variable on which to cluster (if any). If this is not specified, unclustered robust standard errors using `vcovHC` are used. If this is specified, clustered robust standard errors using `vcovCR` are used.
`type`	A character string specifying the estimation type. The default is to use the defaults for `vcovHC` or `vcovCR`, depending on whether `cluster` is specified. For details, see `vcovHC` or `vcovCR`.
`omega`	A vector or a function depending on the arguments ‘residuals’ (the working residuals of the model), ‘diaghat’ (the diagonal of the corresponding hat matrix) and ‘df’ (the residual degrees of freedom). For details, see `vcovHC` or `vcovCR`.
`...`	Any additional arguments to be passed to `coeftest`.

Details

Both functions provide summary output with robust (Huber-White) or clustered robust standard errors based on `vcovHC` or `vcovCR`, respectively. The `summary` method works on objects where the type of the standard errors has already been set by `robustify`. The `robust.summary` function works on unadjusted objects. Thus, `robust.summary(x, ...)` is a shorthand for `summary(robustify(x, ...))`.

For `robust.summary`, if the `cluster` option is specified, clustered robust standard errors are used based on the variance-covariance matrix from `vcovCR` with clustering on `cluster`. If not, robust standard errors are used based on the variance-covariance matrix from `vcovHC`.

Value

An object of class `summary.robustified` containing a `coefficients` object computed using `coeftest` and the `method` attribute specifying the type of standard errors used.

See Also

See Also `robustify`, `vcovHC`, `vcovCR` and `coeftest`.

Examples

```
## With clustering
clotting <- data.frame(
    cl = 1:9,
    u = c(5,10,15,20,30,40,60,80,100),
    lot = c(118,58,42,35,27,25,21,19,18,
            69,35,26,21,18,16,13,12,12))
clot.model <- glm(lot ~ log(u), data = clotting, family = Gamma)
robust.summary(clot.model, cluster=cl)

## Without clustering
data(swiss)
model1 <- lm(Fertility ~ ., data = swiss)
robust.summary(model1)
model1r <- robustify(model1)
summary(model1r)
```

`robustify` *Robustify a model*

Description

Create a robustified object that includes robust or clustered robust standard errors.

Usage

```
robustify(x, cluster, type, omega, ...)
```

Arguments

`x`	A model of class `lm`, `glm`, or any other class which contains a `call` object and methods for `estfun` and `nobs`.
`cluster`	The variable on which to cluster (if any). If this is not specified, unclustered robust standard errors using `vcovHC` are used. If this is specified, clustered robust standard errors using `vcovCR` are used.
`type`	A character string specifying the estimation type. The default for linear models of class `lm` but not `glm` is to use "HC1" for `vcovHC` or "CR1" for `vcovCR`, depending on whether `cluster` is specified. For other models, the default is to use "HC" for `vcovHC` or "CR" for `vcovCR`, depending on whether `cluster` is specified. This mirrors the defaults used by Stata as closely as possible. For details, see `vcovHC` or `vcovCR`.
`omega`	A vector or a function depending on the arguments ‘residuals’ (the working residuals of the model), ‘diaghat’ (the diagonal of the corresponding hat matrix) and ‘df’ (the residual degrees of freedom). For details, see `vcovHC` or `vcovCR`.
`...`	Any additional arguments to be passed to `coeftest`.

Details

This function creates a `robustified` object containing the model, coefficients, and variance-covariance matrix based on `vcovHC` or `vcovCR`, respectively. If the `cluster` option is specified, the variance-covariance matrix is computed using `vcovCR` with clustering on `cluster`. If not, the variance-covariance matrix is computed using `vcovHC`. For generating formatted tables of regression coefficients, the outputted objects should be compatible with the ’texreg’ package. When used with `lm`, `glm` or a few other types of models, these objects are also compatible with the ’stargazer’ package.

Value

An object of class `robustified` with the `method` attribute specifying the type of standard errors used.

References

Cameron, A. Colin, and Douglas L. Miller. “A Practitioner’s Guide to Cluster-Robust Inference.” *Journal of Human Resources* 50, no. 2 (Spring 2015): 317-372. doi: 10.3368/jhr.50.2.317

See Also

See Also `summary.robustified`, `vcovHC`, `vcovCR` and `coeftest`.

Examples

```
## With clustering
clotting <- data.frame(
    cl = 1:9,
    u = c(5,10,15,20,30,40,60,80,100),
    lot = c(118,58,42,35,27,25,21,19,18,
            69,35,26,21,18,16,13,12,12))
```

```
clot.model <- glm(lot ~ log(u), data = clotting, family = Gamma)
robust.clot.model <- robustify(clot.model, cluster=cl)
robust.clot.model
summary(robust.clot.model)

## Without clustering
data(swiss)
model1 <- robustify(lm(Fertility ~ ., data = swiss))
model1
summary(model1)
```

`summary.mi.estimates` *Summary for multiple imputation*

Description

Output summary information for estimates computed on multiply imputed data sets.

Usage

```
## S3 method for class 'mi.estimates'
summary(object, ...)
```

Arguments

`object`	An object of class `mi.estimates`, usually computed with `mi.eval`.
`...`	Any additional arguments to be passed to `coeftest`.

Details

This function provides summary output for models estimated using multiple imputation using `mi.eval`.

Value

An object of class `summary.mi.estimates` containing a `coefficients` object of class `coeftest` and other summary information.

See Also

See Also `mi.eval` and `amelia`.

Examples

```
if (require("Amelia")) {
    data(africa)
    a.out <- amelia(x = africa, cs = "country", ts = "year", logs = "gdp_pc")

    model <- mi.eval(lm(civlib ~ log(gdp_pc), data=a.out, subset=year==1973))

    summary(model)
}
```

`vcovCR` *Clustered Robust Covariance Matrix Estimation*

Description

Robust estimation of the covariance matrix of the coefficient estimates in regression models with clustering.

Usage

```
vcovCR(x, cluster = NULL, type = c("CR", "CR0", "CR1"))
```

Arguments

`x`	A fitted model object.
`cluster`	A variable or expression giving the cluster for each observation.
`type`	A character string specifying the estimation type. For details see below.

Details

The default `type` of "CR" uses the same adjustment as 'Stata'. The values of "CR0" and "CR1" are analogous to "HC0" and "HC1", respectively, in `vcovHC`.

Value

A matrix containing the covariance matrix estimate with attribute `type` giving the `type` option used in estimating it.

See Also

See Also `robust.summary` and `vcovHC`.

Examples

```
clotting <- data.frame(
    cl = rep(1:2,each=9),
    u = c(5,10,15,20,30,40,60,80,100),
    lot = c(118,58,42,35,27,25,21,19,18,
            69,35,26,21,18,16,13,12,12))
clot.model <- glm(lot ~ log(u), data = clotting, family = Gamma)
vcovCR(clot.model, cluster=cl)

data(swiss)
model1 <- lm(Fertility ~ ., data = swiss)
## These should give the same answer
vcovCR(model1, cluster=1:nobs(model1), type="CR0")
sandwich::vcovHC(model1, type="HC0")
```

Index

∗ **cluster**
  robust.summary, 5
  robustify, 6
  vcovCR, 9
∗ **htest**
  mi.eval, 3
  robust.summary, 5
  robustify, 6
  summary.mi.estimates, 8
∗ **models**
  mi.eval, 3
∗ **model**
  summary.mi.estimates, 8
∗ **package**
  bucky-package, 2
∗ **robust**
  robust.summary, 5
  robustify, 6
  vcovCR, 9

amelia, *3*, *4*, *8*

bucky *(*bucky-package*)*, 2
bucky-package, 2

coef, *3*, *4*
coeftest, *3*, *5–7*

df.residual, *3*, *4*

imputationList, *3*

lapply, *4*

mclapply, *4*
mi.eval, 3, *8*
mice, *3*, *4*
MIcombine, *3*
mids, *3*

name, *4*

robust.summary, 5, *9*
robustify, *5*, *6*, 6

sandwich, *3*
summary.mi.estimates, *4*, 8
summary.robustified, *7*
summary.robustified *(*robust.summary*)*, 5

vcov, *3*, *4*
vcovCR, *3*, *5–7*, 9
vcovHC, *3*, *5–7*, *9*

with.mids, *4*