Package 'mKBO'

October 31, 2025

Type Package

Title Multi-Group Kitagawa-Blinder-Oaxaca Decomposition

Version 0.1.0

Description Provides multigroup Kitagawa-Blinder-Oaxaca ('mKBO') decompositions, that allow for more than two groups. Each group is compared to the sample average. For more details see Thaning and Nieuwenhuis (2025) <doi:10.31235/osf.io/6twvj_v1>.

License CC BY 4.0

Encoding UTF-8

LazyData true

Depends R (>= 3.5.0)

Imports dplyr, tidyr, purrr, stats, tibble, tidyselect, stringr, rlang, broom, utils

RoxygenNote 7.3.3

NeedsCompilation no

Author Rense Nieuwenhuis [cre, aut] (ORCID: <https://orcid.org/0000-0001-6138-0463>), Max Thaning [aut] (ORCID: <https://orcid.org/0000-0002-1859-4703>)

Maintainer Rense Nieuwenhuis <rense.nieuwenhuis@sofi.su.se>

Repository CRAN

Date/Publication 2025-10-31 18:20:13 UTC

Contents

mkbo *Estimate the mKBO decomposition*

Description

This is the main function, computing the multi-group Kitagawa-Blinder-Oaxaca decomposition

Usage

```
mkbo(formula, group, w = NULL, data, group_fixed = TRUE, viewpoint = "group")
```

Arguments

formula	A regression formula (as a string) specifying the outcome and explanatory variables.
group	A string naming the grouping variable. This variable should be a factor, and the decomposition will be performed for each level of this factor.
w	A string naming the variable in data that contains observation weights. If NULL, equal weights are used.
data	A data.frame or tibble containing the microdata. The data must not contain missing values in any of the variables used in the decomposition.
group_fixed	Logical. If TRUE (default), group fixed effects are included in the pooled model used to estimate the sample-level coefficients.
viewpoint	Character. Either "group"or "sample". Specifies the decomposition perspective: • "group": How would group outcomes change if they had the endowments/coefficient structure of the full sample? • "sample": How do group characteristics differ from the sample, and how much does this explain outcome differences?

Details

The function performs group-wise regressions and compares them to a pooled regression model. It decomposes the differences in group means of the dependent variable into parts due to differences in observed characteristics (endowments), differences in how those characteristics translate into outcomes (coefficients), and the interaction of both.

The choice of viewpoint changes whether the decomposition is anchored on the sample or group averages, and this influences the interpretation of each component.

Group-specific coefficients are augmented with treatment contrasts to match the pooled model structure.

Value

An object of class "mkbo", which is a list containing:

- RECI A tibble summarizing the mean outcome per group (M), mean difference from the reference (D), and contributions from endowments (E), coefficients (C), and interactions (I).
- E_var A data frame detailing variable-level contributions to the endowments (E) component.
- C_var A data frame detailing variable-level contributions to the coefficients (C) component.
- I_var A data frame detailing variable-level contributions to the interaction (I) component.

Examples

```
mkbo_output <- mkbo("PERNP ~ BACHELOR", group = "RACE", data=pums_subset)
```

---

mkbo_summary *Summarize Components of an mKBO Decomposition*

Description

Provides a summary of the modified Kitagawa-Blinder-Oaxaca (mKBO) decomposition for selected model terms or grouped categories of terms. This function is useful for inspecting how specific variables (or sets of variables) contribute to the overall decomposition across groups.

Usage

```
mkbo_summary(x, term = NULL, term.cat = NULL)
```

Arguments

x	An object of class "mkbo", as returned by the mkbo function.
term	A character vector specifying one or more model terms (e.g., variable names) for which to summarize decomposition results. Use this argument to inspect contributions from specific variables.
term.cat	A character string specifying a common prefix for a group of terms (typically dummy variables from a factor). This will summarize the decomposition results for all terms that match this pattern (e.g., "education" will match "education.Bachelor", "education.Master", etc.).

Value

A data.frame with one row per group and the following columns:

- group Group identifier (from the grouping variable used in mkbo).
- M Group-specific mean of the outcome variable.
- D Difference from the reference (sample mean), as used in the mKBO decomposition.
- R Total explained difference (sum of E + C + I components).
- E Component of the difference due to endowments (differences in covariates).
- C Component due to coefficients (differences in effects).
- I Interaction component (joint difference in covariates and coefficients, conditional on E and C).

Examples

```
mkbo_output <- mkbo("PERNP ~ BACHELOR", group = "RACE", data=pums_subset)
mkbo_summary(mkbo_output, term="BACHELORTRUE")
```

`mkbo_triangle` *Calculates Triangle*

Description

This function calculates all group-differences

Usage

```
mkbo_triangle(
  mkbo_output,
  term = NULL,
  term.cat = NULL,
  components = c("E"),
  percentage = TRUE,
  absolute_gaps = TRUE
)
```

Arguments

`mkbo_output`	placeholder text
`term`	Specify the model term for which the mKBO results should be presented. Can be a vector of terms to present the summed results for those terms. Should be specified in quotation marks.
`term.cat`	Specify a factor variable for which to sum the mKBO results across categories. Should be specified in quotation marks.
`components`	Specify+he decomposition components to be included in the calculation. Can be any combination of c("E", "C", "I"), or "R".
`percentage`	Specify to express the changes in gap as percentage (default) or in absolute differences (expressed in unites of the dependent variables).
`absolute_gaps`	If TRUE, the changes in gaps are expressed in absolute terms even when signs change.

Value

An object of class `tibble`, containing absolute or relative group-differences explained by the variables specified in `mKBO`.

Examples

```
mkbo_output <- mkbo("PERNP ~ BACHELOR", group = "RACE", data=pums_subset)
mkbo_triangle(mkbo_output, term="BACHELORTRUE")
```

pums_subset *American Community Survey (ACS) Public Use Microdata Sample (PUMS)*

Description

This is a 5% random sample of the 2023 subset of the American Community Survey (ACS) Public Use Microdata Sample (PUMS)

Usage

```
pums_subset
```

Format

'pums_subset' A data frame with 71,919 rows and 7 columns:

PERNP Annual Earnings of a person.

RACE Race, in 9 categories: "Alaska Native", "American Indian","Asian","Black","Mixed", "Pacific", "Tribe Specific", "White", and "Other"

BACHELOR Binary indicator of whether an individuals has a Bachelor's degree of higher.

AGEP Age.

PWGTP Person weight. ...

Source

<https://www.census.gov/programs-surveys/acs/microdata/access.html>

Index