Package ‘DisImpact’

April 3, 2020

Title Calculates Disproportionate Impact When Binary Success Data are Disaggregated by Subgroups

Version 0.0.7

Description Implements methods for calculating disproportionate impact: the percentage point gap, proportionality index, and the 80% index.

Depends R (>= 3.4.0)
Imports dplyr, rlang, tidyselect, purrr, tidyr
License GPL-3

URL https://github.com/vinhdizzo/DisImpact

BugReports https://github.com/vinhdizzo/DisImpact/issues

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Author Vinh Nguyen [aut, cre]

Maintainer Vinh Nguyen <nguyenvq714@gmail.com>

Repository CRAN

Date/Publication 2020-04-03 04:10:06 UTC
R topics documented:

di_80_index ......................................................... 2
di_ppg .............................................................. 3
di_ppg_iterate .................................................... 5
di_prop_index ...................................................... 6
ppg_moe ............................................................. 8
student_equity ..................................................... 9

Index 10

di_80_index  Calculate disproportionate impact per the 80% index

Description

Calculate disproportionate impact per the 80% index method.

Usage

di_80_index(success, group, cohort, weight, data)

Arguments

success  A vector of success indicators (1/0 or TRUE/FALSE) or an unquoted reference (name) to a column in data if it is specified. It could also be a vector of counts, in which case weight should also be specified (group size).

group  A vector of group names of the same length as success or an unquoted reference (name) to a column in data if it is specified.

cohort  (Optional) A vector of cohort names of the same length as success or an unquoted reference (name) to a column in data if it is specified. Disproportionate impact is calculated for every group within each cohort. When cohort is not specified, then the analysis assumes a single cohort.

weight  (Optional) A vector of case weights of the same length as success or an unquoted reference (name) to a column in data if it is specified. If success consists of counts instead of success indicators (1/0), then weight should also be specified to indicate the group size.

data  (Optional) A data frame containing the variables of interest. If data is specified, then success, group, and cohort will be searched within it.

Details

This function determines disproportionate impact based on the 80% index method, as described in this reference from the California Community Colleges Chancellor’s Office. It assumes that a higher rate is good ("success"). For rates that are deemed negative (eg, rate of drop-outs, high is bad), then consider looking at the converse of the non-success (eg, non drop-outs, high is good) instead in order to leverage this function properly.
di_ppg

Value

A data frame consisting of: cohort (if used), group, n (sample size), success (number of successes for the cohort-group), pct (proportion of successes for the cohort-group), di_80_index (ratio of pct to the max pct for each cohort), and di_indicator (1 if di_80_index < 0.80).

References


Examples

library(dplyr)
data(student_equity)
di_80_index(success=Transfer, group=Ethnicity, data=student_equity) %>%
as.data.frame

---

di_ppg

*Calculate disproportionate impact per the percentage point gap (PPG) method.*

Description

Calculate disproportionate impact per the percentage point gap (PPG) method.

Usage

```
di_ppg(success, group, cohort, weight, reference = c("overall", "hpg"),
  data, min_moe = 0.03, use_prop_in_moe = FALSE, prop_sub_0 = 0.5,
  prop_sub_1 = 0.5)
```

Arguments

- **success**: A vector of success indicators (1/0 or TRUE/FALSE) or an unquoted reference (name) to a column in data if it is specified. It could also be a vector of counts, in which case weight (group size) should also be specified.
- **group**: A vector of group names of the same length as success or an unquoted reference (name) to a column in data if it is specified.
- **cohort**: (Optional) A vector of cohort names of the same length as success or an unquoted reference (name) to a column in data if it is specified. Disproportionate impact is calculated for every group within each cohort. When cohort is not specified, then the analysis assumes a single cohort.
- **weight**: (Optional) A vector of case weights of the same length as success or an unquoted reference (name) to a column in data if it is specified. If success consists of counts instead of success indicators (1/0), then weight should also be specified to indicate the group size.
reference

Either 'overall' (default), 'hpg' (highest performing group), a single proportion (eg, 0.50), or a vector of proportions. Reference is used as a point of comparison for disproportionate impact for each group. When cohort is specified: 1. 'overall' will use the overall success rate of each cohort group as the reference; 2. 'hpg' will use the highest performing group in each cohort as reference; 3. the specified proportion will be used for all cohorts; 4. the specified vector of proportions will refer to the reference point for each cohort in alphabetical order (so the number of proportions should equal to the number of unique cohorts).

data

(Optional) A data frame containing the variables of interest. If data is specified, then success, group, and cohort will be searched within it.

min_moe

The minimum margin of error (MOE) to be used in the calculation of disproportionate impact and is passed to ppg_moe. Defaults to 0.03.

use_prop_in_moe

A logical value indicating whether or not the MOE formula should use the observed success rates (TRUE). Defaults to FALSE, which uses 0.50 as the proportion in the MOE formula. If TRUE, the success rates are passed to the proportion argument of ppg_moe.

prop_sub_0

For cases where 'proportion' is 0, substitute with 'prop_sub_0' (defaults to 0.5) to account for the zero MOE. This is relevant only when 'use_prop_in_moe=TRUE'.

prop_sub_1

For cases where 'proportion' is 1, substitute with 'prop_sub_1' (defaults to 0.5) to account for the zero MOE. This is relevant only when 'use_prop_in_moe=TRUE'.

Details

This function determines disproportionate impact based on the percentage point gap (PPG) method, as described in this reference from the California Community Colleges Chancellor’s Office. It assumes that a higher rate is good ("success"). For rates that are deemed negative (eg, rate of drop-outs, high is bad), then consider looking at the converse of the non-success (eg, non drop-outs, high is good) instead in order to leverage this function properly. Note that the margin of error (MOE) is calculated using using 1.96*sqrt(0.25^2/n), with a min_moe used as the minimum by default.

Value

A data frame consisting of: cohort (if used), group, n (sample size), success (number of successes for the cohort-group), pct (proportion of successes for the cohort-group), reference (reference used in DI calculation), moe (margin of error), pct_lo (lower 95% confidence interval for pct), pct_hi (upper 95% confidence interval for pct), and di_indicator (1 if there is disproportionate impact, ie, when pct_hi <= reference).

References


Examples

library(dplyr)
data(student_equity)
# di_ppg_iterate

Iteratively calculate disproportionate impact via the percentage point gap (PPG) method for many variables.

## Description

Iteratively calculate disproportionate impact via the percentage point gap (PPG) method for many disaggregation variables.

## Usage

```r
di_ppg_iterate(data, success_vars, group_vars, cohort_vars, reference_groups, repeat_by_vars = NULL, min_moe = 0.03, use_prop_in_moe = FALSE, prop_sub_0 = 0.5, prop_sub_1 = 0.5)
```

## Arguments

- **data**: A data frame for which to iterate DI calculation for a set of variables.
- **success_vars**: A character vector of success variable names to iterate across.
- **group_vars**: A character vector of group (disaggregation) variable names to iterate across.
- **cohort_vars**: A character vector of cohort variable names to iterate across.
- **reference_groups**: 
- **repeat_by_vars**: 
- **min_moe**: 
- **use_prop_in_moe**: 
- **prop_sub_0**: 
- **prop_sub_1**: 

---

# di_ppg_iterate

Iteratively calculate disproportionate impact via the percentage point gap (PPG) method for many variables.

## Description

Iteratively calculate disproportionate impact via the percentage point gap (PPG) method for many disaggregation variables.

## Usage

```r
di_ppg_iterate(data, success_vars, group_vars, cohort_vars, reference_groups, repeat_by_vars = NULL, min_moe = 0.03, use_prop_in_moe = FALSE, prop_sub_0 = 0.5, prop_sub_1 = 0.5)
```

## Arguments

- **data**: A data frame for which to iterate DI calculation for a set of variables.
- **success_vars**: A character vector of success variable names to iterate across.
- **group_vars**: A character vector of group (disaggregation) variable names to iterate across.
- **cohort_vars**: A character vector of cohort variable names to iterate across.
Either 'overall', 'hpg', or a character vector of the same length as 'group_vars' that indicates the reference group value for each group variable in 'group_vars'.

A character vector of variables to repeat DI calculations for across all combination of these variables, including '- All' as a group for each variable. The reference rate used for DI comparison differs for every combination of the variables listed here.

The minimum margin of error to be used in the PPG calculation, passed to 'di_ppg'.

Whether the estimated proportions should be used in the margin of error calculation by the PPG, passed to 'di_ppg'.

Passed to 'di_ppg'.

Passed to 'di_ppg'.

Iteratively calculate disproportionate impact via the percentage point gap (PPG) method for all combinations of 'success_vars', 'group_vars', and 'cohort_vars', for each combination of subgroups specified by 'repeat_by_vars'.

A data frame with all relevant returned fields from 'di_ppg' plus 'success_variable' (elements of 'success_vars'), 'disaggregation' (elements of 'group_vars'), and 'reference_group' (elements of 'reference_groups').

library(dplyr)
data(student_equity)
# Multiple group variables
di_ppg_iterate(data=student_equity, success_vars=c('Transfer'),
               group_vars=c('Ethnicity', 'Gender'), cohort_vars=c('Cohort'),
               reference_groups='overall')

di_prop_index

Calculate disproportionate impact per the proportionality index (PI) method.

Calculate disproportionate impact per the proportionality index (PI) method.

di_prop_index(success, group, cohort, weight, data)
di_prop_index

Arguments

success  A vector of success indicators (1/0 or TRUE/FALSE) or an unquoted reference (name) to a column in data if it is specified. It could also be a vector of counts, in which case weight should also be specified (group size).

group    A vector of group names of the same length as success or an unquoted reference (name) to a column in data if it is specified.

cohort   (Optional) A vector of cohort names of the same length as success or an unquoted reference (name) to a column in data if it is specified. disproportionate impact is calculated for every group within each cohort. When cohort is not specified, then the analysis assumes a single cohort.

weight   (Optional) A vector of case weights of the same length as success or an unquoted reference (name) to a column in data if it is specified. If success consists of counts instead of success indicators (1/0), then weight should also be specified to indicate the group size.

data     (Optional) A data frame containing the variables of interest. If data is specified, then success, group, and cohort will be searched within it.

Details

This function determines disproportionate impact based on the proportionality index (PI) method, as described in this reference from the California Community Colleges Chancellor’s Office. It assumes that a higher rate is good (“success”). For rates that are deemed negative (eg, rate of drop-outs, high is bad), then consider looking at the converse of the non-success (eg, non drop-outs, high is good) instead in order to leverage this function properly.

Value

A data frame consisting of: cohort (if used), group, n (sample size), success (number of successes for the cohort-group), pct_success (proportion of successes attributed to the group within the cohort), pct_group (proportion of sample attributed to the group within the cohort), and di_prop_index (ratio of pct_success to pct_group). When di_prop_index < 1, then there are signs of disproportionate impact.

References


Examples

library(dplyr)
data(student_equity)
di_prop_index(success=Transfer, group=Ethnicity, data=student_equity) %>%
  as.data.frame
ppg_moe

Margin of error for the PPG

Description
Calculate the margin of error (MOE) for the percentage point gap (PPG) method.

Usage
ppg_moe(n, proportion, min_moe = 0.03, prop_sub_0 = 0.5, prop_sub_1 = 0.5)

Arguments
- **n**: Sample size for the group of interest.
- **proportion**: (Optional) The proportion of successes for the group of interest. If specified, then the proportion is used in the MOE formula. Otherwise, a default proportion of 0.50 is used (conservative and yields the maximum MOE).
- **min_moe**: The minimum MOE returned even if the sample size is large. Defaults to 0.03. This equates to a minimum threshold gap for declaring disproportionate impact.
- **prop_sub_0**: For cases where ‘proportion’ is 0, substitute with ‘prop_sub_0’ (defaults to 0.5) to account for the zero MOE.
- **prop_sub_1**: For cases where ‘proportion’ is 1, substitute with ‘prop_sub_1’ (defaults to 0.5) to account for the zero MOE.

Value
The margin of error for the PPG given the specified sample size.

References

Examples
ppg_moe(n=800)
ppg_moe(n=c(200, 800, 1000, 2000))
ppg_moe(n=800, proportion=0.20)
ppg_moe(n=800, proportion=0.20, min_moe=0)
ppg_moe(n=c(200, 800, 1000, 2000), min_moe=0.01)
Fake data on student equity

Description

Data randomly generated to illustrate the use of the package.

Usage

data(student_equity)

Format

An object of class tbl_df (inherits from tbl.data.frame) with 20000 rows and 4 columns.

Examples

data(student_equity)
Index

*Topic datasets
  student_equity, 9

di_80_index, 2
di_ppg, 3
di_ppg_iterate, 5
di_prop_index, 6

ppg_moe, 4, 8

student_equity, 9