Package 'sampsizeval'

October 14, 2022

Title Sample Size for Validation of Risk Models with Binary Outcomes

Version 1.0.0.0

Description Estimation of the required sample size to validate a risk model for binary outcomes, based on the sample size equations proposed by Pavlou et al. (2021) <doi:10.1177/09622802211007522>. For precision-based sample size calculations, the user is required to enter the anticipated values of the C-statistic and outcome prevalence, which can be obtained from a previous study. The user also needs to specify the required precision (standard error) for the C-statistic, the calibration slope and the calibration in the large. The calculations are valid under the assumption of marginal normality for the distribution of the linear predictor.

License MIT + file LICENSE

LazyData false

RoxygenNote 7.1.1

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

Imports stats, sn, pracma, dplyr, plyr

URL https://github.com/mpavlou/sampsizeval

BugReports https://github.com/mpavlou/sampsizeval/issues

NeedsCompilation no

Author Menelaos Pavlou [aut, cre] (<https://orcid.org/0000-0003-1161-1440>)

Maintainer Menelaos Pavlou <m.pavlou@ucl.ac.uk>

Repository CRAN

Date/Publication 2021-05-28 07:30:02 UTC

R topics documented:

| sampsizeval | *sampsizeval: Estimation of Required Sample Size for Validation of Risk Models for Binary Outcomes* |

Description

This function calculates the sample size required in the validation dataset to estimate the C-statistic (C), the calibration Slope (CS) and the Calibration in the Large (CL) with sufficient precision. It takes as arguments the anticipated values of the C-statistic and the outcome prevalence (obtained, for example, from a previous study) and the required standard error for C, CS and CL.

Usage

```
sampsizeval(p, c, se_c, se_cs, se_cl, c_ni = FALSE)
```

Arguments

p	(numeric) The anticipated outcome prevalence, a real number between 0 and 1
c	(numeric) The anticipated C-statistic, a real number between 0.5 and 1
se_c	(numeric) The required standard error of the estimated C-Statistic
se_cs	(numeric) The required standard error of the estimated Calibration Slope
se_cl	(numeric) The required standard error of the estimated Calibration in the Large
c_ni	(logical) Numerical integration is used for the calculations for C-statistic (TRUE) or the closed-form expression (FALSE). Default value is 'FALSE'

Details

The sample size calculations are valid under the assumption of marginal normality for the distribution of the linear predictor.The default sample size calculation based on C uses the closed-form expression in equation (9) as proposed by *Pavlou et al. (2021)*. This is quick to run and accurate for all values of anticipated C and p.The default sample size calculations based on CS and CL use the formulae (12) and (13) that require the use of numerical integration. The parameters of the assumed Normal distribution used in the latter two expressions are obtained using equations (7) and (8) and are fine-tuned for values of anticipated C>0.8.

Sample size calculations from the estimator based on C that uses numerical integration can also be obtained.

Value

size_c: the sample size based on the C-statistic

size_cs: the sample size based on the Calibration Slope

size_cl: the sample size based on the Calibration in the Large

size_recommended: the final sample size recommendation (the largest of the three above)

References

Pavlou M, Chen Q, Omar ZR, Seaman RS, Steyerberg WE, White RI, Ambler G. Estimation of required sample size for external validation of risk models for binary outcomes, SMMR (2021). doi:10.1177/09622802211007522

Examples

```
# Calculate the sample size of the validation data to estimate the
# C-statistic, the Calibration slope and the Calibration in the Large with
# sufficient precision. It is assumed that the anticipated prevalence is 0.1
# and the C-statistic is 0.75. The required SE for the C statistic is 0.025
# (corresponding to a confidence interval of width approximately 0.1) and the
# required SE for the calibration slope and calibration in the large is 0.1
# (corresponding to a confidence interval of width approximately 0.4).

sampsizeval(p=0.1, c=0.75, se_c=0.025, se_cs =0.1, se_cl = 0.1)
```

Index