Package 'pmsesampling'

September 9, 2025

Title Sample Size Determination for Accurate Predictive Linear Regression

Version 0.1.1

Description Provides analytic and simulation tools to estimate the minimum sample size required for achieving a target prediction mean-squared error (PMSE) or a specified proportional PMSE reduction (pPMSEr) in linear regression models. Functions implement the criteria of Ma (2023) <https://digital.wpi.edu/downloads/0g354j58c>, support covariance-matrix handling, and include helpers for root-finding and diagnostic plotting.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.2

Imports Matrix, stats, rootSolve

Suggests rmarkdown, testthat(>= 3.0.0)

BugReports https://github.com/Chenaters/pmsesampling/issues

URL https://github.com/Chenaters/pmsesampling

NeedsCompilation no

Author Louis Chen [aut, cre], Zheyang Wu [aut, ths]

Maintainer Louis Chen <chenaters@gmail.com>

Repository CRAN

Date/Publication 2025-09-09 14:00:02 UTC

Contents

`pmsesampling-package` *pmsesampling: Sample Size Determination for Accurate Predictive Linear Regression*

Description

Tools to estimate the minimum sample size required to achieve a target Prediction Mean-Squared Error (PMSE) or a specified proportional PMSE reduction (pPMSEr). Functions implement the analytic and simulation-based criteria described in Ma (2023) and include helpers for covariance-matrix handling, root-finding and diagnostic plotting.

Core functions

`pmse_samplesize()` Determines sample size from PMSE equation in basic and full models and the efficient sample size

Typical workflow

1. Obtain σ_k^2 and σ_p^2
2. Or import or build a predictor covariance matrix.
3. Or obtain $Cohen's f^2$ and R^2
4. Call `pmse_samplesize` with available inputs to get sample size.

Author(s)

Maintainer: Louis Chen <chenaters@gmail.com>

Authors:

- Zheyang Wu <zheyangwu@wpi.edu> [thesis advisor]

References

Ma Y. (2023) *Predictive Power and Efficient Sample Size in Linear Regression Models*. Worchester Polytechnic Institute

See Also

Useful links:

- https://github.com/Chenaters/pmsesampling
- Report bugs at https://github.com/Chenaters/pmsesampling/issues

`pmse_samplesize` *Compute efficient sample size under user-defined PMSE targets*

Description

`pmse_samplesize` computes a sample size for a prediction model. The function implements the formulas found in the thesis "Predictive Power and Efficient Sample Size in Linear Regression Models" by Yifan Ma (2023).

Usage

```
pmse_samplesize(
  k,
  p,
  PMSE_val_k = 1,
  PMSE_val_p = 1,
  efficiency_level = 0.9,
  sigma_k2 = NULL,
  sigma_p2 = NULL,
  cov = NULL,
  corr = NULL,
  SD = 1,
  f2 = NULL,
  f2_2 = NULL,
  R2_full = NULL,
  R2_basic = NULL
)
```

Arguments

`k`	Integer. Total number of predictors in the full model.
`p`	Integer. Number of basic predictors in the reduced model.
`PMSE_val_k`	Numeric. Target PMSE value for the full model.
`PMSE_val_p`	Numeric. Target PMSE value for the reduced model.
`efficiency_level`	Numeric. Target efficiency level. (default is 0.9, meaning 90% of asymptotic pPMSEr)
`sigma_k2`	Numeric. Predictor error variance for full model. If 'NULL' it is derived.
`sigma_p2`	Numeric. Predictor error variance for basic model. If 'NULL' it is derived.
`cov`	Optional covariance matrix. Must be `(k+1) x (k+1)` with the response 1st row and column.
`corr`	Optional correlation matrix. (Same layout as `cov`).
`SD`	Optional numeric vector of standard deviation for the predictors when a correlation matrix is supplied. Default 1

f2	Numeric. Cohen's f2 for effects of all predictors in full model.
f2_2	Numeric. Cohen's f2 for the effects of new predictors given the basic model.
R2_full	Numeric. Coefficient of determination for full model.
R2_basic	Numeric. Coefficient of determination for basic model.

Details

pmse_samplesize

- Sample Size Calculation for Prediction Models

pmse_samplesize The function calculates predictor error variance for the full model, with all predictors, and the reduced model, with the basic predictors using a provided covariance matrix or correlation matrix. It can also calculate predictor error variance through Cohen's F^2 and R^2 values. With the predictor error variance it determines a sample size from the efficient sample size at a target efficiency level and a sample size from a PMSE value of the full and reduced model. The final returned sample size is the largest out of the outputs.

Value

Numeric representing the required sample size.

References

Ma, Y. (2023). *Predictive Power and Efficient Sample Size in Linear Regression Models*. Master's Thesis, Worcester Polytechnic Institute.

Examples

```
## Example with a 5-predictor model (k = 5) and 2 basic predictors (p = 2)
pmse_samplesize(
  k = 5, p = 2,
  PMSE_val_k    = 1,
  PMSE_val_p    = 1,
  efficiency_level = 0.9,
  sigma_k2 = 0.50,
  sigma_p2 = 0.60
)
```

Index