

Package ‘ebrahim.gof’

September 1, 2025

Type Package

Title Ebrahim-Farrington Goodness-of-Fit Test for Logistic Regression

Version 1.0.0

Date 2025-08-20

Maintainer Ebrahim Khaled Ebrahim <ebrahimkhaled@alexu.edu.eg>

Description Implements the Ebrahim-Farrington goodness-of-fit test for logistic regression models, particularly effective for sparse data and binary outcomes. This test provides an improved alternative to the traditional Hosmer-Lemeshow test by using a modified Pearson chi-square statistic with data-dependent grouping. The test is based on Farrington (1996) theoretical framework but simplified for practical implementation with binary data. Includes functions for both the original Farrington test (for grouped data) and the new Ebrahim-Farrington test (for binary data with automatic grouping). For more details see Hosmer (1980) <doi:10.1080/03610928008827941> and Farrington (1996) <doi:10.1111/j.2517-6161.1996.tb02086.x>.

License GPL-3

URL <https://github.com/ebrahimkhaled/ebrahim.gof>

BugReports <https://github.com/ebrahimkhaled/ebrahim.gof/issues>

Depends R (>= 3.5.0)

Imports stats

Suggests testthat (>= 3.0.0), knitr, rmarkdown, ResourceSelection, ggplot2

Encoding UTF-8

RoxygenNote 7.3.2

VignetteBuilder knitr

NeedsCompilation no

Author Ebrahim Khaled Ebrahim [aut, cre]

Repository CRAN

Date/Publication 2025-09-01 17:10:22 UTC

Contents

ef.gof	2
Index	5

ef.gof	<i>Ebrahim-Farrington Goodness-of-Fit Test for Logistic Regression</i>
--------	--

Description

Performs the Ebrahim-Farrington goodness-of-fit test for logistic regression models. This test is particularly effective for binary data and sparse datasets, providing an improved alternative to the traditional Hosmer-Lemeshow test.

Usage

```
ef.gof(y, predicted_probs, model = NULL, m = NULL, G = 10)
```

Arguments

y	Numeric vector of binary responses (0/1) for binary data, or counts of successes for grouped data.
predicted_probs	Numeric vector of predicted probabilities from the logistic regression model. Must be same length as y.
model	Optional glm object. Required only for the original Farrington test with grouped data (when m is provided and G is NULL).
m	Optional numeric vector of trial counts for each observation (for grouped data). If NULL, data is assumed to be binary.
G	Optional integer specifying the number of groups for binary data grouping. Default is 10. If NULL, no grouping is performed and m must be provided.

Details

The Ebrahim-Farrington test is based on Farrington's (1996) theoretical framework but simplified for practical implementation with binary data. The test uses a modified Pearson chi-square statistic with data-dependent grouping, where observations are grouped by their predicted probabilities.

For binary data (when G is specified), the test automatically groups observations into G groups based on predicted probabilities and applies the simplified Ebrahim-Farrington statistic:

$$Z_{EF} = \frac{T_{EF} - (G - 2)}{\sqrt{2(G - 2)}}$$

where T_{EF} is the modified Pearson chi-square statistic, and G is the number of groups.

For grouped data (when m is provided), the test applies the original Farrington test with full variance calculations.

Value

A data frame with the following columns:

Test	Character string identifying the test performed
Test_Statistic	Numeric value of the standardized test statistic
p_value	Numeric p-value for the test

Note

- For binary data with automatic grouping (G specified): Use the Ebrahim-Farrington test which is computationally efficient and doesn't require the model specification.
- For grouped data (m provided): Use the original Farrington test which requires the fitted model object.
- The test statistic follows a standard normal distribution under the null hypothesis of adequate model fit.
- For binary data with m=1 for all observations and no grouping, the test is not applicable and will return a p-value of 1.

Author(s)

Ebrahim Khaled Ebrahim <ebrahimkhaled@alexu.edu.eg>

References

Farrington, C. P. (1996). On Assessing Goodness of Fit of Generalized Linear Models to Sparse Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(2), 349-360.
 Ebrahim, K. E. (2025). Goodness-of-Fits Tests and Calibration Machine Learning Algorithms for Logistic Regression Model with Sparse Data. *Master's Thesis*, Alexandria University.
 Hosmer, D. W., & Lemeshow, S. (1980). A goodness-of-fit test for the multiple logistic regression model. *Communications in Statistics - Theory and Methods*, 9(10), 1043–1069. <https://doi.org/10.1080/03610928008827941>

See Also

[hoslem.test](#) for the Hosmer-Lemeshow test

Examples

```
# Example 1: Binary data with automatic grouping (Ebrahim-Farrington test)
set.seed(123)
n <- 500
x <- rnorm(n)
linpred <- 0.5 + 1.2 * x
prob <- 1 / (1 + exp(-linpred))
y <- rbinom(n, 1, prob)

# Fit logistic regression
model <- glm(y ~ x, family = binomial())
predicted_probs <- fitted(model)
```

```
# Perform Ebrahim-Farrington test with 10 groups
result <- ef.gof(y, predicted_probs, G = 10)
print(result)

# Example 2: Compare with different number of groups
result_4 <- ef.gof(y, predicted_probs, G = 4)
result_20 <- ef.gof(y, predicted_probs, G = 20)

# Example 3: Grouped data (original Farrington test)
# Note: This requires actual grouped data with trials > 1
# Simulated grouped data
n_groups <- 50
m_trials <- sample(5:20, n_groups, replace = TRUE)
x_grouped <- rnorm(n_groups)
linpred_grouped <- -0.5 + 1.0 * x_grouped
prob_grouped <- 1 / (1 + exp(-linpred_grouped))
y_grouped <- rbinom(n_groups, m_trials, prob_grouped)

# Fit model for grouped data
data_grouped <- data.frame(successes = y_grouped, trials = m_trials, x = x_grouped)
model_grouped <- glm(cbind(successes, trials - successes) ~ x,
  data = data_grouped, family = binomial())
predicted_probs_grouped <- fitted(model_grouped)

# Original Farrington test
result_grouped <- ef.gof(y_grouped, predicted_probs_grouped,
  model = model_grouped, m = m_trials)
print(result_grouped)
```

Index

`ef.gof`, [2](#)

`hoslem.test`, [3](#)