Package 'glsm'

July 14, 2025

Type Package

Title Saturated Model Log-Likelihood for Multinomial Outcomes

Version 0.0.0.6

Date 2025-07-09

Author Jorge Villalba [aut, cre] (ORCID:
<https://orcid.org/0000-0002-2888-9660>),
Humberto Llinas [aut] (ORCID: <https://orcid.org/0000-0002-2976-5109>),
Jorge Borja [aut] (ORCID: <https://orcid.org/0009-0006-4824-5199>),
Jorge Tilano [aut] (ORCID: <https://orcid.org/0009-0005-5793-4183>)

Maintainer Jorge Villalba <jvillalba@utb.edu.co>

Description When the response variable Y takes one of R > 1 values, the function 'glsm()' computes the maximum likelihood estimates (MLEs) of the parameters under four models: null, complete, saturated, and logistic. It also calculates the log-likelihood values for each model. This method assumes independent, non-identically distributed variables. For grouped data with a multinomial outcome, where observations are divided into J populations, the function 'glsm()' provides estimation for any number K of explanatory variables.

Depends R (>= 3.5.0)

Imports stats, dplyr (>= 1.0.0), ggplot2 (>= 1.0.0), VGAM (>= 1.0.0), plyr

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.2

LazyData true

NeedsCompilation no

Repository CRAN

Date/Publication 2025-07-14 17:10:02 UTC

Contents

`confint.glsm` *Confidence Intervals for Coefficients in* `glsm` *Objects*

Description

Calculates confidence intervals for the coefficients in a fitted `glsm` model. Includes exponentiated intervals (Odds Ratios) for easier interpretation.

Usage

```
## S3 method for class 'glsm'
confint(object, parm, level = 0.95, ...)
```

Arguments

`object`	The type of prediction required. The default is on the scale of the linear predictors. The alternative `response` gives the predicted probabilities.
`parm`	calculate confidence intervals for the coefficients
`level`	It gives the desired confidence level for the confidence interval. For example, a default value is level = 0.95, which will generate a 95% confidence interval." The alternative `response` gives the predicted probabilities.
`...`	further arguments passed to or from other methods.

Details

Confint Method for 'glsm'

The saturated model is characterized by the assumptions 1 and 2 presented in section 2.3 by Llinas (2006, ISSN:2389-8976).

Value

An object of class `"confint.glsm"`, which is a list containing:

`object`	a `glsm` object
`parm`	calculate confidence intervals for the coefficients.
`level`	confidence levels

Author(s)

Humberto Llinás (Universidad del Norte, Barranquilla-Colombia; author), Jorge Villalba (Universidad Tecnológica de Bolívar, Cartagena-Colombia; author and creator), Jorge Borja (Universidad del Norte, Barranquilla-Colombia; author and creator), Jorge Tilano (Universidad del Norte, Barranquilla-Colombia; author)

References

Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). *Applied Logistic Regression* (3rd ed.). New York: Wiley. ISBN: 978-0-470-58247-3 Llinás, H. (2006). Precisiones en la teoría de los modelos logísticos. *Revista Colombiana de Estadística*, 29(2), 239–265. Llinás, H., & Carreño, C. (2012). The Multinomial Logistic Model for the Case in Which the Response Variable Can Assume One of Three Levels and Related Models. *Revista Colombiana de Estadística*, 35(1), 131–138. Orozco, E., Llinás, H., & Fonseca, J. (2020). Convergence theorems in multinomial saturated and logistic models. *Revista Colombiana de Estadística*, 43(2), 211–231. Llinás, H., Arteta, M., & Tilano, J. (2016). El modelo de regresión logística para el caso en que la variable de respuesta puede asumir uno de tres niveles: estimaciones, pruebas de hipótesis y selección de modelos. *Revista de Matemática: Teoría y Aplicaciones*, 23(1), 173–197.

Examples

```
# Load the glsm package and example dataset
library(glsm)
data("hsbdemo", package = "glsm")

# Fit a multinomial logistic regression model using glsm()
model <- glsm(prog ~ ses + gender, data = hsbdemo)

# Get confidence intervals for all model coefficients (default 95% level)
confint(model)

# Get confidence intervals for a specific coefficient
params <- names(model$coefficients)

results <- lapply(params, function(p) {
  cat("\nConfidence interval for:", p, "\n")
  print(confint(model, parm = p, level = 0.95))
})
```

`glsm` *Saturated Model Log-Likelihood for Multinomial Outcomes*

Description

When the response variable `Y` takes one of `R > 1` values, the function `"glsm()"` computes the maximum likelihood estimates (MLEs) of the parameters under four models: null, complete, saturated, and logistic. It also calculates the log-likelihood values for each model.

The method assumes independent, non-identically distributed variables. For grouped data with a multinomial outcome variable, where the observations are divided into `J` populations, the function `'"glsm()"` offers reliable estimation for any number `K` of explanatory variables.

Usage

```
glsm(formula, data, ref = NaN)
```

Arguments

formula	An object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted. See 'Details' for more information on model specification.
data	An optional data frame, list, or environment (or object coercible via as.data.frame) containing the variables in the model. If variables are not found in data, they are taken from environment(formula), typically the environment from which glsm() is called.
ref	Optional character string indicating the reference level of the response variable. If not specified, the first level is used by default.

Details

glsm.R

An expression of the form y ~ model is interpreted as a specification that the response variable y is modeled by a linear predictor, symbolically defined by model (the systematic component). The model consists of terms separated by + operators. Each term can include variable or factor names, and interactions between variables are denoted by :. Such a term represents the interaction of all included variables and factors. In this context, y is the outcome variable, which may be binary or polychotomous.

Value

An object of class "glsm", which is a list containing at least the following components:

coefficients	Vector of estimated coefficients, including intercepts and slopes.
coef	Alias for coefficients. Returns the same vector of estimated intercepts and slopes.
Std.Error	Vector of standard errors for the estimated coefficients (intercepts and slopes).
ExpB	Vector containing the exponentiated coefficients (i.e., exp(beta)) for interpretation as odds ratios.
Wald	Wald test statistic used to assess the significance of each coefficient (assumed to follow a chi-squared distribution).
DF	Degrees of freedom associated with the Wald test's chi-squared distribution.
P.value	P-values corresponding to the Wald test statistics.
Log_Lik_Complete	Log-likelihood value of the complete model.
Log_Lik_Null	Log-likelihood value of the null model.
Log_Lik_Logit	Log-likelihood value of the logistic model.
Log_Lik_Saturate	Log-likelihood value of the saturated model.
Populations	Number of populations considered in the saturated model.
Dev_Null_vs_Logit	Deviance statistic comparing the null and logistic models.

`Dev_Logit_vs_Complete`
: Deviance statistic comparing the logistic and complete models.

`Dev_Logit_vs_Saturate`
: Deviance statistic comparing the logistic and saturated models.

`Df_Null_vs_Logit`
: Degrees of freedom for the deviance test comparing the null and logistic models.

`Df_Logit_vs_Complete`
: Degrees of freedom for the deviance test comparing the logistic and complete models.

`Df_Logit_vs_Saturate`
: Degrees of freedom for the deviance test comparing the logistic and saturated models.

`P.v_Null_vs_Logit`
: P-value for the hypothesis test comparing the null and logistic models.

`P.v_Logit_vs_Complete`
: P-value for the hypothesis test comparing the logistic and complete models.

`P.v_Logit_vs_Saturate`
: P-value for the hypothesis test comparing the logistic and saturated models.

`Logit_r`
: Matrix of log-odds values, with respect to the reference category `r` of the outcome variable `Y`.

`p_hat_complete`
: Vector of probabilities that the outcome variable takes the value 1, given the `j`th population (estimated from the complete model, excluding the logistic model).

`p_hat_null`
: Vector of probabilities that the outcome variable takes the value 1, given the `j`th population (estimated from the null model, excluding the logistic model).

`p_rj`
: Matrix containing the estimated values of each `prj`, the probability that the outcome variable takes the value `r`, given the `j`th population (estimated using the logistic model).

`odd`
: Vector containing the odds for each `j`th population.

`OR`
: Vector containing the odds ratios for each variable's coefficient.

`z_rj`
: Vector containing the values of each `Zrj`, defined as the sum of observations in the `j`th population.

`nj`
: Vector containing the number of observations (`nj`) in each `j`th population.

`p_rj_tilde`
: Vector containing the estimated values of each `prj`, the probability that the outcome variable takes the value `r`, given the `j`th population (estimated under the saturated model, without estimating logistic parameters).

`v_rj`
: Vector of variances of the Bernoulli variables in the `j`th population and category `r`.

`m_rj`
: Vector of expected values of `Zj` in the `j`th population and category `r`.

`V_rj`
: Vector of variances of `Zj` in the `j`th population and category `r`.

`V`
: Variance–covariance matrix of `Z`, the vector containing all `Zj` values.

`S_p`
: Score vector computed under the saturated model.

`I_p`
: Fisher information matrix under the saturated model.

`Zast_j`
: Vector of standardized values for the variable `Zj`.

`mcov`	Variance–covariance matrix of the coefficient estimates.
`mcor`	Correlation matrix of the coefficient estimates.
`Esm`	Estimated Saturated Matrix. A data frame containing estimates from the saturated model. For each population `j`, it includes the values of the explanatory variables, `nj`, `Zrj`, `prj_tilde`, and the log-likelihood `Lp_tilde`.
`Elm`	Estimated Logit Matrix. A data frame containing estimates from the logistic model. For each population `j`, it includes the values of the explanatory variables, `nj`, `Zrj`, `prj`, the logit transformation `Logit_rj`, and the variance of the logit (`var_logit_rj`).
`call`	The original function call used to fit the glsm model.

Author(s)

Humberto Llinás (Universidad del Norte, Barranquilla-Colombia; author), Jorge Villalba (Universidad Tecnológica de Bolívar, Cartagena-Colombia; author and creator), Jorge Borja (Universidad del Norte, Barranquilla-Colombia; author and creator), Jorge Tilano (Universidad del Norte, Barranquilla-Colombia; author)

References

Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). *Applied Logistic Regression* (3rd ed.). New York: Wiley. ISBN: 978-0-470-58247-3 Llinás, H. (2006). Precisiones en la teoría de los modelos logísticos. *Revista Colombiana de Estadística*, 29(2), 239–265. Llinás, H., & Carreño, C. (2012). The Multinomial Logistic Model for the Case in Which the Response Variable Can Assume One of Three Levels and Related Models. *Revista Colombiana de Estadística*, 35(1), 131–138. Orozco, E., Llinás, H., & Fonseca, J. (2020). Convergence theorems in multinomial saturated and logistic models. *Revista Colombiana de Estadística*, 43(2), 211–231. Llinás, H., Arteta, M., & Tilano, J. (2016). El modelo de regresión logística para el caso en que la variable de respuesta puede asumir uno de tres niveles: estimaciones, pruebas de hipótesis y selección de modelos. *Revista de Matemática: Teoría y Aplicaciones*, 23(1), 173–197.

Examples

```
library(glsm)
data("hsbdemo", package = "glsm")
model <- glsm(prog ~ ses + gender, data = hsbdemo, ref = "academic")
model
```

`hsbdemo` *hsbdemo: School data for testing.*

Description

Entering high school students make program choices among general program, vocational program and academic program. Their choice might be modeled using their writing score and their social economic status.The data set contains variables on 200 students. The outcome variable is prog, program type. The predictor variables are social economic status, ses, a three-level categorical variable and writing score, write, a continuous variable.

Usage

```
hsbdemo
```

Format

A data frame with 200 rows and 17 columns:

Student Categorical. Student identification code.

id Categorical. Unique identifier for each student.

gender Categorical. Student gender: "female" or "male".

ses Categorical. Socioeconomic status: "low", "middle", "high".

schtyp Categorical. Type of school: "private" or "public".# corregido

prog Categorical. Program of study chosen: 0 = General, 1 = Vocational, 2 = Academic.

read Continuous. Reading test score.

write Continuous. Writing test score.

math Continuous. Math test score.

science Continuous. Science test score.

socst Continuous. Social studies test score.

honors Categorical. Honors enrollment status: "enrolled" or "not enrolled".

awards Integer. Number of awards received, ranging from 0 to 9.

cid Categorical. Unspecified score, ranging from 0 to 20.

prog0 Binary. 1 if prog = General, 0 otherwise.

prog1 Binary. 1 if prog = Vocational, 0 otherwise.

prog2 Binary. 1 if prog = Academic, 0 otherwise.

Source

Simulated dataset inspired by high school program choices.

`summary.glsm` *Summary Method for in* `glsm` *Objects*

Description

Summarizes a fitted `glsm` model, including coefficients, standard errors, odds ratios, Wald tests, and likelihood-ratio comparisons with nested models.

Usage

```
## S3 method for class 'glsm'
summary(object, ...)
```

Arguments

`object`	The `glsm` model to summarize. The details of the model specification are provided under **Details**.
`...`	Other arguments passed to or from other methods.

Details

Summary Method for 'glsm'

Value

"summary.glsm" returns an object of class `summary.glsm`, a list with components:

`Call`	The original call used to fit the model.
`coeff`	A matrix of coefficients with columns for the estimated coefficients (Coef(B)), standard errors (Std.Error), exponentiated coefficients (Exp(B)), Wald test statistics (Wald), degrees of freedom (DF), and the corresponding p-values (P.value).
`comparison test`	A matrix with comparison tests of the logistic model against the following models: Null, Complete, and Saturated. It includes the test statistic (Deviance), degrees of freedom (DF), and p-values (P.value).

#' @details The `glsm` function estimates a multinomial logistic regression model when the response variable takes more than two levels. The model compares the logistic specification against nested models (null, complete, and saturated), and provides maximum likelihood estimates, asymptotic inference for coefficients, and goodness-of-fit measures. This summary method presents the key components of the model in a structured format.

Author(s)

Humberto Llinás (Universidad del Norte, Barranquilla-Colombia; author), Jorge Villalba (Universidad Tecnológica de Bolívar, Cartagena-Colombia; author and creator), Jorge Borja (Universidad del Norte, Barranquilla-Colombia; author and creator), Jorge Tilano (Universidad del Norte, Barranquilla-Colombia; author)

References

Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). *Applied Logistic Regression* (3rd ed.). New York: Wiley. ISBN: 978-0-470-58247-3 Llinás, H. (2006). Precisiones en la teoría de los modelos logísticos. *Revista Colombiana de Estadística*, 29(2), 239–265. Llinás, H., & Carreño, C. (2012). The Multinomial Logistic Model for the Case in Which the Response Variable Can Assume One of Three Levels and Related Models. *Revista Colombiana de Estadística*, 35(1), 131–138. Orozco, E., Llinás, H., & Fonseca, J. (2020). Convergence theorems in multinomial saturated and logistic models. *Revista Colombiana de Estadística*, 43(2), 211–231. Llinás, H., Arteta, M., & Tilano, J. (2016). El modelo de regresión logística para el caso en que la variable de respuesta puede asumir uno de tres niveles: estimaciones, pruebas de hipótesis y selección de modelos. *Revista de Matemática: Teoría y Aplicaciones*, 23(1), 173–197.

Examples

```
data("hsbdemo", package = "glsm")
model <- glsm(prog ~ ses + gender, data = hsbdemo)
summary(model)
```

Index