Package ‘lsm’

March 7, 2020

Type Package

Title Estimation of the log Likelihood of the Saturated Model

Version 0.2.0

Date 2020-03-07

Author Humberto Llinas [aut],
Omar Fabregas [aut],
Jorge Villalba [aut, cre]

Maintainer Jorge Villalba <jlvia1191@gmail.com>

Description When the values of the outcome variable Y are either 0 or 1, the function lsm() calculates the estimation of the log likelihood in the saturated model. This model is characterized by Llinas (2006, ISSN:2389-8976) in section 2.3 through the assumptions 1 and 2. The function LogLik() works (almost perfectly) when the number of independent variables K is high, but for small K it calculates wrong values in some cases. For this reason, when Y is dichotomous and the data are grouped in J populations, it is recommended to use the function lsm() because it works very well for all K.

Depends R (>= 3.5.0)

Imports stats

Encoding UTF-8

License MIT + file LICENSE

LazyData TRUE

RoxygenNote 6.1.0

Collate 'lsm.R'

NeedsCompilation no

Repository CRAN

Date/Publication 2020-03-07 21:00:02 UTC

R topics documented:

`chdage` *Coronary Heart Disease Study*

Description

A dataset containing the age and other attributes of almost 100 subjects selected to participate in a study. The variables are as follows:

Usage

```
chdage
```

Format

A data frame with 100 observations on the following 3 variables.

`ID` identification code

`AGE` age in years

`CHD` presence (1) or absence (0) of evidence of significant coronary heart disease

References

[1] Hosmer, D. (2013). Wiley Series in Probability and Statistics Ser. : Applied Logistic Regression (3). New York: John Wiley & Sons, Incorporated.

Examples

```
data(chdege)
## maybe str(chdege) ; plot(chdege) ...
```

confint.lsm *Confidence Intervals for lsm Objects*

Description

Provides a confint method for `lsm` objects.

Usage

```
## S3 method for class \code{lsm}
## S3 method for class 'lsm'
confint(object, parm, level =0.95, ...)
```

Arguments

`object`	a lsm object
`parm`	parameter
`level`	confidence levels
`...`	additional parameters

Value

An object of class `lsm` analysis.

Author(s)

Jorge Villalba Acevedo

References

[1] Humberto Jesus Llinas. (2006). Accuracies in the theory of the logistic models. Revista Colombiana De Estadistica,29(2), 242-244.

[2] Hosmer, D. (2013). Wiley Series in Probability and Statistics Ser. : Applied Logistic Regression (3). New York: John Wiley & Sons, Incorporated.

[3] Chambers, J. M. and Hastie, T. J. (1992) Statistical Models in S. Wadsworth & Brooks/Cole.

Examples

```
# Hosmer, D. (2013) page 3: Age and coranary Heart Disease (CHD) Status of 20 subjects:

 AGE <- c(20, 23, 24, 25, 25, 26, 26, 28, 28, 29, 30, 30, 30, 30, 30, 30, 30, 32, 33, 33)
 CHD <- c(0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0)
 data <- data.frame (CHD, AGE)
 Ela <- lsm(CHD ~ AGE, family = binomial, data)
confint(Ela)
```

`icu` *The icu Study.*

Description

The `icu` study data set consists of a sample of 200 subjects who were part of a much larger study on survival of patients following admission to an adult intensive care unit (`icu`). The variables are as follows:

Usage

```
icu
```

Format

A data frame with 200 observations on the following 21 variables.

`ID` a numeric vector

`STA` a numeric vector

`AGE` a numeric vector

`GENDER` a numeric vector

`RACE` a numeric vector

`SER` a numeric vector

`CAN` a numeric vector

`CRN` a numeric vector

`INF` a numeric vector

`CPR` a numeric vector

`SYS` a numeric vector

`HRA` a numeric vector

`PRE` a numeric vector

`TYP` a numeric vector

`FRA` a numeric vector

`PO2` a numeric vector

`PH` a numeric vector

`PCO` a numeric vector

`BIC` a numeric vector

`CRE` a numeric vector

`LOC` a numeric vector

References

[1] Hosmer, D. (2013). Wiley Series in Probability and Statistics Ser. : Applied Logistic Regression (3). New York: John Wiley & Sons, Incorporated.

Examples

```
data(icu)
## maybe str(icu) ; plot(icu) ...
```

`lowbwt` *The Low Birth Weight Study.*

Description

This data set contains information on 189 birhs to women seen in the obstetrics clinic. The variables are as follows:

Usage

```
lowbwt
```

Format

A data frame with 189 observations on the following 11 variables.

`ID` a numeric vector

`SMOKE` a numeric vector

`RACE` a numeric vector

`AGE` a numeric vector

`LWT` a numeric vector

`BWT` a numeric vector

`LOW` a numeric vector

`PTL` a numeric vector

`HT` a numeric vector

`UI` a numeric vector

`FTV` a numeric vector

References

[1] Hosmer, D. (2013). Wiley Series in Probability and Statistics Ser. : Applied Logistic Regression (3). New York: John Wiley & Sons, Incorporated.

Examples

```
data(lowbwt)
## maybe str(lowbwt) ; plot(lowbwt) ...
```

`lsm` *Estimation of the log Likelihood of the Saturated Model*

Description

When the values of the outcome variable `Y` are either 0 or 1, the function `lsm()` calculates, among others, the values of the maximum likelihood estimates (ML-estimations) of the corresponding parameters in the null, complete, saturated and logistic models and also the estimations of the log likelihood in each of this models. The models null and complete are described by Llinas (2006, ISSN:2389-8976) in sections 2.1 and 2.2. The saturated model is characterized in section 2.3 of that paper through the assumptions 1 and 2. Finally, the logistic model and its assumptions are explained in section 2.4.

Additionally, based on asymptotic theory for these ML-estimations and the score vector, the function `lsm()` calculates the values of the approximations for different deviations -2 log L, where L is the likelihood function. Based on these approximations, the function obtains the values of statistics for several hypothesis tests (each with an asymptotic chi-squared distribution): Null vs Logit, Logit vs Complete and Logit vs Saturated.

With the function `lsm()`, it is possible calculate confidence intervals for the logistic parameters and for the corresponding odds ratio. The asymptotic theory was developed for the case of independent, non-identically distributed variables. If `Y` is dichotomous and the data are grouped in `J` populations, it is recommended to use the function `lsm()` because it works very well for all `K`, the number of explanatory variables.

Usage

```
lsm(formula, family=binomial, data, na.action )
```

Arguments

`formula`	An expression of the form y ~ model, where `y` is the outcome variable (binary or dichotomous: its values are 0 or 1).
`family`	an optional family for defaul binomial
`data`	an optional data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which `lsm()` is called.
`na.action`	a function which indicates what should happen when the data contain `NA`s.

Details

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

Value

`lsm` returns an object of class "`lsm`".

An object of class "`lsm`" is a list containing at least the following components:

- `coefficients` Vector of coefficients estimations.
- `Std.Error` Vector of the coefficients's standard error.
- `Exp(B)` Vector with the exponential of the coefficients.
- `Wald` Value of the Wald statistic.
- `D.f` Degree of freedom for the Chi-squared distribution.
- `P.value` P-value with the Chi-squared distribution.
- `Log_Lik_Complete` Estimation of the log likelihood in the complete model.
- `Log_Lik_Null` Estimation of the log likelihood in the null model.
- `Log_Lik_Logit` Estimation of the log likelihood in the logistic model.
- `Log_Lik_Saturate` Estimation of the log likelihood in the saturate model.
- `Populations` Number of populations in the saturated model.
- `Dev_Null_vs_Logit` Value of the test statistic (Hypothesis: null vs logistic models).
- `Dev_Logit_vs_Complete` Value of the test statistic (Hypothesis: logistic vs complete models).
- `Dev_Logit_vs_Saturate` Value of the test statistic (Hypothesis: logistic vs saturated models).
- `Df_Null_vs_Logit` Degree of freedom for the test statistic's distribution (Hypothesis: null vs logistic models).
- `Df_Logit_vs_Complete` Degree of freedom for the test statistic's distribution (Hypothesis: logistic vs saturated models).
- `Df_Logit_vs_Saturate` Degree of freedom for the test statistic's distribution (Hypothesis: Logistic vs saturated models).
- `P.v_Null_vs_Logit` `p-values` for the hypothesis test: null vs logistic models.
- `P.v_Logit_vs_Complete` `p-values` for the hypothesis test: logistic vs complete models.
- `P.v_Logit_vs_Saturate` `p-values` for the hypothesis test: logistic vs saturated models.
- `Logit` Estimation of the logit function (the log-odds).
- `p_hat` Estimation of the probability that the outcome variable takes the value 1, given one population.
- `fitted.values` Vector with the values of the log_Likelihood in each `jth` population.

z_j	Vector with the values of each Zj (the sum of the observations in the jth population).
n_j	Vector with the nj (the number of the observations in each jth population).
p_j	Vector with the estimation of each pj (the probability of success in the jth population).
v_j	Vector with the variance of the Bernoulli variables in the jth population.
m_j	Vector with the expected values of Zj in the jth population.
V_j	Vector with the variances of Zj in the jth population.
V	Variance and covariance matrix of Z, the vector that contains all the Zj.
S_p	Score vector in the saturated model.
I_p	Information matrix in the saturated model.
Zast_j	Vector with the values of the standardized variable of Zj.

Author(s)

Humberto Llinas Solano [aut], Universidad del Norte, Barranquilla-Colombia \ Omar Fabregas Cera [aut], Universidad del Norte, Barranquilla-Colombia \ Jorge Villalba Acevedo [cre, aut], Universidad Tecnológica de Bolívar, Cartagena-Colombia.

References

[1] Humberto Jesus Llinas. (2006). Accuracies in the theory of the logistic models. Revista Colombiana De Estadistica,29(2), 242-244.

[2] Hosmer, D. (2013). Wiley Series in Probability and Statistics Ser. : Applied Logistic Regression (3). New York: John Wiley & Sons, Incorporated.

[3] Chambers, J. M. and Hastie, T. J. (1992) Statistical Models in S. Wadsworth & Brooks/Cole.

Examples

```
# Hosmer, D. (2013) page 3: Age and coranary Heart Disease (CHD) Status of 20 subjects:

 AGE <- c(20, 23, 24, 25, 25, 26, 26, 28, 28, 29, 30, 30, 30, 30, 30, 30, 30, 32, 33, 33)
 CHD <- c(0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0)

 data <- data.frame (CHD, AGE)
 lsm(CHD ~ AGE , family = binomial,  data)

 # Other case.

    y <- c(1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1)
   x1 <- c(2, 2, 2, 5, 5, 5, 5, 8, 8, 11, 11, 11)

  data <- data.frame (y, x1)
  ELAINYS <-lsm(y ~ x1, family=binomial, data)
  summary(ELAINYS)

 ## For more ease, use the following notation.
```

```
 lsm(y~., family = binomial, data)

## Other case.

y <- as.factor(c(1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1))
 x1 <- as.factor(c(2, 2, 2, 5, 5, 5, 5, 8, 8, 11, 11, 11))

 data <- data.frame (y, x1)
 ELAINYS1 <-lsm(y ~ x1, family=binomial, data)
 confint(ELAINYS1)

## For more ease, use the following notation.
 lsm(y~. , family = binomial, data)
```

---

pros *The Prostate Cancer Study*

Description

A third data set involves a study of patiens with cancer of the prostate.

Usage

```
pros
```

Format

A data frame with 380 observations on the following 9 variables.

`ID` a numeric vector

`CAPSULE` a numeric vector

`AGE` a numeric vector

`RACE` a numeric vector

`DPROS` a numeric vector

`DCAPS` a numeric vector

`PSA` a numeric vector

`VOL` a numeric vector

`GLEASON` a numeric vector

References

[1] Hosmer, D. (2013). Wiley Series in Probability and Statistics Ser. : Applied Logistic Regression (3). New York: John Wiley & Sons, Incorporated.

Examples

```
data(pros)
## maybe str(pros) ; plot(pros) ...
```

summary.lsm *Summarizing Method for lsm Objects*

Description

Provides a summary method for lsm objects.

Usage

```
## S3 method for class \code{lsm}
## S3 method for class 'lsm'
summary(object, ...)
```

Arguments

object	a lsm object
...	additional parameters

Value

Side effect: a summary table with results from lsm analysis.

References

[1] Humberto Jesus Llinas. (2006). Accuracies in the theory of the logistic models. Revista Colombiana De Estadistica,29(2), 242-244.

[2] Hosmer, D. (2013). Wiley Series in Probability and Statistics Ser. : Applied Logistic Regression (3). New York: John Wiley & Sons, Incorporated.

[3] Chambers, J. M. and Hastie, T. J. (1992) Statistical Models in S. Wadsworth & Brooks/Cole.

Examples

```
# Hosmer, D. (2013) page 3: Age and coranary Heart Disease (CHD) Status of 20 subjects:

 AGE <- c(20, 23, 24, 25, 25, 26, 26, 28, 28, 29, 30, 30, 30, 30, 30, 30, 30, 32, 33, 33)
 CHD <- c(0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0)
 data <- data.frame (CHD, AGE)
 Ela <- lsm(CHD ~ AGE, family = binomial, data)
 summary(Ela)
```

uis *uis*

Description

uis

Usage

```
uis
```

Format

A data frame with 575 observations on the following 9 variables.

`ID` a numeric vector

`AGE` a numeric vector

`BECK` a numeric vector

`IVHX` a numeric vector

`NDRUGTX` a numeric vector

`RACE` a numeric vector

`TREAT` a numeric vector

`SITE` a numeric vector

`DFREE` a numeric vector

References

[1] Hosmer, D. (2013). Wiley Series in Probability and Statistics Ser. : Applied Logistic Regression (3). New York: John Wiley & Sons, Incorporated.

Examples

```
data(uis)
## maybe str(uis) ; plot(uis) ...
```

Index