

Package ‘lsm’

October 13, 2022

Type Package

Title Estimation of the log Likelihood of the Saturated Model

Version 0.2.1.2

Date 2022-02-03

Author Humberto Llinas [aut] (<<https://orcid.org/0000-0002-2976-5109>>),
Omar Fabregas [aut] (<<https://orcid.org/0000-0001-6853-6280>>),
Jorge Villalba [aut, cre] (<<https://orcid.org/0000-0002-2888-9660>>)

Maintainer Jorge Villalba <jlvialba1191@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, dplyr (>= 1.0.0), ggplot2 (>= 1.0.0)

Encoding UTF-8

License MIT + file LICENSE

LazyData TRUE

RoxygenNote 7.1.2

NeedsCompilation yes

Repository CRAN

Date/Publication 2022-02-04 03:10:02 UTC

R topics documented:

chdage	2
confint.lsm	3

icu	4
lowbwt	5
lsm	6
predict.lsm	9
pros	10
summary.lsm	11
survey	12
uis	15

Index	16
--------------	-----------

chdage	<i>Coronary Heart Disease Study</i>
--------	-------------------------------------

Description

Coronary Heart Disease Study

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

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

Examples

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

 confint.lsm

Confidence Intervals for lsm Objects

Description

Provides a confint method for lsm objects.

Usage

```
## S3 method for class 'lsm'
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	further arguments passed to or from other methods.
level	The type of prediction required. The default is on the scale of the linear predictors. The alternative response gives the predicted probabilities.
...	further arguments passed to or from other methods.

Details

confint Method for lsm

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:

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

Author(s)

Jorge Villalba Acevedo [cre, aut], Cartagena-Colombia.

References

- [1] Humberto Jesus Llinas. (2006). Accuracies in the theory of the logistic models. *Revista Colombiana De Estadística*,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 coronary 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, 0, 1, 0, 0, 0, 0)
# data <- data.frame (CHD, AGE)
# Ela <- lsm(CHD ~ AGE, family = binomial, data)
# summary(Ela)
```

icu

icu

Description

icu

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

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

lowbwt

Description

lowbwt

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

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 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. 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 .

Usage

```
lsm(formula, family = binomial, data = environment(formula))
```

Arguments

<code>formula</code>	An expression of the form $y \sim \text{model}$, where y is the outcome variable (binary or dichotomous: its values are 0 or 1).
<code>family</code>	an optional function for example binomial.
<code>data</code>	an optional data frame, list or environment (or object coercible by <code>as.data.frame</code> to a data frame) containing the variables in the model. If not found in data, the variables are taken from <code>environment(formula)</code> , typically the environment from which <code>lsm()</code> is called.

Details

Estimation of the log Likelihood of the Saturated Model

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.
ExpB	Vector with the exponential of the coefficients.
Wald	Value of the Wald statistic.
DF	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	Vector with the log-odds.
p_hat	Vector with the probabilities that the outcome variable takes the value 1, given the jth population.
odd	Vector with the values of the odd in each jth population.

OR	Vector with the values of the odd ratio for each coefficient of the variables.
z_j	Vector with the values of each Z_j (the sum of the observations in the j th population).
n_j	Vector with the n_j (the number of the observations in each j th population).
p_j	Vector with the estimation of each p_j (the probability of success in the j th population).
v_j	Vector with the variance of the Bernoulli variables in the j th population.
m_j	Vector with the expected values of Z_j in the j th population.
V_j	Vector with the variances of Z_j in the j th population.
V	Variance and covariance matrix of Z , the vector that contains all the Z_j .
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 Z_j .
mcov	Variance and covariance matrix for coefficient estimates.
mcor	Correlation matrix for coefficient estimates.
Esm	Estimates in the saturated model.
Elm	Estimates in the logistic model.

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 Estadística*, 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 coronary Heart Disease (CHD) Status of 20 subjects:

#library(lsm)

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

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

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



```

#lsm(y~., 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)

# 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)

```

predict.lsm

Predict Method for lsm Objects

Description

Obtains predictions from a fitted lsm object.

Usage

```

## S3 method for class 'lsm'
predict(
  object,
  newdata,
  type = c("link", "response", "odd"),
  interval = c("none", "confidence", "prediction", "odd"),
  level = 0.95,
  ...
)

```

Arguments

object	A fitted object of class lsm.
newdata	Optionally, a data frame in which to look for variables with which to predict. If omitted, the fitted linear predictors are used.
type	The type of prediction required. The default is on the scale of the linear predictors. The alternative response gives the predicted probabilities.

interval gives the
 level gives the
 ... further arguments passed to or from other methods.

Details

Predict Method for lsm Fits

Value

A vector or matrix of predictions. following components:

pros	<i>pros</i>
------	-------------

Description

pros

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

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 'lsm'
summary(object, ...)
```

Arguments

object	An expression of the form $y \sim \text{model}$, where y is the outcome variable (binary or dichotomous: its values are 0 or 1).
...	further arguments passed to or from other methods.

Details

summary Method for lsm

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 "lsm" is a list containing at least the following components:

object	a lsm object
...	additional parameters

Author(s)

Jorge Villalba Acevedo [cre, aut], Cartagena-Colombia.

References

- [1] Humberto Jesus Llinas. (2006). Accuracies in the theory of the logistic models. *Revista Colombiana De Estadística*,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 coronary 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, 1, 0, 0, 0, 0)
# data <- data.frame (CHD, AGE)
# Ela <- lsm(CHD ~ AGE, family = binomial, data)
# summary(Ela)
```

survey

survey

Description

The data was collected by applying a survey to a sample of university students.

Usage

survey

Format

A data frame (tibble) with 800 observations and 66 variables, which are described below:

Observation Student.

ID Identification code.

Gender Gender of the student, 1 = Female; 2 = Male.

Like What do you do most often in your free time? 1 = Network (Check social networks); 2 = TV (Watch TV).

Age Age of the student (in years), Numeric vector from 12.0 to 30.0

Smoke Do you smoke? 0 = No; 1 = Yes.

Height Height of the student (in meters), Numeric vector from 1.50 to 1.90.

Weight Weight of the student (in kilograms), numeric vector from 49 to 120.

BMI Body mass index of the student (kg/m²), numeric vector from 14 to 54.

School Type of school students come from, 1 = Private; 2 = Public.

SES Socio-economic stratus of the student, 1 = Low; 2 = Medium; 3 = High.

Enrollment What was your type funding to study at the university? 1 = Credit; 2 = Scholarship; 3 = Savings.

Score Percentage of success in a certain test, numeric vector from 0 to 100%

MotherHeight Height of the mother of the student (in meters), numeric vector 1 = Short; 2 = Normal; 3 = Tall.

MotherAge Age of the mother of the student (in years), numeric vector from 39 to 89.

MotherCHD Has your mother had coronary heart disease? 0 = No; 1 = Yes.

- FatherHeight Height of the father of the student (in meters), numeric vector 1 = Short; 2 = Normal; 3 = Tall.
- FatherAge Age of the father of the student (in years), numeric vector from 39 to 89
- FatherCHD Has your father had coronary heart disease, 1 = No; 2 = Yes.
- Status Student's academic status at the end of the previous semester, 1 = Distinguished; 2 = Normal; 3 = Regular.
- SemAcum Average of all final grades in the previous semester, numeric vector from 0.0 to 5.0
- Exam1 First exam taken last semester, numeric vector from 0.0 to 5.0
- Exam2 Second exam taken last semester, numeric vector from 0.0 to 5.0
- Exam3 Third exam taken last semester, numeric vector from 0.0 to 5.0
- Exam4 Last exam taken last semester, numeric vector from 0.0 to 5.0
- ExamAcum Sum of the four exams mentioned above, numeric vector from 0.0 to 5.0
- Definitive Average of the four exams mentioned above, numeric vector from 0.0 to 5.0
- Expense Average of your monthly expenses (in 10 thousand Colombian pesos), numeric vector from 23.0 to 90.0
- Income Father's monthly income (in millions of Colombian pesos), numeric vector from 1.0 to 3.0
- Gas Value paid for gas service in the last month (in thousands of Colombian pesos), numeric vector from 15.0 to 28.0
- Course What type of virtual classes do you prefer? 1 = Virtual; 2 = Face-to-face.
- Law Opinion on a law, 1 = In disagreement; 2=Agree
- Economic How was your family's economy during the pandemic? 1 = Bad; 2 = Regular; 3 = Good.
- Race Does the student belong to an ethnic group? 1=None; 2= Ethnic
- Region Region of the country where the student comes from, 1 = North; 2 = Center; 3 = South.
- EM01 During this period of preventative isolation, you frequently become nervous or restless for no reason, 1 = Never, 2 = Rarely; 3 = Almost always; 4 = Always.
- EM02 During this period of preventative isolation, you are often irritable, 1 = Never, 2 = Rarely; 3 = Almost always; 4 = Always.
- EM03 During this period of preventive isolation, you are often sad or despondent, 1 = Never, 2 = Rarely; 3 = Almost always; 4 = Always
- EM04 During this period of preventive isolation, you are often easily frightened, 1 = Never, 2 = Rarely; 3 = Almost always; 4 = Always
- EM05 During this period of preventative isolation, you often have trouble thinking clearly, 1 = Never, 2 = Rarely; 3 = Almost always; 4 = Always
- GOAL1 I am concerned that I may not be able to understand the contents of my subjects this semester as thoroughly as I would like, 1 = Strongly agree; 2 = Disagree; 3 = Undecided; 4 = Agree; 5 = Strongly agree.
- GOAL2 It is important for me to do better than other students in my subjects this semester, 1 = Strongly agree; 2 = Disagree; 3 = Undecided; 4 = Agree; 5 = Strongly agree.
- GOAL3 I am concerned that I may not learn all that I can learn in my subjects this semester, 1 = Strongly agree; 2 = Disagree; 3 = Undecided; 4 = Agree; 5 = Strongly agree.

Pre_STAT1 I like statistics, 1 = Strongly agree; 2 = Disagree; 3 = Undecided; 4 = Agree; 5 = Strongly agree.

Pre_STAT2 I don't focus when I make problems statistics, 1 = Strongly agree; 2 = Disagree; 3 = Undecided; 4 = Agree; 5 = Strongly agree.

Pre_STAT3 I don't understand statistics much because of my way of thinking, 1 = Strongly agree; 2 = Disagree; 3 = Undecided; 4 = Agree; 5 = Strongly agree.

Pre_STAT4 I use statistics in everyday life, 1 = Strongly agree; 2 = Disagree; 3 = Undecided; 4 = Agree; 5 = Strongly agree.

Post_STAT1 I like statistics, 1 = Strongly agree; 2 = Disagree; 3 = Undecided; 4 = Agree; 5 = Strongly agree.

Post_STAT2 I don't focus when I make problems statistics, 1 = Strongly agree; 2 = Disagree; 3 = Undecided; 4 = Agree; 5 = Strongly agree.

Post_STAT3 I don't understand statistics much because of my way of thinking, 1 = Strongly agree; 2 = Disagree; 3 = Undecided; 4 = Agree; 5 = Strongly agree.

Post_STAT4 I use statistics in everyday life, 1 = Strongly agree; 2 = Disagree; 3 = Undecided; 4 = Agree; 5 = Strongly agree.

Pre_IDARE1 I feel calm, 1=Nothing; 2= Little; 3= Quite a bit; 4= A lot.

Pre_IDARE2 I feel safe, 1=Nothing; 2= Little; 3= Quite a bit; 4= A lot.

Pre_IDARE3 I feel nervous, 1=Nothing; 2= Little; 3= Quite a bit; 4= A lot.

Pre_IDARE4 I'm stressed, 1=Nothing; 2= Little; 3= Quite a bit; 4= A lot.

Pre_IDARE5 I am comfortable, 1=Nothing; 2= Little; 3= Quite a bit; 4= A lot.

Post_IDARE1 I feel calm, 1=Nothing; 2= Little; 3= Quite a bit; 4= A lot.

Post_IDARE2 I feel safe, 1=Nothing; 2= Little; 3= Quite a bit; 4= A lot.

Post_IDARE3 I feel nervous, 1=Nothing; 2= Little; 3= Quite a bit; 4= A lot.

Post_IDARE4 I'm stressed, 1=Nothing; 2= Little; 3= Quite a bit; 4= A lot.

Post_IDARE5 I am comfortable, 1=Nothing; 2= Little; 3= Quite a bit; 4= A lot.

PSIC01 I feel good, 1=Almost never; 2= Sometimes; 3= Frequently; 4= Almost always.

PSIC02 I get tired quickly, 1=Almost never; 2= Sometimes; 3= Frequently; 4= Almost always.

PSIC03 I feel like crying, 1=Almost never; 2= Sometimes; 3= Frequently; 4= Almost always.

PSIC04 I would like to be as happy as others seem to be, 1=Almost never; 2= Sometimes; 3= Frequently; 4= Almost always.

PSIC05 I lose opportunities for not being able to decide quickly, 1=Almost never; 2= Sometimes; 3= Frequently; 4= Almost always.

Details

survey

Examples

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

*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

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

* **cedage**

chdage, 2

* **data**

icu, 4

pros, 10

survey, 12

uis, 15

* **lowbwt**

lowbwt, 5

chdage, 2

confint.lsm, 3

icu, 4

lowbwt, 5

lsm, 6

predict.lsm, 9

pros, 10

summary.lsm, 11

survey, 12

uis, 15