

# Package ‘prLogistic’

February 20, 2015

**Type** Package

**Version** 1.2

**Date** 2013-09-19

**Title** Estimation of Prevalence Ratios using Logistic Models

**Author** Raydonal Ospina <raydonal@de.ufpe.br> and Leila D. Amorim <leiladen@ufba.br>.

**Maintainer** Raydonal Ospina <raydonal@de.ufpe.br>

**Depends** R (>= 2.10.0), boot, methods, stats, stats4, lme4, Hmisc

**Description** Estimation of prevalence ratios using logistic models and confidence intervals with delta and bootstrap methods.

**License** GPL (>= 2)

**LazyLoad** yes

**LazyData** yes

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2013-09-19 18:47:33

## R topics documented:

downer	2
LBW	3
prLogisticBootCond	4
prLogisticBootMarg	6
prLogisticDelta	8
Thailand	10
titanic	11
Toenail	12
UIS	13
<b>Index</b>	<b>15</b>

---

downer

*Downer Data*

---

### Description

Data are related to dairy cows with a serious illness that may lead them to death. The dairy cows are also called downers because they sometimes lay down. A study was conducted with over 400 downer cows at the Ruakura New Zealand Animal Health Laboratory during 1983-1984. In this study a variety of blood tests were performed. The main goal of the study was to evaluate if the survival of the cows can be predicted from the blood measurements. For the illustration considered here, we use a subset of the original dataset. The variables on this dataset are serum aspartate amino transferase, calving, serum creatine phosphokinase, Myopathy and Survival. These data are originally from Clark, Henderson, Hoggard, Ellison, and Young (1987). There were missing for some variables. We consider here only the complete cases.

### Usage

```
data(downer)
```

### Format

A data frame with 216 observations on the following 5 variables:

**AST** Serum aspartate amino transferase, in U/l at 30C, recoded to 1 if equal or less than 492, 0= greater than 492.

**CK** Serum creatine phosphokinase, in U/l at 30C, recoded to 1 if equal or less than 5467, 0= greater than 5467.

**Calving** Calving (0 if measured before calving, 1 if after).

**Myopathy** Muscle disorder indicator (1=yes, 0=no).

**Survival** Survival indicator (1=survived, 0=died or killed).

### Source

The data come from  
Weisberg, S. (2005) Applied Linear Regression. Wiley-Interscience.

These data are copyrighted by John Wiley & Sons Inc. We acknowledged it and use the data to illustrate the methodology included in this library.

### References

Clark R, Henderson HV, Hoggard GK, Ellison R and Young B (1987) The ability of biochemical and haematological tests to predict recovery in periparturient recumbent cows. New Zealand Vet. Journal, 32, 126-133

## Examples

```
data("downer", package = "prLogistic")
attach(downer)
# prop.table(table(Survival))
fit.logist1=glm(Survival ~ Calving + CK + AST + Myopathy,
family=binomial, data=downer)
# summary(fit.logist1)
exp(fit.logist1$coef)
```

---

LBW

*Low Birth Weight Data*

---

## Description

Data are from a hypothetical study to evaluate factors associated to low birth weight. This data set contains information on 488 births to 188 women seen in an obstetrics clinic, comprising 1 to 3 births for each woman and yielding an average of 2.6 births per women. The outcome binary variable "low" indicates low birth weight defined as birth weight less than 2500 grams. These data are copyrighted by John Wiley & Sons Inc. and are used here to illustrate the calculation of prevalence ratios and corresponding confidence intervals.

## Usage

```
data(LBW)
```

## Format

A data frame with 488 observations on the following 6 variables:

**ID** Subject's identification.

**birth** Birth order.

**smoke** Smoking Status During Pregnancy (1=Yes, 0= No).

**race** Race (1=White, 0= Others).

**age** Age of Mother (1= age greater than 35 years, 0= age less or equal to 35 years).

**low** Low Birth Weight (1 if birth weight less than 2500 grams, 0 if birth weight greater or equal to 2500 grams).

## Source

The data come from

Data courtesy of University of Massachusetts Amherst (<http://www.umass.edu/statdata/statdata/stat-logistic.html>)

These data are copyrighted by John Wiley & Sons Inc. We acknowledged it and use the data to illustrate the methodology included in this library.

## References

Hosmer and Lemeshow (2000) Applied Logistic Regression. New York: Wiley Inter-Science. Second Edition.

## Examples

```
data("LBW", package = "prLogistic")
attach(LBW)
summary(LBW)
```

---

prLogisticBootCond	<i>Estimation of Prevalence Ratios using Logistic Models and Bootstrap Confidence Intervals for Conditional Standardization</i>
--------------------	---

---

## Description

This function estimates prevalence ratios (PRs) and bootstrap confidence intervals using logistic models for conditional standardization. The estimation of standard errors for PRs is obtained through use of bootstrapping. Confidence intervals of (1-alpha)% for PRs are available for standard logistic regression and for random-effects logistic models (Santos et al, 2008). The function prLogisticBootCond allows estimation of PRs using conditional standardization procedure (Wilcosky and Chambless, 1985).

glm, glmer, prLogisticDelta, prLogisticBootMarg

## Usage

```
prLogisticBootCond(object, data, conf = 0.95, R = 99, ...)
```

## Arguments

object	any fitted model object from which fixed effects estimates can be extracted. The details of model specification are given below.
data	a required data frame containing the variables named in object.
conf	scalar or vector specifying confidence level(s) for estimation. The default is conf= 0.95.
R	the number of bootstrap replicates. The default is R=99.
...	optional additional arguments. Currently none are used in any methods.

## Details

The fitted model object can be obtained using glm() function for binary responses when unit samples are independent. The glmer() function should be used for correlated binary responses. Only binary predictors are allowed. If categorization for predictors is other than (0,1), factor() should be considered.

**Value**

Returns prevalence ratio and its 95% bootstrap confidence intervals for conditional standardization. Both normal and percentile bootstrap confidence intervals are presented.

**Author(s)**

Raydonal Ospina, Department of Statistics, Federal University of Pernambuco, Brazil  
(raydonal@de.ufpe.br)

Leila D. Amorim, Department of Statistics, Federal University of Bahia, Brazil  
(leiladen@ufba.br).

**References**

Localio AR, Margolis DJ, Berlin JA (2007). Relative risks and confidence intervals were easily computed indirectly from multivariate logistic regression. *Journal of Clinical Epidemiology*, **60**, 874-882.

Oliveira NF, Santana VS, Lopes AA (1997). Ratio of proportions and the use of the delta method for confidence interval estimation in logistic regression. *Journal of Public Health*, **31**(1), 90-99.

Santos CAST et al (2008). Estimating adjusted prevalence ratio in clustered cross-sectional epidemiological data. *BMC Medical Research Methodology*, **8** (80). Available from <http://www.biomedcentral.com/1471-2280/8/80>.

Wilcosky TC, Chambless LE (1985). A comparison of direct adjustment and regression adjustment of epidemiologic measures. *Journal of Chronic Diseases*, **34**, 849-856.

**See Also**

[glm](#), [glmer](#), [prLogisticDelta](#), [prLogisticBootMarg](#)

**Examples**

```
# Estimates from logistic regression with bootstrap confidence intervals -
## conditional standardization
# Not run:
# data("titanic", package = "prLogistic")
# attach(titanic)
# fit.logistic=glm(survived~ sex + pclass + embarked, family=binomial,
# data = titanic)
# prLogisticBootCond(fit.logistic, data = titanic)
# End (Not run:)

# Another way for fitting the same model:
# Not run:
# prLogisticBootCond(glm(survived~ sex + pclass + embarked, family=binomial,
# data = titanic), data=titanic)
# End (Not run:)

### For clustered data
# Estimates from random-effects logistic regression
## with bootstrap confidence intervals - conditional standardization
```

```
# Not run:
# library(lme4)
# data("Thailand", package = "prLogistic")
# attach(Thailand)
# ML = glmer(rgi ~ sex + pped + (1|schoolid),
# family = binomial, data = Thailand)

# prLogisticBootCond(ML, data = Thailand)
# End (Not run:)
```

---

prLogisticBootMarg      *Estimation of Prevalence Ratios using Logistic Models and Bootstrap Confidence Intervals for Marginal Standardization*

---

### Description

This function estimates prevalence ratios (PRs) and bootstrap confidence intervals using logistic models for marginal standardization. The estimation of standard errors for PRs is obtained through use of bootstrapping. Confidence intervals of (1-alpha)% for PRs are available for standard logistic regression and for random-effects logistic models (Santos et al, 2008). The function prLogisticBootMarg allows estimation of PRs using marginal standardization procedure (Wilcosky and Chambless, 1985).

### Usage

```
prLogisticBootMarg(object, data, conf = 0.95, R = 99, ...)
```

### Arguments

object	any fitted model object from which fixed effects estimates can be extracted. The details of model specification are given below.
data	a required data frame containing the variables named in object.
conf	scalar or vector specifying confidence level(s) for estimation. The default is conf= 0.95.
R	the number of bootstrap replicates. The default is R=99.
...	optional additional arguments. Currently none are used in any methods.

### Details

The fitted model object can be obtained using glm() function for binary responses when unit samples are independent. The glmer() function should be used for correlated binary responses. Only binary predictors are allowed. If categorization for predictors is other than (0,1), factor() should be considered.

### Value

Returns prevalence ratio and its 95% bootstrap confidence intervals for marginal standardization. Both normal and percentile bootstrap confidence intervals are presented.

**Author(s)**

Raydonal Ospina, Department of Statistics, Federal University of Pernambuco, Brazil  
(raydonal@de.ufpe.br)

Leila D. Amorim, Department of Statistics, Federal University of Bahia, Brazil  
(leiladen@ufba.br).

**References**

Localio AR, Margolis DJ, Berlin JA (2007). Relative risks and confidence intervals were easily computed indirectly from multivariate logistic regression. *Journal of Clinical Epidemiology*, **60**, 874-882.

Oliveira NF, Santana VS, Lopes AA (1997). Ratio of proportions and the use of the delta method for confidence interval estimation in logistic regression. *Journal of Public Health*, **31**(1), 90-99.

Santos CAST et al (2008). Estimating adjusted prevalence ratio in clustered cross-sectional epidemiological data. *BMC Medical Research Methodology*, **8** (80). Available from <http://www.biomedcentral.com/1471-2280/8/80>.

Wilcosky TC, Chambless LE (1985). A comparison of direct adjustment and regression adjustment of epidemiologic measures. *Journal of Chronic Diseases*, **34**, 849-856.

**See Also**

[glm](#), [glmer](#), [prLogisticDelta](#), [prLogisticBootCond](#)

**Examples**

```
### For independent observations:
## Estimates from logistic regression with bootstrap confidence intervals -
## marginal standardization
# Not run:
# data("titanic", package = "prLogistic")
# attach(titanic)
# fit.logistic=glm(survived~ sex + pclass + embarked, family=binomial,
# data = titanic)
# prLogisticBootMarg(fit.logistic, data = titanic)
# End (Not run:)

# Another way for fitting the same model:
# Not run:
# prLogisticBootMarg(glm(survived~ sex + pclass + embarked,
# family=binomial, data = titanic), data=titanic)
# End (Not run:)

### For clustered data
## Estimates from random-effects logistic regression
## with bootstrap confidence intervals - marginal standardization
# Not run:
# library(lme4)
# data("Thailand", package = "prLogistic")
# attach(Thailand)
# ML = glmer(rgi ~ sex + pped + (1|schoolid),
```

```
# family = binomial, data = Thailand)

# prLogisticBootMarg(ML, data = Thailand)
# End (Not run:)
```

---

prLogisticDelta	<i>Estimation of Prevalence Ratios using Logistic Models and Confidence Intervals with Delta Method</i>
-----------------	---

---

## Description

This function estimates prevalence ratios (PRs) and their confidence intervals using logistic models. The estimation of standard errors for PRs is obtained through use of delta method. Confidence intervals of (1-alpha)% for PRs are available for standard logistic regression and for random-effects logistic models (Santos et al, 2008). The function prLogisticDelta allows estimation of PRs using two standardization procedures: conditional or marginal (Wilcosky and Chambless, 1985).

glm, glmer, prLogisticBootCond, prLogisticBootMarg

## Usage

```
prLogisticDelta(formula, cluster = FALSE,
  pattern = c("conditional", "marginal"),
  conf = 0.95, dataset, ...)
```

## Arguments

formula	a symbolic description of the model to be fitted. The details of model specification are given below.
cluster	logical argument specifying data clustering. The default is cluster=FALSE. If data is clustered or longitudinal, it should be set to cluster=TRUE.
pattern	the standardization procedure. If pattern is missing then conditional standardization is used. The standardization is set to be the marginal if pattern="marginal".
conf	scalar or vector specifying confidence level(s) for estimation. The default is conf = 0.95.
dataset	a required data frame containing the variables named in formula
...	optional additional arguments. Currently none are used in any methods.

## Details

A typical form used with glm() function is included in the formula argument as response ~ terms where response is the (binary) response vector and terms is a series of terms which specifies a linear predictor for response. The prLogisticDelta assumes a binomial family associated to the model. The glmer() function is used when a vertical bar character "|" separates an expression for a model matrix and a grouping factor. Currently only binary predictors are allowed. If categorization for predictors is other than (0,1), factor() should be considered.



**Value**

Returns prevalence ratio and its 95% confidence intervals.

**Author(s)**

Raydonal Ospina, Department of Statistics, Federal University of Pernambuco, Brazil  
(raydonal@de.ufpe.br)

Leila D. Amorim, Department of Statistics, Federal University of Bahia, Brazil  
(leiladen@ufba.br).

**References**

Localio AR, Margolis DJ, Berlin JA (2007). Relative risks and confidence intervals were easily computed indirectly from multivariate logistic regression. *Journal of Clinical Epidemiology*, **60**, 874-882.

Oliveira NF, Santana VS, Lopes AA (1997). Ratio of proportions and the use of the delta method for confidence interval estimation in logistic regression. *Journal of Public Health*, **31**(1), 90-99.

Santos CAST et al (2008). Estimating adjusted prevalence ratio in clustered cross-sectional epidemiological data. *BMC Medical Research Methodology*, **8** (80). Available from <http://www.biomedcentral.com/1471-2280/8/80>.

Wilcosky TC, Chambless LE (1985). A comparison of direct adjustment and regression adjustment of epidemiologic measures. *Journal of Chronic Diseases*, **34**, 849-856.

**See Also**

[glm](#), [glmer](#), [prLogisticBootCond](#), [prLogisticBootMarg](#)

**Examples**

```
### For independent observations:
# Estimates from logistic regression with conditional standardization -
# delta method
# Not run:
# data("titanic", package = "prLogistic")
# attach(titanic)
# prLogisticDelta(survived~ sex + pclass + embarked, data = titanic)
# End (Not run:)

# Estimates from logistic regression with marginal standardization -
# delta method
prLogisticDelta(survived~ sex + pclass + embarked,
data = titanic, pattern="marginal")

### For clustered data
# Estimates from random-effects logistic regression with conditional
# standardization - delta method
# Not run:
# data("Thailand", package = "prLogistic")
# prLogisticDelta(rgi~ sex + pped + (1|schoolid),
```

```
# data = Thailand, cluster=TRUE)
# End (Not run:)

# Estimates from random-effects logistic regression with marginal
# Not run:
# standardization - delta method
# prLogisticDelta(rgi ~ sex + pped + (1|schoolid), data = Thailand,
# pattern="marginal", cluster=TRUE)
# End (Not run:)
```

---

 Thailand

*Thailand Data*


---

### Description

Data are from a national survey of primary education in Thailand, including information for 8,582 sixth graders nested within 411 schools (Raudenbush & Bhumirat, 1992). The outcome binary variable "rgi" indicates whether a pupil has repeated a grade during primary education (0=no, 1=yes). The predictor variables are: child's sex (0=male, 1=female) and child's pre-primary education (0=no, 1=yes). Every level-1 record corresponds to a student.

### Usage

```
data(Thailand)
```

### Format

A data frame with 8,582 observations on the following 4 variables:

- schoolid** School identification.
- sex** Child's sex (0 if boy, 1 if girl).
- pped** Pre-primary education indicator (1=yes, 0=no).
- rgi** Repeated grade indicator (1=yes, 0=no).

### Source

The data come from  
 Raudenbush, S, Bhumirat C. (1992) The distribution of resources for primary education and its consequences for educational achievement in Thailand. *International Journal of Educational Research*, 17, 143-164.

### References

- Raudenbush et al (2000) HLM 5. Hierarchical Linear and Nonlinear Modeling. Chicago: Scientific Software International.
- Hox J (2002). Multilevel Analysis. Techniques and Applications. Mahwah, NJ: Lawrence Erlbaum Associates

## Examples

```
data("Thailand", package = "prLogistic")
attach(Thailand)
summary(Thailand)
```

---

titanic

*Titanic Data*

---

## Description

Data describes the survival status of individual passengers on the British luxury passenger liner Titanic that sank when it struck an iceberg on April 1912, on its maiden voyage to New York City from Southampton, England (Weisberg, 2003). Information were obtained from Encyclopedia Titanica. The variables on the dataset called titanic3 were made available by Harrell Jr (2001) and contains pclass, sex, embarked and survival status for each person on the board of the Titanic. pclass refers to passenger class (1st, 2nd, 3rd), and is a proxy for socio-economic class. embarked indicates the place where the passenger embarked on the ship. The variables pclass and embarked were recoded in the present version of the dataset. These data were used to compute prevalence ratios and corresponding confidence intervals.

## Usage

```
data(titanic)
```

## Format

A data frame with 1307 observations on the following 4 variables:

**pclass** The passenger class (1st, 2nd, 3rd) in the ship. Recoded to 1=1st class and 0=2nd or 3rd classes.

**survived** The survival status of the passenger (1=yes, 0=no).

**sex** Sex of the passenger (1=female, 0=male).

**embarked** The place where the passenger embarked for the trip (Q=Queenstown, S=Southampton, C=Cherbourg). Recoded to 1=Queenstown or Cherbourg and 0=Southampton.

## Source

The data were originally obtained from Dawson, Robert J. MacG. (1995), The Unusual Episode Data Revisited. Journal of Statistics Education, 3. <http://www.amstat.org/publications/jse/v3n3/datasets.dawson.html>

## References

Weisberg, S. (2005) Applied Linear Regression. Third Edition. New Jersey: John-Wiley & Sons.  
Frank E. Harrell Jr (2001) Regression Modelling Strategies. New York: Springer-Verlag.

### Examples

```
data("titanic", package = "prLogistic")
attach(titanic)
dim(titanic)
library(Hmisc)
describe(as.factor(pclass))
table(survived,pclass)
glm(survived ~ as.factor(pclass), family=binomial, data=titanic)
```

---

Toenail

*Toenail Data*

---

### Description

Data are from a multicenter study comparing two oral treatments for toe-nail infection, including information for 294 patients measured at 7 weeks, comprising a total of 1908 measurements. The outcome binary variable "onycholysis" indicates the degree of separation of the nail plate from the nail-bed (none or mild versus moderate or severe). Patients were evaluated at baseline (week 0) and at weeks 4, 8, 12, 24, 36, and 48 thereafter.

### Usage

```
data(Toenail)
```

### Format

A data frame with 1,908 observations on the following 5 variables:

**ID** Subject's identification.

**Response** Degree of onycholysis (0 if none or mild, 1 if moderate or severe).

**Treatment** Oral treatment indicator (1=Treatment A, 0= Treatment B).

**Month** Exact timing of measurements, in months.

**Visit** Visit number (visit numbers 1-7 correspond to scheduled visits at 0, 4, 8, 12, 24, 36, and 48 weeks).

### Source

The data come from

Data courtesy of Novoartis, Belgium (<http://www.blackwellpublishing.com/rss/datasets/C4827r.txt>)

## References

- De Backer, M., De Vroey, C., Lesaffre, E., Scheys, I., and De Keyser, P. (1998). Twelve weeks of continuous oral therapy for toenail onychomycosis caused by dermatophytes: A double-blind comparative trial of terbinafine 250 mg/day versus itraconazole 200 mg/day. *Journal of the American Academy of Dermatology*, 38, 57-63.
- Lesaffre, E. and Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: An example. *Journal of the Royal Statistical Society, Series C*, 50, 325-335.
- Rabe-Hesketh S, Skrondal A (2008). *Multilevel and Longitudinal Modeling Using Stata*. Mahwah, NJ: Lawrence Erlbaum Associates. Second Edition.

## Examples

```
data("Toenail", package = "prLogistic")
attach(Toenail)
summary(Toenail)
```

---

 UIS

*UIS - The UMARU IMPACT Study - Data*


---

## Description

Data were from randomized trials related to treatment for drug abuse obtained by the University of Massachusetts Aids Research Unit (UMARU) IMPACT Study (UIS). The main goal of the study was to compare treatment programs of different durations in the reduction of drug abuse and in the prevention of high-risk HIV behavior (Hosmer and Lemeshow, 2000). The variables on the dataset included here are age at enrollment, drug use history at admission, race, treatment group, treatment site, and patient's status at the end of the treatment program (remained drug free or otherwise). These data are copyrighted by John Wiley & Sons Inc. and are used here to illustrate the calculation of prevalence ratios and corresponding confidence intervals.

## Usage

```
data(UIS)
```

## Format

A data frame with 575 observations on the following 7 variables:

**ID** Identification code.

**Age** Age at enrollment (in years) recoded to 1= 32 years or younger, 0= more than 32 years-old.

**DrugUse** Drug use history at admission (1=never, 2=previous, 3=recent). Recoded to 1= never and 0= previous or recent.

**race** Subject's race (0=white, 1=other).

**trt** Treatment group (0=short, 1=long).

**site** Treatment site (0=A, 1=B).

**drugFree** Returned to drug use prior to the scheduled end of the treatment program (1=remained drug free, 0=otherwise).

**Source**

The data come from  
Hosmer and Lemeshow (2000) Applied Logistic Regression. 2nd Edition, New York: John-Wiley  
& Sons.

These data are copyrighted by John Wiley & Sons Inc.

We acknowledged it and use the data to illustrate the methodology included in this library.

**References**

Weisberg, S. (2005) Applied Linear Regression. Wiley-Interscience.

**Examples**

```
data("UIS", package = "prLogistic")
attach(UIS)
tmp=table(drugFree, trt)
options(digits=2)
prop.table(tmp, 2)
```

# Index

## \*Topic **datasets**

downer, 2

LBW, 3

Thailand, 10

titanic, 11

Toenail, 12

UIS, 13

## \*Topic **distribution**

prLogisticBootCond, 4

prLogisticBootMarg, 6

prLogisticDelta, 8

## \*Topic **regression**

prLogisticBootCond, 4

prLogisticBootMarg, 6

prLogisticDelta, 8

downer, 2

glm, 5, 7, 9

glmer, 5, 7, 9

LBW, 3

prLogisticBootCond, 4, 7, 9

prLogisticBootMarg, 5, 6, 9

prLogisticDelta, 5, 7, 8

Thailand, 10

titanic, 11

Toenail, 12

UIS, 13