Package ‘gamlss.data’

February 15, 2013

Description  Data for GAMLSS models.

Title  GAMLSS Data.

LazyData  yes

Version  4.2-0

Date  2012-09-29

Depends  R (>= 2.10)

Author  Mikis Stasinopoulos <d.stasinopoulos@londonmet.ac.uk>, Bob Rigby <r.rigby@londonmet.ac.uk>

Maintainer  Mikis Stasinopoulos <d.stasinopoulos@londonmet.ac.uk>

License  GPL-2 | GPL-3

URL  http://www.gamlss.org/

Repository  CRAN

Date/Publication  2012-09-29 21:02:01

NeedsCompilation  no

R topics documented:

abdom .................................................. 2
acidity .................................................. 3
aep ...................................................... 4
aids ..................................................... 5
alveolar ............................................... 6
CD4 .................................................... 7
computer ............................................. 7
db ...................................................... 8
dbbmi .................................................. 9
fabric ................................................. 10
Description

The abdom data frame has 610 rows and 2 columns. The data are measurements of abdominal circumference (response variable) taken from fetuses during ultrasound scans at Kings College Hospital, London, at gestational ages (explanatory variable) ranging between 12 and 42 weeks.

Usage

data(abdom)

Format

This data frame contains the following columns:

- y abdominal circumference: a numeric vector
- x gestational age: a numeric vector

Details

The data were used to derived reference intervals by Chitty et al. (1994) and also for comparing different reference centile methods by Wright and Royston (1997), who also commented that the distribution of Z-scores obtained from the different fitted models 'has somewhat longer tails than the normal distribution'.
**acidity**

**Source**

Dr. Eileen M. Wright, Department of Medical Statistics and Evaluation, Royal Postgraduate Medical School, Du Cane Road, London, W12 0NN.

**References**


**Examples**

```r
data(abdom)
attach(abdom)
plot(x,y)
detach(abdom)
```

---

**acidity**  
*The Acidity Data files for GAMLSS*

**Description**

The data shows the acidity index for 155 lakes in the Northeastern United States (previously analysed as a mixture of gaussian distributions on the log scale by Crawford *et al.* (1992, 1994)). These 155 observations are the log acidity indices for the lakes.

**Usage**

```r
data(acidity)
```

**Format**

A data frame with 155 observations on the following variable.

- **y** a numeric vector showing the acidity index for 155 lakes in the Northeastern United States

**References**


Examples

```r
data(acidity)
with(acidity, hist(y))
```

---

**The Hospital Stay Data**

**Description**

The data, 1383 observations, are from a study at the Hospital del Mar, Barcelona during the years 1988 and 1990, Gange et al. (1996).

**Usage**

```r
data(aep)
```

**Format**

A data frame with 1383 observations on the following 8 variables.

- `los`: the total number of days patients spent in hospital: a discrete vector
- `noinap`: the number of inappropriate days spent in hospital: a discrete vector
- `loglos`: the log(los/10): a numeric vector
- `sex`: the gender of patient: a factor with levels 1=male, 2=female
- `ward`: the type of ward in the hospital: a factor with levels 1=medical, 2=surgical, 3=others
- `year`: the specific year 1988 or 1990: a factor with levels `88` and `90`
- `age`: the age of the patient subtracted from 55: a numeric vector
- `y`: the response variable a matrix with 2 columns, the first is `noinap` the second is equal to (los-noinap)

**Details**

Gange et al. (1996) used a logistic regression model for the number of inappropriate days (`noinap`) out of the total number of days spent in hospital (`los`), with binomial and beta binomial errors and found that the later provided a better fit to the data. They modelled both the mean and the dispersion of the beta binomial distribution (BB) as functions of explanatory variables.

**Source**


**References**

**Examples**

```r
data(aep)
attach(aep)
pro<-noinap/los
plot(ward,pro)
rm(pro)
detach(aep)
```

---

**aids**

*Aids Cases in England and Wales*

---

**Description**

The quarterly reported AIDS cases in the U.K. from January 1983 to March 1994 obtained from the Public Health Laboratory Service, Communicable Disease Surveillance Centre, London.

**Usage**

```r
data(aids)
```

**Format**

A data frame with 45 observations on the following 3 variables.

- **y**: the number of quarterly aids cases in England and Wales: a numeric vector
- **x**: time in months from January 1983, 1:45: a numeric vector
- **qrt**: the quarterly seasonal effect a factor with 4 levels, [1=Q1 (Jan-March), 2=Q2 (Apr-June), 3=Q3 (July-Sept), 4=Q4 (Oct-Dec)]

**Details**

The counts y can be modelled using a (smooth) Poisson regression model in time x with the quarterly effects i.e. cs(x,df=7)+qrt. Overdispersion persists, so use a Negative Binomial distribution of type I or II. The data also can be used to find a break point in time, see Rigby and Stasinopoulos (1992).

**Source**

Public Health Laboratory Service, Communicable Disease Surveillance Centre, London.

**References**

alveolar

The Alveolar Data files for GAMLSS

Description

alveolar: alveolar-bronchiolar adenomas data used by Tamura and Young (1987) and also reproduce in Hand et al. (1994), data set 256. The data are the number of mice out of certain number of mice (the binomial denominator) in 23 independent groups, having alveolar-bronchiolar adenomas.

Usage

data(alveolar)

Format

Data frames each with the following variable.

r  a numeric vector showing the number of mice out of n number of mice (the binomial denominator below) in 23 independent groups, having alveolar-bronchiolar adenomas.

n  a numeric vector showing the total number of mice

Details

Data sets usefull for the GAMLSS booklet

References


Examples

data(alveolar)
with(alveolar, hist(r/n))
Description

CD4: The data were given by Wade and Ader (1994) and refer to cd4 counts from uninfected children born to HIV-1 mothers and the age of the child.

Usage

data(CD4)

Format

Data frames each with the following variable.

- cd4 a numeric vector showing the CD4 counts
- age the age of the child

Details

Data sets useful for the GAMLSS booklet

References


Examples

data(CD4)
with(CD4, plot(cd4~age))

computer

The Computer Failure Data files for GAMLSS

Description

computing: The data relate to DEC-20 computers which operated at the Open University in the 1980. They give the number of computers that broke down in each of the 128 consecutive weeks of operation, starting in late 1983, see Hand *et al.* (1994) page 109 data set 141.
Usage

data(computer)

Format

Data frames each with the following variable.

failure a numeric vector showing the number of times computers failed

Details

Data sets useful for the GAMLSS booklet

References


Examples

data(computer)
with(computer, plot(table(failure)))

---

db                     Head Circumference of Dutch Boys

Description

The data are coming from the Fourth Dutch Growth Study, Fredriks et al. (2000a, 2000b), which is a cross-sectional study that measures growth and development of the Dutch population between the ages 0 and 21 years. The study measured, among other variables, height, weight, head circumference and age for 7482 males and 7018 females. Here we have the only the head circumference of Dutch boys.

Usage

data(db)

Format

A data frame with 7040 observations on the following 2 variables.

head head circumference
age age in years

Source

The data were kindly given by professor Stef. van Buuren.
References


Examples

data(db)
attach(db)
plot(age,head)
detach(db)

dbbmi  
BMI of Dutch Boys

Description

The data are comming from the Fourth Dutch Growth Study, Fredriks et al. (2000a, 2000b), which is a cross-sectional study that measures growth and development of the Dutch population between the ages 0 and 21 years. The study measured, among other variables, height, weight, head circumference and age for 7482 males and 7018 females. Here we have the only the BMI of Dutch boys.

Usage

data(dbbmi)

Format

A data frame with 7294 observations on the following 2 variables.

age  a numeric vector
bmi  a numeric vector

Source

The data were kindly given by professor Stef. van Buuren.
References

Fredriks, A.M. van Buuren, S. Burgmeijer, R.J.F. Meulmeester, J.F. Beuker, R.J. Brugman, E.
Roede, M.J. Verloove-Vanhorick, S.P. and Wit, J. M. (2000a), Continuing positive secular change
Fredriks, A.M. van Buuren, S. Wit, J.M. and Verloove-Vanhorick, S. P. (2000b) Body index mea-
surments in 1996-7 compared with 1980, Archives of Childhood Diseases, 82, 107–112
reference curves. Statistics in Medicine, 20, 1259–1277

Examples

```r
data(dbbmi)
plot(bmi~age, data=dbbmi)
```

The Fabric Data

Description

The data are 32 observations on faults in rolls of fabric

Usage

```r
data(fabric)
```

Format

A data frame with 32 observations on the following 3 variables.

- `leng` the length of the roll : a numeric vector
- `y` the number of faults in the roll of fabric : a discrete vector
- `x` the log of the length of the roll : a numeric vector

Details

The data are 32 observations on faults in rolls of fabric taken from Hinde (1982) who used the EM
algorithm to fit a Poisson-normal model. The response variable is the number of faults in the roll of
fabric and the explanatory variable is the log of the length of the roll.

Source

John Hinde

References

Hinde, J. (1982) Compound Poisson regression models: in GLIM 82, Proceedings of the Inter-
York.
Examples

data(fabric)
attach(fabric)
plot(x,y)
detach(fabric)

---

**film30**

*Film revenue data for the 1930’s*

Description

Data from film revenues from the 1930s’.

Usage

```r
 data(film30)
```

Format

A data frame with 969 observations on the following 3 variables.

- **film**: a factor with the name of the film
- **total**: a numeric vector
- **opening**: a numeric vector

Source

The data were collected by Prof. John Sedgwick

References


Examples

```r
 data(film30)
 ## maybe str(film30) ; plot(film30) ...
 ```
Film revenue data for the 1990's

Description
Data from film revenues from the 1990s.'

Usage
data(film90)

Format
A data frame with 4031 observations on the following 14 variables.

- **time**: Date of release of the film
- **year**: the year of release of the film
- **month**: a factor, the month of release of the film
- **title**: a factor indicating the title of the film
- **borev**: real box office revenues, calculated in 1987 prices
- **lborev**: the log of real box office revenues, calculated in 1987 prices
- **nosc**: the number of screens
- **lnosc**: the log of the number of screens
- **boopen**: box office opening revenues calculated in 1987 prices
- **lboopen**: the log of box office opening revenues calculated in 1987 prices
- **borev1**: the box office revenues after the first week, borev0-boopen
- **lborev1**: the log of box office revenues after the first week
- **dist**: a factor indicating whether Independent or Major distributor
- **whetherCost**: a factor with levels 0 1 indication whether the original data has the cost of the film to make

Details
Those data are analysed in Voudouris et. al. (2011)

Source
Data collected by Prof. John Sedgwick

References
The Glass Data files for GAMLSS

Description

glass: show the strength of glass fibres, measured at the National Physical Laboratory, England, see Smith and Naylor (1987), (the unit of measurement were not given in the paper).

Usage

data(glass)

Format

Data frames each with the following variable.

strength a numeric vector showing the strength of glass fibres

Details

Data sets useful for the GAMLSS booklet

References


Examples

data(glass)
with(glass, hist(strength))
**Description**

There two data sets contain data used in Hodges (1998). In addition to the data used in that manuscript, it contains other data items.

The original data consists of two matrices of dimensions of 341x6 and a 45x4 respectively.

The first matrix hodges describes plans. The information for each plan is: the state, a two-character code that identifies plans within state, the total premium for an individual, the total premium for a family, the total enrollment of federal employees as individuals, and the total enrollment of federal employees as families.

The second matrix, hodges, describes states. The information for each state is: its two-letter abbreviation, the state average expenses per admission (from American Medical Association 1991 Annual Survey of Hospitals), population (1990 Census), and the region (from the Marion Merrill Dow Managed Care Digest 1991).

The Hodges manuscript used these variables: Plan level: individual premium, individual enrollment. State level: expenses per admission, region.

**Usage**

data(hodges)

**Format**

Two data frames the first with 341 observations on the following 6 variables.

- **state**: a factor with 45 levels AL AZ CA CO CT DC DE FL GA GU HI IA ID IL IN KS KY LA MA MD ME MI MN MO NC ND NE NH NJ NM NV NY OH OK OR PA PR RI SC TN TX UT VA WA WI
- **plan**: a two-character code that identifies plans within state declared here as factor with 325 levals.
- **prind**: a numeric vector showing the total premium for an individual
- **prfam**: a numeric vector showing the total premium for a family
- **enind**: a numeric vector showing the total enrollment of federal employees as individuals
- **enfam**: a numeric vector showing the total enrollment of federal employees as families.

and the second with 45 observations on the following 4 variables

- **State**: a factor with levels same as state above
- **expe**: a numeric vector showing the state average expenses per admission (from American Medical Association 1991 Annual Survey of Hospitals)
- **pop**: a numeric vector showing the population (1990 Census)
- **region**: the region (from the Marion Merrill Dow Managed Care Digest 1991), a factor with levels MA MT NC NE PA SA SC
Source

http://www.biostat.umn.edu/~hodges/

References


Examples

data(hodges)
attach(hodges)
plot(prind~state, cex=1, cex.lab=1.5, cex.axis=1, cex.main=1.2)
str(hodges)
data(hodges1)
str(hodges1)

LGAclaims

The LGA Claims Data files for GAMLSS

Description

These are several small data files useful for gamlss fits.

LGAclaims: the data were given by Gillian Heller and can be found in de Jong and Heller (2007). This data set records the number of third party claims, *Claims*, in a twelve month period between 1984-1986 in each of 176 geographical areas (local government areas) in New South Wales, Australia. Areas are grouped into thirteen statistical divisions (SD). Other recorded variables are the number of accidents, *Accidents*, the number of people killed or injured and population with all variables classified according to area.

Usage

data(LGAclaims)

Format

Data frames each with the following variable.

- **Claims**: the number of third party claims
- **LGA**: Local government areas in New South Wales
- **SD**: statistical divisions
- **Pop\_\_density**: population density
- **KI**: the number of people killed or injured
\textbf{Accidents}  the number of accidents  \\
\textbf{Population} population size  \\
L_{\text{\$KI}}  log of KI  \\
L_{\text{\$Accidents}}  the log of the number of accidents  \\
L_{\text{\$Population}}  log Population  \\

\textbf{Details}  \\
Data sets usefull for the GAMLSS booklet  \\

\textbf{References}  \\

\textbf{Examples}  \\
data(LGAclaims)  \\
with(LGAclaims, plot(data.frame(Claims, Pop\_density, KI, Accidents, Population)))  \\

\begin{verbatim}
 lice          Data files for GAMLSS
\end{verbatim}  \\

\textbf{Description}  \\
lice : The data come from Williams (1944) (also used by Stein and Juritz (1988).) and they are lice per head of Hindu male prisoners in Cannamore, South India, 1937-1939.  \\

\textbf{Usage}  \\

data(lice)  \\

\textbf{Format}  \\
Data frames each with the following variable.  \\
\textbf{head}  a numeric vector showing the number lice per head of Hindu male prisoners in Cannamore, South India, 1937-1939.  \\
\textbf{freq}  a numeric vector showing the frequency of lice per head  \\

\textbf{Details}  \\
Data sets usefull for the GAMLSS booklet
References


Examples
data(lice)

---

The Margolin Data files for GAMLSS

Description

Margolin: Margolin et al. (1981) present data from an Ames Salmonella assay, where y is the number of revertant colonies observed on a plate given a dose x of quinoline. The data were subsequently analysed by Breslow (1984), Lawless (1987) and Saha and Paul (2005).

Usage
data(margolin)

Format

Data frames each with the following variable.

- **y**: a numeric vector showing the number of revertant colonies observed on a plate given a dose x of quinoline.
- **x**: a numeric vector showing a dose x of quinoline.

Details

Data sets useful for the GAMLSS booklet

References


Examples
data(margolin)
with(margolin, plot(y~x))
Description

Mothers encouragement for participation in Higher Education. The response variable is `mums` a three level factor which can be used in a multinomial Logistic model or `mumsB` a two level factor suitable for binary logistic model.

Usage

data(Mums)

Format

A data frame with 871 observations on the following 7 variables.

- `mums` mothers encouragement: factor with levels 1 is for strong encouragement, 2 is for some encouragement and 3 for no encouragement/discouragement
- `class` social class: a factor with levels 1 is C1, 2 is C2, 3 is D and 4 is E
- `age` age of the participants: a factor with levels 1 is 16-18, 2 is 19-20 and 3 is 20-30
- `gender` a factor with levels 1 is male and 2 is female
- `ethn` ethnicity of the participants: a factor with levels 1 is white, 2 is black, 3 is asian and 4 is other
- `qual` qualifications of the participants: a factor with levels, 1 is greater or equal to 2 A levels, 2 is HND or more than 5 GCSE’s, 3 is less than 5 GSCSE’s ar none above and 4 no formal qualification
- `mumsB` mothers encouragement: a factor with levels, 0 is no encouragement or some encouragement 1 is for strong encouragement

Details

The data were collected as part of the Social Class and widening Participation in Higher Education Project based at the University of North London (now London Metropolitan University) and supported by the University’s Development and Diversity Fund over the period 1998-2000.

Source

Professor Robert Gilchrist director of STORM at London Metropolitan

References

Examples

data(Mums)
MM<-xtabs(~mums+qual, data=Mums)
mosaicplot(MM, color=TRUE)
MM<-xtabs(~mums+ethn+gender, data=Mums)
mosaicplot(MM, color=TRUE)

parzen

The Parzen Data File for GAMLSS

Description

Parzen: Parzen (1979) and also contained in Hand et al. (1994), data set 278. The data give the annual snowfall in Buffalo, NY (inches) for the 63 years, from 1910 to 1972 inclusive.

Usage

data(parzen)

Format

Data frames each with the following variable.

snowfall the annual snowfall in Buffalo, NY (inches) for the 63 years, from 1910 to 1972 inclusive, 63 observations

Details

Data sets usefull for the GAMLSS booklet

References


Examples

data(parzen)
with(parzen, hist(snowfall))
Poliomyelitis cases in US

Description

Poliomyelitis cases reported to the U.S. Centers for Disease Control for the years 1970 to 1983, that is, 168 observations.

Usage

data(polio)

Format

The format is: Time-Series [1:168] from 1970 to 1984: 0 1 0 0 1 3 9 2 3 5 ...

Details

The data were originally modelled by Zeger (1988) who used a parameter driven approach, in which a first order autoregressive model was used for the latent process, to conclude that there is evidence of a decrease in the polio infection rate. The data were analysed also by Li (1994), Zeger and Qaqish (1988), Davis et al. (1999), and by Benjamin et al (2003).

Source

Zeger (1988) w

References


Examples

data(polio)
plot(polio)
**Rent data**

**Description**

A survey was conducted in April 1993 by Infratest Sozialforschung. A random sample of accommodation with new tenancy agreements or increases of rents within the last four years in Munich was selected including: i) single rooms, ii) small apartments, iii) flats, iv) two-family houses. Accommodation subject to price control rents, one family houses and special houses, such as penthouses, were excluded because they are rather different from the rest and are considered a separate market. For the purpose of this study, 1967 observations of the variables listed below were used, i.e. the rent response variable $R$ followed by the explanatory variables found to be appropriate for a regression analysis approach by Fahrmeir *et al.* (1994, 1995):

**Usage**

data(rent)

**Format**

A data frame with 1969 observations on the following 9 variables.

- **R**: rent response variable, the monthly net rent in DM, i.e. the monthly rent minus calculated or estimated utility cost
- **Fl**: floor space in square meters
- **A**: year of construction
- **Sp**: a variable indicating whether the location is above average, 1, (550 observations) or not, 0, (1419 observations)
- **Sm**: a variable indicating whether the location is below, 1, average (172 obs.) or not, 0, (1797 obs.)
- **B**: a factor with levels indicating whether there is a bathroom, 1, (1925 obs.) or not, 0, (44 obs.)
- **H**: a factor with levels indicating whether there is central heating, 1, (1580 obs.) or not, 0, (389 obs.)
- **L**: a factor with levels indicating whether the kitchen equipment is above average, 1, (161 obs.) or not, 0, (1808 obs.)
- **loc**: a factor (combination of Sp and Sm) indicating whether the location is below, 1, average, 2, or above average 3

**Details**

This set of data were used by Stasinopoulos *et al.* (2000) to fit a model where both the mean and the dispersion parameter of a Gamma distribution were modelled using the explanatory variables.

**Source**

Provide by Prof. L. Fahrmeir
References


Examples

data(rent)
attach(rent)
plot(F1,R)

---

species

*The Fish Species Data files for GAMLSS*

Description

species: The number of different fish species ($y=fish$) was recorded for 70 lakes of the world together with explanatory variable $x=log(lake)$ area. The data are given and analyzed by Stein and Juritz (1988).

Usage

data(species)

Format

Data frames each with the following variable.

- **fish** a numeric vector showing the number of different species in 70 lakes in the word
- **lake** a numeric vector showing the lake area

Details

Data sets usefull for the GAMLSS booklet
References


Examples

```r
data(species)
with(species, plot(fish~log(lake)))
```

Data frames each with the following variable.

- **word**: a numeric vector showing the number a word appearing in a text.
- **freq**: a numeric vector showing the frequency of the number a word appearing in a text.

Details

Data sets useful for the GAMLSS booklet.

References


Examples

```r
data(stylo)
plot(freq~word, type="h", data=stylo)
```
tensile  

The Tensile Data files for GAMLSS

Description

tensile: These data come from Quesenberry and Hales (1980) and were also reproduced in Hand et al. (1994), data set 180, page 140. They contain measurements of tensile strength of polyester fibres and the authors were trying to check if they were consistent with the lognormal distribution. According to Hand et al. (1994) "these data follow from a preliminary transformation. If the lognormal hypothesis is correct, these data should have been uniformly distributed".

Usage

data(tensile)

Format

Data frames each with the following variable.

str  a numeric vector showing the tensile strength

Details

Data sets usefull for the GAMLSS booklet

References


Examples

data(tensile)
with(tensile,hist(str))
The Turkish stock exchange index

Description

The Turkish stock exchange index was recorded daily from 1/1/1988 to 31/12/1998. The daily returns, \( \text{ret} = \log \left( \frac{I_{i+1}}{I_i} \right) \), were obtained for \( i = 1,2,\ldots,2868 \).

Usage

data(tse)

Format

A data frame with 2868 observations on the following 4 variables.

- year the year
- month the month
- day the day
- ret day returns \( \text{ret}[t] = \log(\text{currency}[t]) - \log(\text{currency}[t-1]) \)
- currency the currency exchange rate
- tl day return \( \text{ret}[t] = \log_{10}(\text{currency}[t]) - \log_{10}(\text{currency}[t-1]) \)

References


Examples

data(tse)
plot(ts(tse$ret))

US air pollution data set

Description


Usage

data(usair)
Format

A data frame with 41 observations on the following 7 variables.

- **y**: a numeric vector: sulphur dioxide concentration in air mgs. per cubic metre in 41 cities in the USA
- **x1**: a numeric vector: average annual temperature in degrees F
- **x2**: a numeric vector: number of manufacturers employing >20 workers
- **x3**: a numeric vector: population size in thousands
- **x4**: a numeric vector: average annual wind speed in miles per hour
- **x5**: a numeric vector: average annual rainfall in inches
- **x6**: a numeric vector: average number of days rainfall per year

Source

Hand et al. (1994) data set 26, USAIR.DAT, originally from Sokal and Rohlf (1981)

References


Examples

data(usair)
str(usair)
plot(usair)

# a possible gamlss model
# gamlss(library)
# ap<-gamlss(y~cs(x1,2)+x2+x3+cs(x4,2)+x5+cs(x6,3)+x4:x5,
# data=usair, family=GA(mu.link="inverse"))
#

---

**vas5**

**Visual analog scale (VAS) data**

Description

In the original data 368 patients, measured at 18 times after treatment with one of 7 drug treatments (including placebo), plus a baseline measure (time=0) and one or more pre-baseline measures (time=-1). Here for illustration we will ignore the repeated measure nature of the data and we shall use data from time 5 only (364 observations). The VAS scale response variable, Y, is assumed to be distributed as BEINF(mu, sigma, nu, tau) where any of the distributional parameters mu, sigma, nu and tau are modelled as a constant or as a function of the treatment.

Usage

data(vas5)
VictimsOfCrime

Format
A data frame with 364 observations on the following 3 variables.

- patient: a factor indicating the patient
- treat: the treatment factor with levels 1 2 3 4 5 6 7
- vas: the response variable

Details
The Visual analog scale is used to measure pain and quality of life. For example, patients are required to indicate a value from 0 to 100 for the amount of discomfort they have. This can be easily translated to a value from 0 to 1 and consequently analyzed using the beta distribution. Unfortunately if 0’s or 100’s are recorded the beta distribution is not appropriate since the values 0 and 1 are not allowed in the definition of the beta distribution. Note that the inflated beta distribution allows values at 0 and 1. This is a mixed distribution (continuous and discrete) having four parameters, nu for modelling the probability at zero \( p(Y=0) \) relative to \( p(0<Y<1) \), tau for modelling the probability at one \( p(Y=1) \) relative to \( p(0<Y<1) \), and mu and sigma for modelling the between values, \( 0<Y<1 \), using a beta distributed variable \( BE(mu, sigma) \) with mean \( mu \) and variance \( sigma*mu*(1-mu) \).

Source
The data were provided by Dr. Peter Lane

Examples
data(vas5)

Description
The data shows whether victims of crime were reported in the local media.

Usage
data(VictimsOfCrime)

Format
A data frame with 10590 observations on the following 2 variables.

- reported: Whether the crime was reported in local media.
- age: the age of the victim

Details
Whether the crime was reported in local media.
Source

The data were given by Prof Brian Francis of Lancaster University. They can be used to demonstrate the usefulness of smoothing techniques with a binary response variable.

References


Examples

data(VictimsOfCrime)
Index

*Topic datasets
  abdom, 2
  acidity, 3
  aep, 4
  aids, 5
  alveolar, 6
  CD4, 7
  computer, 7
  db, 8
  dbbmi, 9
  fabric, 10
  film30, 11
  film90, 12
  glass, 13
  hodges, 14
  hodges1 (hodges), 14
  LGAclaims, 15
  lice, 16
  margolin, 17
  Mums, 18
  parzen, 19
  polio, 20
  rent, 21
  species, 22
  stylo, 23
  tensile, 24
  tse, 25
  usair, 25
  vas5, 26
  VictimsOfCrime, 27

abdom, 2
acidity, 3
aep, 4
aids, 5
alveolar, 6
CD4, 7
computer, 7
db, 8