Package 'mdhglm'

October 13, 2022

Type Package

Title Multivariate Double Hierarchical Generalized Linear Models

Version 1.8

Date 2018-10-01

Author Youngjo Lee, Marek Molas, Maengseok Noh

Maintainer Maengseok Noh <msnoh@pknu.ac.kr>

Description Allows various models for multivariate response variables where each response is assumed to follow double hierarchical generalized linear models. In double hierarchical generalized linear models, the mean, dispersion parameters for variance of random effects, and residual variance can be further modeled as random-effect models.

Depends R (>= 3.4.0), methods, Matrix, boot, mvtnorm

License Unlimited

NeedsCompilation no

Repository CRAN

Date/Publication 2018-10-25 08:00:19 UTC

R topics documented:

---

mdhglm-package   *Multivariate Double Hierarchical Generalized Linear Models*

Description

This package allows different models for multivariate response variables where each response is assumed to follow DHGLMs (Lee and Nelder, 2006; Lee, Nelder and Pawitan, 2006, 2016). With DHGLMs, the mean and dispersion parameters for the variance of random effects and residual variance (overdispersion) can be modeled as random-effect models. Furthermore, the mdhglm package allows various correlation structures among the random effects in DHGLMs of different response variables. Functions are provided to format and summarize the mdhglm results. The estimates of fixed effects, random effects and their variances and correlations are calculated.

Details

Package:	mdhglm
Type:	Package
Version:	1.6
Date:	2016-09-19
License:	Unlimited
LazyLoad:	yes

This is version 1.4 of the mdhglm package.

Author(s)

Youngjo Lee, Marek Molas, Maengseok Noh

Maintainer: Maengseok Noh <msnoh@pknu.ac.kr>

References

Lee, Y. and Nelder, J.A. (2006). Double hierarchical generalized linear models (with discussion), Applied Statistics 55, 139–185.

Lee, Y., Nelder, J.A. and Pawitan, Y. (2006). Generalized Linear Models with Random Effects. Chapman and Hall, Boca Raton, 1st edition.

Lee, Y., Nelder, J.A. and Pawitan, Y. (2016). Generalized Linear Models with Random Effects (2nd Edition). Chapman and Hall, Boca Raton, 2nd edition.

---

aids   *Repateard Measures on AIDS Data*

Description

Both longitudinal and survival data were collected in a recent clinical trial to compare the efficacy and safety of two antiretroviral drugs in treating patients who had failed or were intolerant of zidovudine (AZT) therapy (Rizopoulos, 2015). In this trial, 467 HIV-infected patients were enrolled and randomly assigned to receive either didanosine (ddI) or zalcitabine (ddC). The number of CD4 cells per cubic millimeter of blood were recorded at study entry, and again at the 2, 6, 12, and 18 month visits.

Usage

```
data("aids")
```

Format

A data frame with 1405 observations on the following 12 variables.

patient patients identifier for 467 patients

Time time to death or censoring

death status with 0 denoting censoring and 1 death for patient

CD4 CD4 cells count

obstime time points at which the CD4 cells count was recorded

drug a factor with levels ddC (zalcitabine) and ddI (didanosine)

gender a factor with levels female and male

prevOI a factor with levels noAIDS (previous opportunistic infection at study entry) and AIDS (no previous infection)

AZT a factor with levels `intolerance` (AZT intolerance) and `failure` (AZT failure)

start start of time in the interval

stop end of time in the interval

event status with 0 denoting censoring and 1 death in the interval

References

Rizopoulos, D. (2015). JM: Joint Modeling of Longitudinal and Survival Data. R package version 1.4-2.

---

aids.id	*Survival Time Data for each of Patients on the AIDS Data*

Description

In the AIDS data, survival data can be summarized for each of patients. Times to death were also recorded with 40 percent of censoring rate.

Usage

```
data("aids.id")
```

Format

A data frame with 467 observations on the following 12 variables.

patient patients identifier for 467 patients

Time time to death or censoring

death status with 0 denoting censoring and 1 death for patient

CD4 CD4 cells count

obstime time points at which the CD4 cells count was recorded

drug a factor with levels `ddC` (zalcitabine) and `ddI` (didanosine)

gender a factor with levels `female` and `male`

prevOI a factor with levels `noAIDS` (previous opportunistic infection at study entry) and `AIDS` (no previous infection)

AZT a factor with levels `intolerance` (AZT intolerance) and `failure` (AZT failure)

start start of time in the first interval

stop end of time in the first interval

event status with 0 denoting censoring and 1 death in the first interval

References

Rizopoulos, D. (2015). JM: Joint Modeling of Longitudinal and Survival Data. R package version 1.4-2.

---

bacteria   *Bacteria Data*

Description

Tests of the presence of the bacteria H. influenzae in children with otitis media in the Northern Territory of Australia (Ripley, et al., 2016).

Usage

```
data("bacteria")
```

Format

A data frame with 220 observations on the following 6 variables.

y a factor with levels n (absence of bacteria) and y (preence of bacteria)

ap a factor with levels a (active) and p (placebo)

hilo a factor with levels hi (high compliance) and lo (low compliance)

week week of test

ID subjects identifier

trt a factor with levels placebo, drug and drug+

References

Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A. and Firsth, D. (2016). MASS: Support Functions and Datasets for Venables and Ripley's MASS. R package version 7.3-45.

---

bladder   *Bladder Cancer Data*

Description

Therneau and Lumley (2015) reported data on recurrences of bladder cancer, which were used to demonstrate methodology for recurrent event modelling (Wei et al., 1989). This is the data set from 85 subjects who were assigned to either thiotepa or placebo, and reports the first four recurrences for any patient. The stop variable is the tumor recurrence or censoring time in month. The event variable is 1 for recurrence and 0 for everything else (including death for any reason).

Usage

```
data("bladder")
```

Format

A data frame with 340 observations on the following 7 variables.

id patients identifier for 85 patients

rx treatment; 1 (=placebo) and 2 (=thiotepa)

number initial number of tumours (8=8 or more)

size size (cm) of largest initial tumour

stop recurrence or censoring time

event event variable with 1 for recurrence and 0 for everythis else (including death for any reason)

enum which recurrence (up to 4)

References

Therneau, T. and Lumley, T. (2015). survival: Survival Analysis. R package version 2.38-3.

Wei, L.J., Lin, D.Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the Americal Statistical Association, 84, 1065–1073.

---

cake *Cake Data*

Description

In an experiment on the preparation of chocolate cakes, conducted at Iowa State College, 3 recipes for preparing the batter were compared (Cochran and Cox, 1957). Recipes I and II differed in that the chocolate was added at 40C and 60C, respectively, while recipe III contained extra sugar. In addition, 6 different baking temperatures were tested: these ranged in 10C steps from 175C to 225C. For each mix, enough batter was prepared for 6 cakes, each of which was baked at a different temperature. Thus the recipes are the whole-unit treatments, while the baking temperatures are the sub-unit treatments. There were 15 replications, and it will be assumed that these were conducted serially according to a randomized blocks scheme: that is, one replication was completed before starting the next, so that differences among replicates represent time differences. A number of measurements were made on the cakes. The measurement presented here is the breaking angle. One half of a slab of cake is held fixed, while the other half is pivoted about the middle until breakage occurs. The angle through which the moving half has revolved is read on a circular scale. Since breakage is gradual, the reading tends to have a subjective element.

Usage

```
data("cake")
```

Format

A data frame with 270 observations on the following 5 variables.

`Replicate` replication number for 1...15 replicates

`Batch` recipes number for 1...3 recipes

`Recipe` recipes number for 1...3 recipes

`Temperature` baking temperature

`Angle` breaking angle

References

Cochran, W.G. and Cox, G.M. (1957). Experimental Designs. John Wiley and Sons, New York, 2nd edition.

---

cgd   *CGD Infection Data*

Description

The CGD data set presented by Fleming and Harrington (1991) and Therneau and Lumley (2015) consists of a placebo-controlled randomized trial of gamma interferon (rIFN-g) in the treatment of chronic granulomatous disease (CGD). 128 patients from 13 centers were tracked for around 1 year. The number (i.e., cluster size) of patients per center ranged from 4 to 26. The survival times are the recurrent infection times of each patient from the different centers. Censoring occurred at the last observation for all patients, except one, who experienced a serious infection on the data he left the study; in the CGD study about 63 per cent of the data were censored. The recurrent infection times for a given patient are likely to be correlated. However, each patient belongs to one of the 13 centers; hence, the correlation may also be attributed to a center effect.

Usage

```
data("cgd")
```

Format

A data frame with 203 observations on the following 16 variables.

`id` subjects identifier for 128 patients

`center` enrolling center for 13 centers

`random` date of randomization

`treat` a factor with levels `placebo` and `rIFN-g` (gamma interferon)

`sex` a factor with levels `male` and `female`

`age` age in years at study entry

`height` height in cm at study entry

weight weight in kg at study entry

inherit a factor for pattern of inheritance with levels X-linked and autosomal

steroids use of steroids at study entry with 1 (=yes) and 0(=no)

propylac use of prophylactic antibiotics at study entry

hos.cat a factor for centers into 4 goups with levels US:NIH, US:other, Europe:Amsterdam and Europe:other

tstart start of each time interval

enum observation number within subject

tstop end of each time interval

status status with 1 for infection and 0 for censoring

References

Fleming, T.R. and Harrington, D.P. (1991). Counting Processes and Survival Analysis. John Wiley and Sons, New York.

Therneau, T. and Lumley, T. (2015). survival: Survival Analysis. R package version 2.38-3.

---

circuit   *Integrated Circuit Data*

Description

An experiment on integrated circuits was reported by Phadke et al. (1983). The width of lines made by a photoresist-nanoline tool were measured in five different locations on silicon wafers, measurements being taken before and after an etching process being treated separately. Here, the pre-etching data are analyzed. The eight experimental factors (A-H) were arranged in an L18 orthogonal array and produced 33 measurements at each of five locations, giving a total of 165 observations. There were no whole-plot (i.e. between-wafer) factors.

Usage

```
data("circuit")
```

Format

A data frame with 165 observations on the following 13 variables.

Width width of lines made by a photoresist-nanoline tool

A a factor A with levels 1, 2 and 3

B a factor B with levels 1, 2 and 3

C a factor C with levels 1, 2 and 3

D a factor D with levels 1, 2 and 3

E a factor E with levels 1, 2 and 3

F a factor F with levels 1, 2 and 3

G a factor G with levels 1, 2 and 3

H a factor H with levels 1, 2 and 3

I a factor I with levels 1, 2 and 3

Wafer wafers indenfifier

waf replication number for two replicates

exp experiment number for differenct array

References

Phadke, M.S., Kacka, R.N., Speeney, D.V. and Grieco, M.J. (1983). Off-line quality control for integrated circuit fabrication using experimental design. Bell System Technical Journal, 62, 1273–1309.

---

cog *Vascular Cognitive Impairment Data*

Description

Lee, Nelder, and Pawitan (2016) considered the Vascular Cognitive Impairment (VCI) data. As the elderly population increases, the proportion of stroke, one of main geriatric diseases also rises. The increase of the number of patients with stroke results in the rise of that of patients with the VCI which cognitive function is declined due to stroke. Through an early intervention for the VCI, the cognitive function can be improved. They considered standardized VCI scores for four domains (response variables): executive (y1), memory (y2), visuospatial (y3) and language (y4). The purpose of the study is to examine the effects of ten demographic and ten acute neuroimaging variables on the cognitive function in the ischemic stroke patients.

Usage

```
data("cog")
```

Format

A data frame with 1139 observations on the following 26 variables.

x1 integer of age/10

x2 =1 (male), =0 (female)

x3 =0 (none), =1 (elementary) , =2 (middle), =3 (high), =4 (over college)

x4 =1 (hypertension), =0 (none)

x5 =1 (diabetes mellitus), =0 (none)

x6 =1 (atrial brillation), =0 (none)

x7 =1 (history of stroke), =0 (none)

x8 national institute of health stroke scale score at admission

x9 time interval from stroke onset to
       rst K-VCIHS-NP

x10 =1 (IQCODE>=3.6), =0 (otherwise)

x11 left or bilateral involvement

x12 lesion multiplicity in acute DWI imaging

x13 cortical involvement of acute lesions

x14 cortical involvement of chronic territorial infarction

x15 Periventricular white matter lesions (PVWM); =0 (PVWN 0, 1), =1 (PVWN 2, 3)

x16 Subcortical white matter lesions (SCWM); = 0 (SCWM 0, 1), = 1 (SCWM 2, 3)

x17 presence of lacunes

x18 presence of cerebral microbleeds

x19 Medial temporal lobe atrophy (MTA); = 1 (MTA 2), = 0 (not 2)

x20 = 1 (MTA 3, 4), = 0 (not 3, 4)

y1 executive

y2 memory

y3 visuospatial

y4 language

id patients identifier

References

Lee, Y., Nelder, J.A. and Pawitan, Y. (2016). Generalized Linear Models with Random Effects (2nd Edition). Chapman and Hall, Boca Raton, 2nd edition.

---

crack_growth   *Crack Growth Data*

Description

Hudak et al. (1978) presented data on crack growth data, which have been listed by Lu and Meeker (1993). There are 21 metallic specimens, each subjected to 120,000 loading cycles, with the crack lengths recorded every 10,000 cycles. We take t=no. cycles/100,000 here, so $t(j)=j/1000$ for $j=1,...,12$ coded by cycle. The crack increment sequences look rather irregular. Let $l(i,j)$ be the crack length of the ith specimen at the jth observation and let $y(i,j)=l(i,j)-l(i,j-1)$ be the corresponding increment of crack length, which has always a positive value. $l(i,j-1)$ is coded by crack0 to use as a covariate.

Usage

```
data("crack_growth")
```

Format

A data frame with 241 observations on the following 5 variables.

ID id for each 241 observations

y increment in crack length, y(i,j)=l(i,j)-l(i,j-1), where l(i,j) is the crack length of the ith specimen at the jth observation

crack0 previous crack length, l(i,j-1)

specimen specimen number for 21 metallic specimens

cycle number of cycles/100,000; t(j)=j/100, j=1,...,12

References

Hudak, S.J., Saxena, A. Bucci, R.J. and Malcom, R.C. (1978). Development of standard methods of testing and analyzing fatigue crack growth rate data. Technical report. AFML-TR-78-40. Westinghouse R and D Center, Wesinghouse Electric Corp., Pittsburgh, PA.

Lu, C.J. and Meeker, W.Q. (1993). Using degeneration measurements to estimate a time-to-failure distribution, Technometrics, 35, 161–174.

DHGLMMODELING	*Defining the Fixed and Random Models for the Mean and Dispersion parameters in DHGLMs*

Description

The DHGLMMODELING specifies a GLM, HGLM, DHGLM model for the mean parameters (mu), and a GLM, HGLM model for the overdispersion parameters (phi). GLM for mu, and GLM for phi are specified by adding only fixed terms to the linear predictors for the mu and phi, respectively. These are extended to HGLM for mu and HGLM for phi by adding some random terms. The DHGLM for mu is specified by allowing random effects for the variance of random effects in HGLM for mu. The LMatrix argument allows correlation structures to be defined for random terms. This is done by setting LMatrix to a matrix L that is used as a post-multiplier for the Z matrix of the corresponding random term.

Usage

```
DHGLMMODELING(Model="mean",Link=NULL,LinPred="constant",RandDist=NULL,
Offset=NULL,LMatrix=NULL,LinkRandVariance=NULL,LinPredRandVariance=NULL,
RandDistRandVariance="gaussian",
LinkRandVariance2=NULL,LinPredRandVariance2=NULL)
```

Arguments

Model	This option specifies a GLM, HGLM or DHGLM model for mu when Model="mean" (default), and a GLM or HGLM for phi when Model="dispersion".
Link	The link function for the linear predictor is specified by the option Link. For Model="mean", Link can be "identity", "logit", "probit", "cloglog", "log", or "inverse". For Model="dispersion", the choice is either "log" or "inverse". The default, specified as NULL link, is "identity" for Model="mean" and "log" for Model="dispersion".
LinPred	The option LinPred specifies the fixed and random terms for the linear predictor for mu when specified as Model="mean" or for phi when Model="dispersion". For Model="mean", LinPred=y~x1+x2+(1|id1)+(1|id2) specifies y as the main response, x1 and x2 as fixed covariates and id1 and id2 as random terms. For Model="dispersion", the main response should be phi, e.g. phi~x1+x2+(1|id1)+(1|id2). This option can specify the model without random effects, e.g., LinPred=phi~x1+x2. The default is "constant", which is set to intercept only the model for the corresponding linear predictors.
RandDist	The option RandDist specifies the distributions of the random terms represented in the option LinPred. It is set as a vector of distribution names from "gaussain" (default), "beta", "gamma", or "inverse-gamma" when Model="mean". For Model="dispersion", the choice is "gaussian" (default), "gamma", or "inverse-gamma". When more than one random terms are specified, e.g., y~x1+x2+(1|id1)+(1|id2) in the option LinPred, the different distributions for each random term can be specified, e.g., c("gaussian", "gamma"), which assumes normal distribution for the random term "id1" and gamma distribution for the random term "id2", respectively.
Offset	The option Offset can be used to specify a known component to be included in the linear predictor specified by LinPred during fitting. This should be the default (NULL) or a numeric vector of length equal to that of the appropriate data.
LMatrix	The option LMatrix sets a matrix that is used as a post-multiplier for the model matrix of the corresponding random effects. This option allows correlation structures to be defined for random effects. For example, when specified as Model="mean" and Lmatrix=L1+L2, the linear predictor for mu takes X beta + Z1 L1 r1 + Z2 L2 r2, where Z1 and Z2 are the model matrices for the random effects v1=L1 r1 and v2=L2 r2, specified in the option LinPred.
LinkRandVariance	The option LinkRandVariance specifies the link function for the linear predictor of the random-effect variances. The choice is either "log" (default) or "inverse". When more than two random terms are specified in the option LinPred, the user can set different link functions, e.g., LinkRandVariance=c("log","inverse") for each random term.
LinPredRandVariance	The option LinPredRandVariance specifies the fixed and random terms for the linear predictor of the random-effect variances for Model="mean". When y~x1+x2+(1|id1)+(1|id2) is specified in the option LinPred, LinPredRandVariance=c(lambda~xx1+(1|id11),lambda~xx2+(1|id12)) specifies xx1 and xx2 as fixed covariates and id11 and id12 as random terms in the lienar

predictors for the variances of the random terms id1 and id2, respectively. For Model="dispersion", the random term is not allowed in the linear predictor of the random-effect variance. The default (NULL) is set to intercept only model for the corresponding linear predictors.

RandDistRandVariance

The option RandDistRandVariance specifies the distributions for the random terms in the LinPredRandVariance. The choice is "gaussian" (default), "gamma", or "inverse-gamma".

LinkRandVariance2

This option specifies the link function for the linear predictor of the variance of random effects, which are specified in the option LinPredRandVariance. The choice is either "log" (default) or "inverse".

LinPredRandVariance2

This option specifies the fixed terms for the linear predictor of the variance of random effects, which is specified in the option LinPredRandVariance. For example, when LinPredRandVariance=c(lambda~xx1+(1|id11),lambda~xx1+(1|id12)) is specified, LinPredRandVariance2=c(~xxx1,~xxx2) specifies xxx1 and xxx2 as fixed covariates for the linear predictor of random-effect variances for id11 and id12, respectively. The default (NULL) is set to constant variance for the random effects in LinPredRandVariance.

---

eg *Ethylene Glycol Data*

Description

Price et al. (1985) presented data from a study on the developmental toxicity, ethylene glycol (EG),in mice. Responses are the malformation (binary response) and fetal weight (continuous response) outcomes for the experiment, which shows clear dose-related trends with respect to both outcomes.

Usage

```
data("eg")
```

Format

A data frame with 1028 observations on the following 5 variables.

litter litter number for 94 mice

dose dose input

y1 fetal weight (g)

y2 malformation (0; No, 1; Yes)

dose2 dose*dose

References

Price, C.J., Kimmel, C.A., Tyl, R.W. and Marr. M.C. (1985). The developmental toxicity of ethylene glycol in rats and mice. Toxicolal Applied Pharmacology, 81, 113—127.

---

epilepsy	*Epilepsy Seizures Data*

Description

Thall and Vail (1990) presented longitudinal data from a clinical trial of 59 epileptics, who were randomized to a new drug or a placebo (T=0 or T=1). Baseline data were available at the start of the trial; the trial included the logarithm of the average number of epileptic seizures recorded in the 8-week period preceding the trial (B), the logarithm of age (A), and visit (V: a linear trend, coded (-3,-1,1,3)/10). A multivariate response variable (y) consists of the seizure counts during 2-week periods before each of four visits to the clinic.

Usage

```
data("epilepsy")
```

Format

A data frame with 236 observations on the following 7 variables.

y   seizure counts during 2-week periods before each of four visits to the clinic

T   treatment (0=new drug, 1=placebo)

B   average number of epileptic seizures in the 8-week period preceding the trial

A   logarithm of age (in years) of each patient

V   linear trend coded (-3, -1, 1, 3)/10 for four visits of each patient

patient   patient number for 59 epileptics

id   observation number for 236 observations

References

Thall, P.F. and Vail, S.C. (1990). Some covariance models for longitudinal count data with overdispersion, Biometrics, 46, 657–671.

---

exch   *Pound-dollar Exchange-rate Data*

Description

Harvey et al. (1994) presented daily observations for the weekday closing exchange rates for the U.K. Sterling/U.S. Dollar from 1/10/81 to 28/6/85.

Usage

```
data("exch")
```

Format

A data frame with 944 observations on the following 4 variables.

rt exchange rate at time t

yt mean-correted returns

yt1 one lag of yt

yt12 yt1*yt1

References

Harvey, A.C., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models. Review of Economic Studies, 61, 247–264.

---

fabric   *Fabric Data*

Description

In the fabric data (Bissell, 1972), the main response is the number of faults in a bolt of fabric of length.

Usage

```
data("fabric")
```

Format

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

x log of length

y number of faults

rf obervation number for 32 obervations

References

Bissell, A.F. (1972). A negative binomial model with varying element sizes. Biometrika, 59, 435–441.

---

factor	*Binary Factor Data*

Description

Lee, Nelder and Pawitan considered the binary factor data with six items taking 1 or 0 for 350 subjects.

Usage

```
data("factor")
```

Format

A data frame with 350 observations on the following 7 variables.

y1 item 1

y2 item 2

y3 item 3

y4 item 4

y5 item 5

y6 item 6

subject subjects idenfier

References

Lee, Y., Nelder, J.A. and Pawitan, Y. (2016). Generalized Linear Models with Random Effects (2nd Edition). Chapman and Hall, Boca Raton, 2nd edition.

---

gas   *Gas Consumption Data*

Description

Durbin and Koopman (2000) considered the lagged quarterly-demand for gas in the UK from 1960 to 1986.

Usage

```
data("gas")
```

Format

A data frame with 108 observations on the following 4 variables.

y lagged quarterly-demand for gas

year year; 1960...1986

quarter quarter; 1...4

time data number for 108 obervations

References

Durbin, J. and Koopman, S.J. (2000). Time series analysis of nongaussian observations based on state space models from both classical and bayesian perspectives (with discussion). Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62, 3–56.

---

injection   *Injection Moulding data*

Description

Engel (1992) presented data from an injection-moulding experiment. An industrial Taguchi experiment was performed to study the in uence of several controllable factors on the mean value and the variation in the percentage of shrinkage of products made by injection moulding. The responses y were percentages of shrinkage of products made by injection moulding. There are seven controllable factors (A-G), in a 2^(7-4) fractional factorial design. There are 8 combinations of levels for seven controllable factors (A-G) to allow the estimates of seven main effects for controllable factors. At each setting of the controllable factors, 4 observations were obtained from a 2^(3-1) fractional factorial with three noise factors (M-O).

Usage

```
data("injection")
```

Format

A data frame with 32 observations on the following 11 variables.

y percentages of shrinkage of products

a a controllable factor A with levels 1 and 2

b a controllable factor B with levels 1 and 2

c a controllable factor C with levels 1 and 2

d a controllable factor D with levels 1 and 2

e a controllable factor E with levels 1 and 2

f a controllable factor F with levels 1 and 2

g a controllable factor G with levels 1 and 2

m a noise factor M with levels 1 and 2

n a noise factor N with levels 1 and 2

o a noise factor O with levels 1 and 2

References

Engel, J. (1992). Modelling variation in industrial experiments. Applied Statistics, 41, 579–593.

---

jointfit   *Fitting Multivariate Double Hierarchical Generalized Linear Models*
                                  using h-likelihood Approach

---

Description

The jointfit is used to fit a multivariate double hierarchical generalized linear models (MDHGLMs) allowing different models for multivariate response variables where each response follow DHGLM. A variety of distributions and link functions for both response and the random effects are allowed. Fixed and random effects can also be fitted in both the mean and the dispersion components. To call the fitting function jointfit, models for the mean and dispersion must be specified by DHGLM-MODELING object preferably created by calling the DHGLMMODELING function.

Usage

```
jointfit(RespDist="gaussian",BinomialDen=NULL, DataMain, MeanModel,DispersionModel=NULL,
PhiFix=NULL,LamFix=NULL,structure="correlated",mord=0,dord=1,convergence=1e-05,
Init_Corr=NULL, EstimateCorrelations=TRUE, ZZCorr=NULL, factor=NULL, REML=TRUE,order=1)
```

Arguments

RespDist	The distribution of the response is set by the option RespDist. The user can set it to: "gaussian" (default), "binomial", "poisson", or "gamma".
BinomialDen	When RespDist="binomial", one should use the option BinomialDen to specify the denominator for the binomial distribution. This should be "NULL" (default) or a numeric vector of length equal to the length of DataMain. When specified as BinomialDen=NULL and RespDist="binomial", the denominator is 1.
DataMain	The option DataMain determines the data frame to be used (non-optional).
MeanModel	For the mean model, this option requires DGHLMMODELING object which should specified by the option Model="mean".
DispersionModel	
	For the overdispersion model, this option requires DGHLMMODELING object which should be specified by the option Model="dispersion".
PhiFix	The option for overdispersion parameters (phi) to be estimated or maintaned constant. Specifying defaults such as PhiFix =NULL implies that phi is to be estimated. If not, phi is fixed at a value specified by PhiFix.
LamFix	The option for random-effect variance (lambda) to be estimated or maintaned constant. Specifying defaults such as LamFix =NULL implies that lambda is to be estimated. If not, lambda is fixed at a value specified by LamFix.
structure	The option structure determines structure of random effects. When structure="correlated" (or "shared"), correlated (or "shared") random-effects model is specified. Factor analysis can be fitted by specifying structure="factor". Furthremore, selection model for missing data can be also fitted by specifying structure="selection".
mord	The option mord specifies the order of Laplace approximation to the marginal likelihood for fitting the mean parameters. The choice is either 0 or 1 (default).
dord	The option dord specifies the order of adjusted profile likelihood for fitting the dispersion parameters. The choice is either 1 (default) or 2.
convergence	Setting this option determines the criterion for convergence, which is computed as the absolute difference between the values of all the estimated parameters in the previous and current iterations. The default criterion is 1e-06.
Init_Corr	Setting initial values of correlation (or shared parameters) between random effects
EstimateCorrelations	
	Correlation are estimated or fixed when EstimateCorrelations=TRUE (default) or EstimateCorrelations=FALSE
ZZCorr	List of model matrices for random effects
factor	factor structure when structure="factor"
REML	Giving REML estimates when REML=TRUE (default) or ML estimates when REML=FALSE
order	first order approximation when order=1 (default) or second order approximation when order=2 for factor analysis

Examples

```
data(eg)
eg1<-eg[1:100,] ## using sampled data to have faster results

jm1<-DHGLMMODELING(Link="identity", LinPred=y1~dose+dose2+(1|litter),RandDist="gaussian")
jm2<-DHGLMMODELING(Link="logit", LinPred=y2~dose+dose2+(1|litter),RandDist="gaussian")

Init_Corr=list(c(0))
SSC=list(as.factor(c(eg1$litter,eg1$litter)),as.factor(c(eg1$litter,eg1$litter)))
EstimateOverDisp=c(TRUE,FALSE)
LaplaceFixed=c(TRUE,TRUE)
ZZ1<-model.matrix(~as.factor(eg1$litter)-1)
ZZCorr=list(ZZ1,ZZ1)

#### independent random-effects model ####
res_ind<-jointfit(RespDist=c("gaussian","binomial"),DataMain=list(eg1,eg1),
    MeanModel=list(jm1,jm2),structure="correlated",
    Init_Corr=Init_Corr,EstimateCorrelations=FALSE,convergence=1,ZZCorr=ZZCorr)

#### correlated random-effects model ####
res_corr<-jointfit(RespDist=c("gaussian","binomial"),DataMain=list(eg1,eg1),
    MeanModel=list(jm1,jm2),structure="correlated",
    Init_Corr=Init_Corr,convergence=1,ZZCorr=ZZCorr)

#### shared random-effects model ####
Init_Corr=c(1,-10)
res_saturated<-jointfit(RespDist=c("gaussian","binomial"),DataMain=list(eg1,eg1),
    MeanModel=list(jm1,jm2),structure="shared",
    Init_Corr=Init_Corr,convergence=1)
```

---

kidney   *Kidney Infection Data*

Description

This data consist of times until the first and second recurrences of kidney infection in 38 patients using a portable dialysis machine (McGilchrist and Aisbett, 1991; Therneau and Lumley, 2015). Each survival time is the time until infection since the insertion of the catheter.

Usage

```
data("kidney")
```

Format

A data frame with 76 observations on the following 7 variables.

id patients identifier for 38 patients

time time until infection since the insertion of the catheter

`status` event stuatus (=1; infection, =0; censoring)

`age` age of patients in years

`sex` 1=male, 2=female

`disease` a factor for disease type with levels `Other`, `GN`, `AN` and `PKD`

`frail` frailty estimate from original paper

References

McGilchrist, C.A. and Aisbett, C.W. (1991). Regression with frailty in survival analysis. Biometrics, 47, 461–466.

Therneau, T. and Lumley, T. (2015). survival: Survival Analysis. R package version 2.38-3.

`lip`   *Scottish Lip Cancer Data*

Description

Clayton and Kaldor (1987) analyzed observed and expected numbers of lip cancer cases in the 56 administrative areas of Scotland with a view to produce a map that would display regional variations in cancer incidence and yet avoid the presentation of unstable rates for the smaller areas. The expected numbers had been calculated allowing for the different age distributions in the areas by using a fixed-effects multiplicative model; these were regarded for the purpose of analysis as (Intercept)s based on an external set of standard rates. Presumably the spatial aggregation is due in large part to the effects of environmental risk factors. Data were available on the percentage of the work force in each area employed in agriculture, fishing, or forestry. This covariate exhibits spatial aggregation paralleling that for lip cancer itself. Because all three occupations involve outdoor work, exposure to sunlight, the principal known risk factor for lip cancer, might be the explanation.

Usage

```
data("lip")
```

Format

A data frame with 56 observations on the following 4 variables.

`y` observed number of lip cancer

`n` expected number of lip cancer

`x` percentage of the work force in each area employed in agriculture, fishing, or forestry

`county` county number for 56 areas

References

Clayton, D.G. and Kaldor, J. (1987). Empirical bayes estimates of agestandardized relative risks for use in disease mapping. Biometrics, 43, 671–681.

Loaloa   *Loaloa Data*

Description

Rousset et al. (2016) presented the loaloa data set which describes prevalence of infection by the nematode Loa loa in North Cameroon, 1991- 2001. The study investigated the relationship between altitude, vegetation indices, and prevalence of the parasite.

Usage

```
data("Loaloa")
```

Format

A data frame with 197 observations on the following 11 variables.

longitude longitude, in degrees

latitude latitude, in degrees

ntot sample size per location

npos number of infected individuals per location

maxNDVI maximum normalised-difference vegetation index (NDVI) from repeated satellite scans

seNDVI standard error of NDVI

elev1 altitude, in m

elev2 additional altitude variables derived from the previous one, provided for convenience respectively, positive values of altitude-650,

elev3 positive values of altitude-1000

elev4 positive values of altitude-1300

maxNDVI1 a copy of maxNDVI modified as maxNDVI1 (maxNDVI1=0.8 if maxNDVI>0.8, maxNDVI1=maxNDVI otherwise)

References

Rousset, F., Ferdy, J.B. and Courtiol, A. (2016). spaMM: Mixed Models, Particularly Spatial GLMMs. R package version 1.7.2.

---

motherStress   *Mother's Stress and Children's Morbidity Data*

Description

Asar and Ilk (2014) considered the longitudinal data set from mother's stress and children's morbidity study (MSCM). In MSCM study, 167 mothers and their preschool children were enrolled for 28 days. Investigation of the serial dependence structures of the two longitudinal responses suggested a weak correlation structure for the period of days 1~16. Therefore, only the period of days 17~28 is considered in this data set.

Usage

```
data("motherStress")
```

Format

A data frame with 2004 observations on the following 14 variables.

id subjects identifier for 167 mothers

stress mother's stress (0=absence, 1=presence)

illness children's illness status (0=absence, 1=presence)

married mother's marriage status (0=other, 1=married)

education mother's highest education level (0=less than high school, 1=more than high school)

employed mother's employment status (0=unemployed, 1=employed)

chlth health status of children at baeline (0=very poor/poor, 1=fair, 2=good, 3=very good)

mhlth health status of mother at baeline (0=very poor/poor, 1=fair, 2=good, 3=very good)

race children's race (0=white, 1=non-white)

csex children's gender (0=male, 1=female)

housize household size (0=2-3 people, 1=more than 3 people)

bstress the average mother's stress of the 1~16 days

billness the average children's illnes of the 1~16 days

week study time (=(day-22)/7)

References

Asar, O. and Ilk, O. (2013). mmm: An R package for analyzing multivariate longitudinal data with multivariatemarginal models. Computer Methods and Programs in Biomedicine, 112, 649–654.

nmsqt	*National Merit Scholarship Qualifying Test Data for Twins*

Description

Loehlin and Nichols (1976) presented the national merit twins data including extensive questionnaires from 839 adolescent twins. The twins were identified among the roughly 600,000 US high school juniors who took the national merit scholarship qualifying test (NMSQT) in 1962. They were diagnosed as identical (509 pairs) or same-sex fraternal (330 pairs) by a brief mail questionnaire and later completed a 1082-item questionnaire covering a variety of behaviors, attitudes, personality, life experiences, health, vocational preferences, etc., plus the 480-item California psychological inventory. Twins' scores on the NMSQT and their five subscales are also included. The 285-item questionnaire filled out by the parent was mainly focused on the life histories and experiences of the twins.

Usage

```
data("nmsqt")
```

Format

A data frame with 1536 observations on the following 11 variables.

number data number for 1536 obervations

pairnum twin number for 768 pairs

x1 =0 (male), =1 (female)

x2 mother's educational level; =1 (less than or equal to 8th grade), =2 (part high school), =3 (high school grad), =4 (part college), =5 (college grad), =6 (graduate degree)

x3 father's educational level; =1 (less than or equal to 8th grade), =2 (part high school), =3 (high school grad), =4 (part college), =5 (college grad), =6 (graduate degree)

x4 family income level; =1 (less than equal to $5000), =2 ($5000 to $7499), =3 ($7500 to $9999), =4 ($10000 to $14999), =5 ($15000 to $19999), =6 ($20000 to $24999), =7 (more than $25000).

y1 NMSQT score for English (0-100)

y2 NMSQT score mathematics (0-100)

y3 NMSQT score social science (0-100)

y4 NMSQT score natural science (0-100)

x5 zygosity; =0 (identical), =1 (fraternal)

References

Loehlin, J.C. and Nichols, R.C. (1976). Heredity, Environments and Personality: A Study of 850 Twins. University of Texas Press, Austin.

Orthodont   *Orthodontic Growth Data*

Description

The orthodontic growth data (Pinheiro and Bates, 2000) contain the growth measurements of 27 children for 11 girls and 16 boys from age 8 until age 14. Every two years, the distance between the pituitary and the pterygomaxillary fissure was recorded at ages 8, 10, 12 and 14.

Usage

```
data("Orthodont")
```

Format

A data frame with 108 observations on the following 4 variables.

distance distance between the pituitary and the pterygomaxillary fissure

age age in year (8, 10, 12 and 14 for each subject)

Subject subjects indenfier for 11 girs and 16 boys

Sex Gender (Male and Female)

References

Pinheiro, J.C. and Bates, D. (2000). Mixed-Effects Models in S and S-PLUS. Springer, New York.

PBC910   *Primary Biliary Cirrhosis Data*

Description

Komarek (2015) presented a longitudinal data set from a Mayo Clinic trial on 312 patients with primary biliary cirrhosis (PBC) conducted in 1974-1984. There are 1 to 5 visits per subject performed at time of months. At each visit, measurements of three response variables are observed: continuous logarithmic bilirubin, discrete platelet count and dichotomous indication of blood vessel malformations.

Usage

```
data("PBC910")
```

Format

A data frame with 918 observations on the following 9 variables.

subject patients idenfier

day visiting time in day

month visiting time in month

lbili continuous logarithmic bilirubin

platelet discrete platelet count

spiders dichotomous indication of blood vessel malformation

References

Komarek, A. (2015). mixAK: Multivariate Normal Mixture Models and Mixtures of Generalized
Linear Mixed Models Including Model Based Clustering. R package version 4.2.

Prestige   *Prestige Data*

Description

Fox, et al. (2016) considered an additive non-parametric regression model with the prestige data.

Usage

```
data("Prestige")
```

Format

A data frame with 102 observations on the following 6 variables.

education average education of occupational incumbents, years, in 1971

income average income of incumbents, dollars, in 1971

women percentage of incumbents who are women

prestige Pineo-Porter prestige score for occupation, from a social survey conducted in the mid-
1960s

census Canadian Census occupational code

type a factor for type of occupation with levels bc (Blue Collar), prof (Professional, Managerial
and Technical) and wc (White Collar)

References

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., Firth, D., Friendly, M.,
Gorjanc, G., Graves, S., Heiberger, R., Monette, G., Murdoch, D., Nilsson, H., Ogle, D., Riply, B.,
Venables, W., Winsemius, D. and Zeileis, A. (2016). car: Companion to Applied Regression. R
package version 2.1-2.

---

rats	*Rat Data*

Description

The data set presented by Mantel, et al. (1977) is based on a tumorigenesis study of 50 litters of female rats. For each litter, one rat was selected to receive the drug and the other two rats were placebo-treated controls. Here each litter is treated as a cluster. The survival time is the time to development of tumor, measured in weeks. Death before occurrence of tumor yields a right-censored observation; forty rats developed a tumor, leading to censoring of about 73 percent.

Usage

```
data("rats")
```

Format

A data frame with 300 observations on the following 5 variables.

litter cluster number for 50 litters

rx tratment (1=drug, 0=placebo)

time time to development of tumor, measured in weeks

status status (1=occurrence of event, 0=censoring)

sex a factor with levels f (female) and m (male)

References

Mantel, N., Bohidar, N.R. and Ciminera, J.L. (1977). Mantel-haenszel analyses of litter-matched time-to-response data, with modifications for recovery of interlitter information. Cancer Research, 37, 3863–3868.

---

ra_data	*Rheumatoid Arthritis Data*

Description

The Rheumatoid Arthritis Patients rePort Onset Re-activation sTudy (RAPPORT study) is a longitudinal study that aims to identify an increase in disease activity by self-reported questionnaires. A cohort of 159 patients is followed throughout one year. Self-reported questionnaires are provided for patients every three months together with clinical evaluations of patients disease status. Health assessment questionnaires (HAQ) (20 questions from eight categories) and the Rheumatoid Arthritis Disease Activity Index (RADAI) (five items, e.g., todays disease activity in terms of swollen and tender joints) were used for patients to self-report their functional status. We use binarized versions of HAQ and RADAI, with cut-off points of 0.5 (HAQ) and 2.2 (RADAI). Lower values indicate disease stability, while higher values show increased disease activity. A clinical examination was

recorded using the disease activity score with 28 joint counts (DAS28), which is a composite score that includes for example the swollen joints count. The DAS28 score varies between 0 and 10, and we assume Gaussian model to approximate its distribution. Information about DAS28, RADAI, and HAQ taken at months 0, 3, 6, 9, and 12 were considered in this example. The analysis include gender and baseline age of the patients as covariates.

Usage

```
data("ra_data")
```

Format

A data frame with 2089 observations on the following 13 variables.

`id` subjects idenfier

`sex` a factor for gender with levels `M` and `V`

`age` age in year

`yearsRA` years of RA

`ethnic` a factor with levels `Dutch` and `Other`

`edu` a factor with levels `Master/Bachelor` and `Other than Master/Bachelor`

`X_NAME_` a factor with levels das0, das12, das3, das6, das9, haq0, haq12, haq3, haq6, haq9, radai0, radai12, radai3, radai6 and radai9

`X_LABEL_` a factor with levels `Das score (clinical) T=0`, `Das score (clinical) T=12`, `Das score (clinical) T=3`, `Das score (clinical) T=6`, `Das score (clinical) T=9`, `HAQ T=0`, `HAQ T=12`, `HAQ T=3`, `HAQ T=6`, `HAQ T=9`, `RADAI T=0`, `RADAI T=12`, `RADAI T=3`, `RADAI T=6` and `RADAI T=9`

`outcome` outcome responses

`outno` number for outcomes

`outbin` outbin

`outlabel` a factor with levels `remission` and `stable`

`time` time in month

---

respiratory   *Respiratory Data*

Description

This example is from a clinical trial comparing two treatments for a respiratory illness (Strokes et al., 1995). In each of two centers, eligible patients were randomly assigned to active treatment or placebo. During treatment, respiratory status was determined at four visits. Potential explanatory variables were centre, sex, and baseline respiratory status (all dichotomous), as well as age(in years) at the time of study entry. There were 111 patients (54 active, 57 placebo).

Usage

```
data("respiratory")
```

Format

A data frame with 444 observations on the following 8 variables.

patient patients identifier for 111 patients

treatment a factor for treatment with levels active and placebo

sex a factor for gender with levels female and male

age age in years

center center number for two centers

baseline baseline respiratory status (0 or 1)

past respiratory status in the previous visit

y respiratory status (0=poor, 1=good)

References

Strokes, M.E., Davis, C.S., and Koch, G.G. (1995). Categorical Data Analysis using the SAS System. SAS Institute, Cary, NC.

---

salamander   *Salamander Data*

Description

McCullagh and Nelder (1989) presented a data set on salamander mating. Three experiments were conducted: two were done with the same salamanders in the summer and autumn and another one in the autumn of the same year using different salamanders. The response variable is binary, indicating success of mating. In each experiment, 20 females and 20 males from two populations called whiteside, denoted by W, and rough butt, denoted by R, were paired six times for mating with individuals from their own and the other population, resulting in 120 observations in each experiment.

Usage

```
data("salamander")
```

Format

A data frame with 360 observations on the following 7 variables.

Season a factor with levels Fall and Summer

Experiment experiment number

TypeM a factor for male with levels R and W

TypeF a factor for female with levels R and W

Male males identifier for 20 males

Female females idenfier for females

Mate 1=success of mating, 0=failness of mating

Source

McCullagh, P. and Nelder, J.A. (1989). Generalized Linear Models. Chapman and Hall, London.

---

sch   *Schizophrenic Behaviour Data*

Description

Rubin and Wu (1997) analyzed schizophrenic behaviour data from an eye-tracking experiment with a visual target moving back and forth along a horizontal line on a screen. The outcome measurement is called the gain ratio, which is eye velocity divided by target velocity, and it is recorded repeatedly at the peak velocity of the target during eye-tracking under three conditions. The first condition is plain sine (PS), which means the target velocity is proportional to the sine of time and the colour of the target is white. The second condition is colour sine (CS), which means the target velocity is proportional to the sine of time, as for PS, but the colours keep changing from white to orange or blue. The third condition is triangular (TR), in which the target moves at a constant speed equal to the peak velocity of PS, back and forth, but the colour is always white. There are 43 nonschizophrenic subjects, 22 females and 21 males, and 43 schizophrenic subjects, 13 females and 30 males. In the experiment, each subject is exposed to five trials, usually three PS, one CS, and one TR. During each trial, there are 11 cycles, and a gain ratio is recorded for each cycle. However, for some cycles, the gain ratios are missing because of eye blinks, so that there are, on average, 34 observations out of 55 cycles for each subject. In this paper, we show how to analyze the data, assuming missing at random which means the probability of being missing does not depend on the values of missing data.s

Usage

```
data("sch")
```

Format

A data frame with 2906 observations on the following 7 variables.

y Gain Ratios for the each measurements of the subject

x1 Effect of PS versus CS (-0.5=CS, 0.5=PS, 0=TR)

x2 Effect of TR versus CS and PS (-1/3=CS, -1/3=PS, 2/3=TR)

sex Gender of subjects (1=male, 0=female)

time Linear trend coded -5, ..., 5 for 11 measurement times of each subject

schiz Indicator of schizophrenic subjects (=1) or not (=0)

v1 Subject number for 43 non-schizophrenic and 43 schizophrenic subjects

References

Rubin, D.B. and Wu, Y.N. (1997) Modeling schizophrenic behavior using general mixture components. Biometrics 53, 243–261.

---

scr	*Serum Creatinine data*

Description

In clinical trials, various response variables of interest are measured repeatedly over time on the same subject, which can be analysed by using the mdhglm package. At the same time, an event time representing recurrent or terminating time is also obtained. For example, consider a clinical study to investigate the chronic renal allograft dysfunction in renal transplants (Sung et al., 1998). The renal function was evaluated from the serum creatinine (sCr) values. Since the time interval between the consecutive measurements differs from patient to patient, we focus on the mean creatinine levels over six months. In addition, a single terminating survival time (graft-loss time), measured by month, is observed from each patient. During the study period, there were 13 graft losses due to the kidney dysfunction. For other remaining patients, we assumed that the censoring occurred at the last follow-up time. Thus, the censoring rate is about 88 percent.

Usage

```
data("scr")
```

Format

A data frame with 1395 observations on the following 9 variables.

id subjects identifier

month visiting time in months

cr serum creatinine value

sex gender (1=male, 0=female)

age age in years

icr 1/cr

sur_time graft loss time

status =1 for occurrence of event, 0=censoring

first indicator first visit or not (1=first visit, 0=oterwise)

References

Sung, K.H., Kang, K.W., Kang, C.M., Kwak, J.Y., Park, T.S. and Lee, S.Y. (1998). Study on the factors affecting the chronic renal allograft dysfunction. The Korean Journal of Nephrology, 17, 483–493.

---

semicon	*Semiconductor Data*

Description

This example is taken from Myers et al. (2002). It involves a designed experiment in a semicon-ductor plant. Six factors are employed, and it is of interest to study the curvature or camber of the substrate devices produced in the plant. There is a lamination process, and the camber measurement is made four times on each device produced. The goal is to model the camber taken in 10E-04 in. as a function of the design variables. Each design variable is taken at two levels and the design is a 2^(6-2) fractional factorial. The camber measurement is known to be nonnormal.

Usage

```
data("semicon")
```

Format

A data frame with 64 observations on the following 8 variables.

Device Devise number

x1 a x1 variable with two values -1 and 1

x2 a x2 variable with two values -1 and 1

x3 a x3 variable with two values -1 and 1

x4 a x4 variable with two values -1 and 1

x5 a x5 variable with two values -1 and 1

x6 a x6 variable with two values -1 and 1

y camber measurement

References

Myers, P.H., Montgomery, D.C., and Vining, G.G. (2002). Generalized Linear Models with Applications in Engineering and the Sciences. John Wiley & Sons, New York.

SmokeOnset	*Smoke Onset Data*

Description

For 1,556 students in the Los Angels area, onset of smoking are observed at each of three timepoints a1; a2 and a3 where a1 is starting time for investigation, a2 is the 1 year follow-up and a3 is the 2 year follow-up (Hedeker, et al., 2015). These event times are grouped at the three intervals [0, a1), [a1, a2) and [a2, a3).

Usage

```
data("SmokeOnset")
```

Format

A data frame with 1556 observations on the following 10 variables.

school schools identifier in the Los Angels area

class classrooms identifier

student students identifier

smkonset coded by i when the event occurs in the ith time interval

event =1 (smoked), =0 (otherwise)

int vector of ones representing intercept

SexMale gender of the student (0=Female, 1=Male)

cc indicating whether the school was randomized to a social-resistance classroom curriculum (1=Yes, 0=No)

tv indicating whether the school was randomized to a media (television) intervention (1=Yes, 0=No)

cctv indicating whether the school was randomized to CC combined with TV

References

Hedeker, D., Archer, K.J., Nordgren, R. and gibbons, R.D. (2015). mixor: Mixed-Effects Ordinal Regression Analysis. R package version 1.0.3.

---

snoring	*Snoring and Hear Diseae Data*

Description

Agresti (2007) presented the data on an epidemiological survey of 2,484 subjects to investigate snoring as a possible risk factor for heart disease. The subjects were classified according to their snoring level, as reported by their spouses.

Usage

```
data("snoring")
```

Format

A data frame with 4 observations on the following 3 variables.

yes number of subjects for occurrence of heart disease

no number of subjects for non-occurrence of heart diease

x snoring level (0, 2, 4, 5)

References

Agresti, A. (2007). An Introduction to Categorical Data Analysis. John Wiley & Sons, New York, 2nd edition.

---

stroke	*Stroke Patients Data*

Description

Approximate 30 percent of hospitalized patients due to acute ischemic stroke are placed in the risk of early neurologic deterioration (END) at their hospital stay. About the patients with acute ischemic stroke, real-time tracking of END would be actualized with monitoring the individual risk assumption with continuous blood pressure (BP). This data-set has systolic BP (SBP) observed repeatedly after arriving at emergency room for two stroke patients (one is END; the other is non-END).

Usage

```
data("stroke")
```

Format

A data frame with 72 observations on the following 3 variables.

time observed time

y1 SBP for END patient

y2 SBP for non-END patient

---

train	*Train Accident Data*

Description

Agresti (2007) presented the train accident data in the UK between 1975 and 2003. The respone variable y is annual number of collisions between trains and road vehicles, for t million kilometers of train travel. The covariate x is number of years since 1975.

Usage

```
data("train")
```

Format

A data frame with 29 observations on the following 4 variables.

x number of years since 1975

y annual number of collisions between trains and road vehicles

t million kilometers of annual train travel

id oberservations identifier

References

Agresti, A. (2007). An Introduction to Categorical Data Analysis. John Wiley & Sons, New York, 2nd edition.

Index