Package: mlmRev (via r-universe)

August 22, 2024

Version 1.0-8

Date 2020-04-04

Title Examples from Multilevel Modelling Software Review

Author Douglas Bates <bates@stat.wisc.edu>, Martin Maechler <maechler@R-project.org> and Ben Bolker <bolker@mcmaster.ca>

Contact LME4 Authors <lme4-authors@lists.r-forge.r-project.org>

Maintainer Steve Walker <steve.walker@utoronto.ca>

Description Data and examples from a multilevel modelling software review as well as other well-known data sets from the multilevel modelling literature.

Depends lme4, R (>= 2.10)

Suggests lattice

LazyData yes

License GPL (>= 2)

NeedsCompilation no

Date/Publication 2020-04-05 04:50:05 UTC

Repository https://stevencarlislewalker.r-universe.dev

RemoteUrl https://github.com/cran/mlmRev

RemoteRef HEAD

RemoteSha c8aa7c705e1fb8ad5d009e8891ba15b3dd96130b

Contents

---

bdf *Language Scores of 8-Graders in The Netherlands*

Description

The bdf data frame has 2287 rows and 25 columns of language scores from grade 8 pupils in elementary schools in The Netherlands.

Usage

```
data(bdf)
```

Format

schoolNR a factor denoting the school.

pupilNR a factor denoting the pupil.

IQ.verb a numeric vector of verbal IQ scores

IQ.perf a numeric vector of IQ scores.

sex Sex of the student.

Minority a factor indicating if the student is a member of a minority group.

repeatgr an ordered factor indicating if one or more grades have been repeated.

aritPRET a numeric vector

classNR a numeric vector

aritPOST a numeric vector

langPRET a numeric vector

langPOST a numeric vector

ses a numeric vector of socioeconomic status indicators.

denomina a factor indicating of the school is a public school, a Protestant private school, a Catholic private school, or a non-denominational private school.

schoolSES a numeric vector

satiprin a numeric vector

natitest a factor with levels `0` and `1`

meetings a numeric vector

currmeet a numeric vector

mixedgra a factor indicating if the class is a mixed-grade class.

percmino a numeric vector

aritdiff a numeric vector

homework a numeric vector

classsiz a numeric vector

groupsiz a numeric vector

References

Snijders, Tom and Bosker, Roel (1999) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, Sage.

Examples

```
summary(bdf)
```

`Chem97` *Scores on A-level Chemistry in 1997*

Description

Scores on the 1997 A-level Chemistry examination in Britain. Students are grouped into schools within local education authories. In addition some demographic and pre-test information is provided.

Usage

```
data(Chem97)
```

Format

A data frame with 31022 observations on the following 8 variables.

lea Local Education Authority - a factor

school School identifier - a factor

student Student identifier - a factor

score Point score on A-level Chemistry in 1997

gender Student's gender

age Age in month, centred at 222 months or 18.5 years

gcsescore Average GCSE score of individual.

gcsecnt Average GCSE score of individual, centered at mean.

Details

This data set is relatively large with 31,022 individuals in 2,280 schools. Note that while this is used, illustratively, to fit Normal response models, the distribution of the response is not well described by a Normal distribution.

Source

http://www.bristol.ac.uk/cmm/learning/mmsoftware/data-rev.html

References

Yang, M., Fielding, A. and Goldstein, H. (2002). Multilevel ordinal models for examination grades (submitted to *Statistical Modelling*).

Examples

```
str(Chem97)
summary(Chem97)
(fm1 <- lmer(score ~ (1|school) + (1|lea), Chem97))
(fm2 <- lmer(score ~ gcsecnt + (1|school) + (1|lea), Chem97))
```

Contraception *Contraceptive use in Bangladesh*

Description

These data on the use of contraception by women in urban and rural areas come from the 1988 Bangladesh Fertility Survey.

Usage

```
data(Contraception)
```

Format

A data frame with 1934 observations on the following 6 variables.

woman Identifying code for each woman - a factor

district Identifying code for each district - a factor

use Contraceptive use at time of survey

livch Number of living children at time of survey - an ordered factor. Levels are `0`, `1`, `2`, `3+`

age Age of woman at time of survey (in years), centred around mean.

urban Type of region of residence - a factor. Levels are `urban` and `rural`

Source

http://www.bristol.ac.uk/cmm/learning/mmsoftware/data-rev.html

References

Steele, F., Diamond, I. And Amin, S. (1996). Immunization uptake in rural Bangladesh: a multilevel analysis. *Journal of the Royal Statistical Society, Series A* (159): 289-299.

Examples

```
str(Contraception)
summary(Contraception)
(fm1 <- glmer(use ~ urban+age+livch+(1|district), Contraception, binomial))
(fm2 <- glmer(use ~ urban+age+livch+(urban|district), Contraception, binomial))
```

Early *Early childhood intervention study*

Description

Cognitive scores of infants in a study of early childhood intervention. The 103 infants from low income African American families were divided into a treatment group (58 infants) and a control group (45 infants). Starting at 0.5 years of age the infants in the treatment group were exposed to an enriched environment. Each infant's cognitive score on an age-specific, normalized scale was recorded at ages 1, 1.5, and 2 years.

Usage

```
data(Early)
```

Format

This `groupedData` object contains the following columns

id An ordered factor of the id number for each infant.

cog A numeric cognitive score.

age The age of the infant at the measurement.

trt A factor with two levels, `"N"` and `"Y"`, indicating if the infant is in the early childhood intervention program.

References

Singer, Judith D. and Willett, John B. (2003), *Applied Longitudinal Data Analysis*, Oxford University Press. (Ch. 3)

Examples

```
str(Early)
```

egsingle *US Sustaining Effects study*

Description

A subset of the mathematics scores from the U.S. Sustaining Effects Study. The subset consists of information on 1721 students from 60 schools

Usage

```
data(egsingle)
```

Format

A data frame with 7230 observations on the following 12 variables.

schoolid a factor of school identifiers

childid a factor of student identifiers

year a numeric vector indicating the year of the test

grade a numeric vector indicating the student's grade

math a numeric vector of test scores on the IRT scale score metric

retained a factor with levels `0` `1` indicating if the student has been retained in a grade.

female a factor with levels `Female` `Male` indicating the student's sex

black a factor with levels `0` `1` indicating if the student is Black

hispanic a factor with levels `0` `1` indicating if the student is Hispanic

size a numeric vector indicating the number of students enrolled in the school

lowinc a numeric vector giving the percentage of low-income students in the school

mobility a numeric vector

Source

These data are distributed with the HLM software package (Bryk, Raudenbush and Congdon, 1996). Conversion to the R format is described in Doran and Lockwood (2004).

References

Doran, Harold C. and Lockwood, J.R. (2004), *Fitting value-added models in R*, (submitted).

Examples

```
str(egsingle)
(fm1 <- lmer(math~year*size+female+(1|childid)+(1|schoolid), egsingle))
```

Exam *Exam scores from inner London*

Description

Exam scores of 4,059 students from 65 schools in Inner London.

Usage

```
data(Exam)
```

Format

A data frame with 4059 observations on the following 9 variables.

school School ID - a factor.

normexam Normalized exam score.

schgend School gender - a factor. Levels are `mixed`, `boys`, and `girls`.

schavg School average of intake score.

vr Student level Verbal Reasoning (VR) score band at intake - a factor. Levels are `bottom 25%`, `mid 50%`, and `top 25%`.

intake Band of student's intake score - a factor. Levels are `bottom 25%`, `mid 50%` and `top 25%`./

standLRT Standardised LR test score.

sex Sex of the student - levels are `F` and `M`.

type School type - levels are `Mxd` and `Sngl`.

student Student id (within school) - a factor

Source

http://www.bristol.ac.uk/cmm/learning/mmsoftware/data-rev.html

References

Goldstein, H., Rasbash, J., et al (1993). A multilevel analysis of school examination results. *Oxford Review of Education* 19: 425-433

Examples

```
str(Exam)
summary(Exam)
(fm1 <- lmer(normexam ~ standLRT + sex + schgend + (1|school), Exam))
(fm2 <- lmer(normexam ~ standLRT*sex + schgend + (1|school), Exam))
(fm3 <- lmer(normexam ~ standLRT*sex + schgend + (1|school), Exam))
```

Gcsemv *GCSE exam score*

Description

The GCSE exam scores on a science subject. Two components of the exam were chosen as outcome variables: written paper and course work. There are 1,905 students from 73 schools in England.

Usage

```
data(Gcsemv)
```

Format

A data frame with 1905 observations on the following 5 variables.

school School ID - a factor

student Student ID - a factor

gender Gender of student

written Total score on written paper

course Total score on coursework paper

Source

http://www.bristol.ac.uk/cmm/learning/mmsoftware/data-rev.html

References

Multivariate response models. (2000). In Rasbash, J., et al, *A user's guide to MLwiN*, Institute of Education, University of London.

Examples

```
str(Gcsemv)
```

guImmun *Immunization in Guatemala*

Description

Immunizations received by children in Guatemala.

Usage

```
data(guImmun)
```

Format

A data frame with 2159 observations on the following 13 variables.

kid a factor identifying the child

mom a factor identifying the family.

comm a factor identifying the community.

immun a factor indicating if the child received a complete set of immunizations. All children in this data frame received at least one immunization.

kid2p a factor indicating if the child was 2 years or older at the time of the survey.

mom25p a factor indicating if the mother was 25 years or older.

ord an factor indicating the child's birth's order within the family. Levels are `01` - first child, `23` - second or third child, `46` - fourth to sixth child, `7p` - seventh or later child.

ethn a factor indicating the mother's ethnicity. Levels are `L` - Ladino, `N` - indigenous not speaking Spanish, and `S` - indigenous speaking Spanish.

momEd a factor describing the mother's level of eduation. Levels are `N` - not finished primary, `P` - finished primary, `S` - finished secondary

husEd a factor describing the husband's level of education. Levels are the same as for `momEd` plus `U` - unknown.

momWork a factor indicating if the mother had ever worked outside the home.

rural a factor indicating if the family's location is considered rural or urban.

pcInd81 the percentage of indigenous population in the community at the 1981 census.

Source

These data are available at http://data.princeton.edu/multilevel/guImmun.dat. Multiple indicator columns in the original data table have been collapsed to factors for this data frame.

References

Rodriguez, Germán and Goldman, Noreen (1995), "Improved estimation procedures for multilevel models with binary response: a case-study", *Journal of the Royal Statistical Society, Series A*, **164**, 339-355.

Examples

```
data(guImmun)
summary(guImmun)
```

---

guPrenat *Prenatal care in Guatemala*

Description

Data on the prenatal care received by mothers in Guatemala.

Usage

```
data(guPrenat)
```

Format

A data frame with 2449 observations on the following 15 variables.

kid a factor identifying the birth

mom a factor identifying the mother or family

cluster a factor identifying the community

prenat a factor indicating if traditional or modern prenatal care was provided for the birth.

childAge an ordered factor of the child's age at the time of the survey.

motherAge a factor indicating if the mother was older or younger. The cut-off age is 25 years.

birthOrd an ordered factor for the birth's order within the family.

indig a factor indicating if the mother is Ladino, or indigenous not speaking Spanish, or indigenous speaking Spanish.

momEd a factor describing the mother's level of eduation.

husEd a factor describing the husband's level of education.

husEmpl a factor describing the husband's employment status.

toilet a factor indicating if there is a modern toilet in the house.

TV a factor indicating if there is a TV in the house and, if so, the frequency with which it is used.

pcInd81 the percentage of indigenous population in the community at the 1981 census.

ssDist distance from the community to the nearest clinic.

Source

These data are available at http://data.princeton.edu/multilevel/guPrenat.dat. Multiple indicator columns in the original data table have been collapsed to factors for this data frame.

References

Rodriguez, Germán and Goldman, Noreen (1995), "Improved estimation procedures for multilevel models with binary response: a case-study", *Journal of the Royal Statistical Society, Series A*, **164**, 339-355.

Examples

```
data(guPrenat)
summary(guPrenat)
```

Hsb82 *High School and Beyond - 1982*

Description

Data from the 1982 study “High School and Beyond”.

Usage

```
data(Hsb82)
```

Format

A data frame with 7185 observations on students including the following 8 variables.

school an ordered factor designating the school that the student attends.

minrty a factor with levels

sx a factor with levels `Male` and `Female`

ses a numeric vector of socio-economic scores

mAch a numeric vector of Mathematics achievement scores

meanses a numeric vector of mean `ses` for the school

sector a factor with levels `Public` and `Catholic`

cses a numeric vector of centered `ses` values where the centering is with respect to the `meanses` for the school.

Details

Each row in this data frame contains the data for one student.

References

Raudenbush, Stephen and Bryk, Anthony (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage (chapter 4).

Examples

```
data(Hsb82)
summary(Hsb82)
```

Mmmec *Malignant melanoma deaths in Europe*

Description

Malignant Melanoma Mortality in the European Community associated with the impact of UV radiation exposure.

Usage

```
data(Mmmec)
```

Format

A data frame with 354 observations on the following 6 variables.

nation a factor with levels `Belgium`, `W.Germany`, `Denmark`, `France`, `UK`, `Italy`, `Ireland`, `Luxembourg`, and `Netherlands`

region Region ID - a factor.

county County ID - a factor.

deaths Number of male deaths due to MM during 1971–1980

expected Number of expected deaths.

uvb Centered measure of the UVB dose reaching the earth's surface in each county.

Source

http://www.bristol.ac.uk/cmm/learning/mmsoftware/data-rev.html

References

Langford, I.H., Bentham, G. and McDonald, A. 1998: Multilevel modelling of geographically aggregated health data: a case study on malignant melanoma mortality and UV exposure in the European community. *Statistics in Medicine* 17: 41-58.

Examples

```
str(Mmmec)
summary(Mmmec)
(fm1 <- glmer(deaths ~ uvb + (1|region), Mmmec, poisson, offset = log(expected)))
```

Oxboys *Heights of Boys in Oxford*

Description

The Oxboys data frame has 234 rows and 4 columns.

Format

This data frame contains the following columns:

Subject an ordered factor giving a unique identifier for each boy in the experiment

age a numeric vector giving the standardized age (dimensionless)

height a numeric vector giving the height of the boy (cm)

Occasion an ordered factor - the result of converting age from a continuous variable to a count so these slightly unbalanced data can be analyzed as balanced.

Details

These data are described in Goldstein (1987) as data on the height of a selection of boys from Oxford, England versus a standardized age.

Source

Pinheiro, J. C. and Bates, D. M. (2000) *Mixed-Effects Models in S and S-PLUS*, Springer, New York. (Appendix A.19)

Examples

```
data(Oxboys)
```

s3bbx *Covariates in the Rodriguez and Goldman simulation*

Description

The s3bbx data frame has 2449 rows and 6 columns of the covariates in the simulation by Rodriguez and Goldman of multilevel dichotomous data.

Usage

```
data(s3bbx)
```

Format

This data frame contains the following columns:

child a numeric vector identifying the child

family a numeric vector identifying the family

community a numeric vector identifying the community

chldcov a numeric vector of the child-level covariate

famcov a numeric vector of the family-level covariate

commcov a numeric vector of the community-level covariate

Source

http://data.princeton.edu/multilevel/simul.htm

References

Rodriguez, Germán and Goldman, Noreen (1995) An assessment of estimation procedures for multilevel models with binary responses, *Journal of the Royal Statistical Society, Series A* **158**, 73–89.

Examples

```
str(s3bbx)
```

---

s3bby *Responses simulated by Rodriguez and Goldman*

Description

A matrix of the results of 100 simulations of dichotomous multilevel data. The rows correspond to the 2449 births for which the covariates are given in s3bbx. The elements of the matrix are all 0, indicating no modern prenatal care, or 1, indicating model prenatal care. These were simulated with "large" variances for both the family and the community random effects.

Usage

```
data(s3bby)
```

Format

An integer matrix with 2449 rows and 100 columns.

Source

http://data.princeton.edu/multilevel/simul.htm

References

Rodriguez, Germán and Goldman, Noreen (1995) An assessment of estimation procedures for multilevel models with binary responses, *Journal of the Royal Statistical Society, Series A* **158**, 73–89.

Examples

```
str(s3bby)
```

`ScotsSec` *Scottish secondary school scores*

Description

Scores attained by 3435 Scottish secondary school students on a standardized test taken at age 16. Both the primary school and the secondary school that the student attended have been recorded.

Usage

```
data(ScotsSec)
```

Format

A data frame with 3435 observations on the following 6 variables.

verbal The verbal reasoning score on a test taken by the students on entry to secondary school.

attain The score attained on the standardized test taken at age 16.

primary A factor indicating the primary school that the student attended.

sex A factor with levels `M` and `F`

social The student's social class on a numeric scale from low to high social class.

second A factor indicating the secondary school that the student attended.

Details

These data are an example of cross-classified grouping factors.

Source

http://www.bristol.ac.uk/cmm/learning/mmsoftware/data-rev.html

References

Paterson, L. (1991). Socio economic status and educational attainment: a multidimensional and multilevel study. *Evaluation and Research in Education* **5**: 97-121.

Examples

```
str(ScotsSec)
```

Socatt *Social Attitudes Survey*

Description

These data come from the British Social Attitudes (BSA) Survey started in 1983. The eligible persons were all adults aged 18 or over living in private households in Britain. The data consist of completed results of 264 respondents out of 410.

Usage

```
data(Socatt)
```

Format

A data frame with 1056 observations on the following 9 variables.

district District ID - a factor

respond Respondent code (within district) - a factor

year A factor with levels `1983`, `1984`, `1985`, and `1986`

numpos An ordered factor giving the number of positive answers to seven questions.

party Political party chosen - a factor. Levels are `conservative`, `labour`, `Lib/SDP/Alliance`, `others`, and `none`.

class Self assessed social class - a factor. Levels are `middle`, `upper working`, and `lower working`.

gender Respondent's sex. (1=male, 2=female)

age Age in years

religion Religion - a factor. Levels are `Roman Catholic`, `Protestant/Church of England`, `others`, and `none`.

Details

These data are provided as an example of multilevel data with a multinomial response.

Source

http://www.bristol.ac.uk/cmm/learning/mmsoftware/data-rev.html

References

McGrath, K. and Waterton, J. (1986). *British Social Attitudes 1983-1986 panel survey.* London, Social and Community Planning Research.

Examples

```
str(Socatt)
summary(Socatt)
```

star *Student Teacher Achievement Ratio (STAR) project data*

Description

Data from Tennessee's Student Teacher Achievement Ratio (STAR) project which was a large-scale, four-year study of the effect of reduced class size.

Usage

```
data(star)
```

Format

A data frame with 26796 observations on the following 18 variables.

`id` a factor - student id number

`sch` a factor - school id number

`gr` grade - an ordered factor with levels `K` < `1` < `2` < `3`

`cltype` class type - a factor with levels `small`, `reg` and `reg+A`. The last level indicates a regular class size with a teachers aide.

`hdeg` highest degree obtained by the teacher - an ordered factor with levels `ASSOC` < `BS/BA` < `MS/MA/MEd` < `MA+` < `Ed.S` < `Ed.D/Ph.D`

`clad` career ladder position of the teacher - a factor with levels `NOT APPR PROB PEND 1 2 3`

`exp` a numeric vector - the total number of years of experience of the teacher

`trace` teacher's race - a factor with levels `W`, `B`, `A`, `H`, `I` and `O` representing white, black, Asian, Hispanic, Indian (Native American) and other

`read` the student's total reading scaled score

`math` the student's total math scaled score

`ses` socioeconomic status - a factor with levels `F` and `N` representing eligible for free lunches or not eligible

`schtype` school type - a factor with levels `inner`, `suburb`, `rural` and `urban`

`sx` student's sex - a factor with levels `M F`

`eth` student's ethnicity - a factor with the same levels as `trace`

`birthq` student's birth quarter - an ordered factor with levels `1977:1` < ... < `1982:2`

`birthy` student's birth year - an ordered factor with levels `1977:1982`

`yrs` number of years of schooling for the student - a numeric version of the grade `gr` with Kindergarten represented as 0. This variable was generated from `gr` and does not allow for a student being retained.

`tch` a factor - teacher id number

Details

Details of the original data source and the process of conversion to this representation are given in the vignette.

Source

http://www.heros-inc.org/data.htm

Examples

```
str(star)
```

Index