Package 'MLMusingR'

May 10, 2022

Type Package

Title Practical Multilevel Modeling

Version 0.2.0

Date 2022-05-10

Maintainer Francis Huang <flhuang2000@yahoo.com>

Description Convenience functions and datasets to be used with Practical Multilevel Modeling using R. The package includes functions for calculating group means, group mean centered variables, and displaying some basic missing data information. A function for computing robust standard errors for linear mixed models based on Liang and Zeger (1986) <doi:10.1093/biomet/73.1.13> and Bell and 'McCaffrey' (2002) <https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2002002/article/9058-eng.pdf?st=NxMjN1YZ> is included as well as a function for checking for level-one homoskedasticity (Raudenbush & Bryk, 2002, ISBN:076191904X).

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.1.2

URL https://github.com/flh3/MLMusingR

BugReports https://github.com/flh3/MLMusingR/issues

Depends R (>= 3.5)

Imports lme4, stats, nlme, Matrix, methods

NeedsCompilation no

Author Francis Huang [aut, cre]

Repository CRAN

Date/Publication 2022-05-10 13:40:02 UTC

R **topics documented:**

---

cdata.ex	*Clustered dataset for centering example*

Description

Dataset of 60 observations from 3 clusters.

Usage

```
cdata.ex
```

Format

A wide data frame of 60 observations. Used for discussing within and between group effects.

x The predictor.

y The outcome of interest.

---

engage   *Student engagement dataset (complete data).*

Description

Example data used to investigate missing data (this is the complete dataset).

Usage

```
data(engage)
```

Format

A data frame with 528 observations from 40 groups and 7 variables:

eng Student engagement.

mot Student motivation.

gpa Student grade point average.

grade Student grade level (6-8; a factor).

rural School level rural variable indicator; 1 = yes/0 = no.

frpm Percent of students eligible for free or reduced price meals at the school.

school School indicator (clustering variable).

---

engage.miss   *Student engagement dataset (with missing data).*

Description

Example data used to investigate missing data (this has missing data).

Usage

```
data(engage.miss)
```

Format

A data frame with 528 observations from 40 groups and 7 variables:

eng Student engagement.

mot Student motivation.

gpa Student grade point average.

grade Student grade level (6-8; a factor).

rural School level rural variable indicator; 1 = yes/0 = no.

frpm Percent of students eligible for free or reduced price meals at the school.

school School indicator (clustering variable).

---

group_center   *Group-mean center a variable*

Description

Also referred to as centering within cluster (or within context) or demeaning the variable. By default, uses na.rm = TRUE when computing group means.

Usage

```
group_center(x, grp)
```

Arguments

x	Variable to center (e.g., dataframe$varname).
grp	Cluster/grouping variable (e.g., dataframe$cluster).

Value

A vector of group-mean centered variables.

Examples

```
data(mtcars)
#create a group centered variable
mtcars$mpg.gpc <- group_center(mtcars$mpg, mtcars$cyl)
```

---

group_mean   *Computes the group mean of a variable*

Description

Computes the group means of a variable by a specified cluster/group. Can also be used with factors that have two levels.

Usage

```
group_mean(x, grp, lm = FALSE)
```

Arguments

x	Variable to compute the mean for (e.g., dataframe$varname).
grp	Cluster/grouping variable (e.g., dataframe$cluster).
lm	Compute reliability (lambda) adjusted means.

Value

Outputs a vector of group means.

Examples

```
data(mtcars)
#create a group mean aggregated variable
mtcars$mpg.barj <- group_mean(mtcars$mpg, mtcars$cyl)
```

---

hdp   *Hospital, doctor, patient (hdp) dataset*

Description

This dataset has a three-level, hierarchical structure with patients nested within doctors within hospitals. The simulation code can be found at <https://stats.idre.ucla.edu/r/codefragments/mesimulation/#setup>.

Usage

```
data(hdp)
```

Format

A data frame with 8,525 rows and 17 variables:

Age Continuous in years but recorded at a higher degree of accuracy.

Married Binary, married/living with partner or single.

FamilyHx Binary (yes/no), does the patient have a family history (Hx) of cancer?

SmokingHx Categorical with three levels, current smoker, former smoker, never smoked.

Sex Binary (female/male).

CancerStage Categorical with four levels, stages 1-4.

LengthofStay Count number of days patients stayed in the hospital after surgery.

WBC Continuous, white blood count. Roughly 3,000 is low, 10,000 is middle, and 30,000 per microliter is high.

RBC Continuous, red blood count.

BMI Body mass index given by the formula (frackgmeters^2.

IL6 Continuous, interleukin 6, a proinflammatory cytokine commonly examined as an indicator of inflammation, cannot be lower than zero.

CRP Continuous, C-reactive protein, a protein in the blood also used as an indicator of inflammation. It is also impacted by BMI.

HID Hospital identifier.

DID Doctor identifier

Experience Years as a doctor.

School Whether the school doctor trained at was high quality or not.

remission Cancer in remission? 1 = yes, 0 = no.

Source

<https://stats.oarc.ucla.edu/r/codefragments/mesimulation/>

Htest	*Test for homoskedasticity at level one*

Description

Based on Raudenbush and Bryk (2002) and Hoffman (2007). A statistically significant Chisq indicates heteroskedasticity. Output shows the H statistic, degrees of freedom, and p value.

Usage

```
Htest(newdata, fml, group)
```

Arguments

newdata	data to be used.
fml	level 1 formula.
group	grouping variable (in quotes).

Value

Returns a data frame which contains:

H	The computed H statistic.
df	The degrees of freedom.
p	The p-value (< .05 indicates heteroskedasticity is present).

References

Hoffman, L. (2007). Multilevel models for examining individual differences in within-person variation and covariation over time. *Multivariate Behavioral Research, 42*(4), 609–629.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd ed.)*. Sage.

Examples

```
set.seed(123)
x1 <- rnorm(400)
y <- x1 * .3 + rnorm(400)
gr <- rep(1:20, each = 20)
dat <- data.frame(x1, y, gr)
Htest(dat, y ~ x1, 'gr') #no violation
y <- x1 * .3 + rnorm(400, 0, sqrt(x1^2)) #add violation
dat <- data.frame(x1, y, gr)
Htest(dat, y ~ x1, 'gr')
```

MatSqrtInverse   *Compute the inverse square root of a matrix*

Description

From Imbens and Kolesar (2016).

Usage

```
MatSqrtInverse(A)
```

Arguments

A   The matrix object.

---

nmiss   *Amount of missing data per variable*

Description

Amount of missing data per variable

Usage

```
nmiss(dat)
```

Arguments

dat   Data frame that you want to inspect.

Value

By default, this function will print the following items to the console

- The percent of missing data per variable.
- The percent of complete cases (range: 0 to 1).
- Suggested number of datasets to impute when using multiple imputation.

Examples

```
data(mtcars)
mtcars[c(2:3), 4] <- NA #create NAs
nmiss(mtcars)
```

---

ri_test1   *Sample dataset 1 for testing the likelihood ratio test*

Description

Example data for testing the need for a random intercept. Illustrates the need to adjust the p values for a modified LRT.

Usage

```
data(ri_test1)
```

Format

A data frame with 900 observations from 30 groups and 4 variables:

y The outcome variable.

w1 A level-2 predictor.

x1 A level-1 predictor

group The cluster identifier

---

ri_test2   *Sample dataset 2 for testing the likelihood ratio test (LRT)*

Description

Example data for testing the need for a random intercept. LRT results show that a random slope is not warranted.

Usage

```
data(ri_test2)
```

Format

A data frame with 3,000 observations from 30 groups and 4 variables:

y The outcome variable.

w1 A level-2 predictor.

x1 A level-1 predictor

group The cluster identifier

---

robust_mixed	*Cluster robust standard errors with degrees of freedom adjustments for lmerMod/lme objects*

Description

Function to compute the CR2/CR0 cluster robust standard errors (SE) with Bell and McCaffrey (2002) degrees of freedom (dof) adjustments. Suitable even with a low number of clusters. The model based (mb) and cluster robust standard errors are shown for comparison purposes.

Usage

```
robust_mixed(m1, digits = 3, type = "CR2", satt = TRUE, Gname = NULL)
```

Arguments

m1	The `lmerMod` or `lme` model object.
digits	Number of decimal places to display.
type	Type of cluster robust standard error to use ("CR2" or "CR0").
satt	If Satterthwaite degrees of freedom are to be computed (if not, between-within df are used).
Gname	Group/cluster name if more than two levels of clustering (does not work with lme).

Value

A data frame (`results`) with the cluster robust adjustments with p-values.

Estimate	The regression coefficient.
mb.se	The model-based (regular, unadjusted) SE.
cr.se	The cluster robust standard error.
df	degrees of freedom: Satterthwaite or between-within.
p.val	p-value using CR0/CR2 standard error.
stars	stars showing statistical significance.

Author(s)

Francis Huang, <huangf@missouri.edu>

Bixi Zhang, <bixizhang@missouri.edu>

References

Bell, R., & McCaffrey, D. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. Survey Methodology, 28, 169-182. *(link)*

Liang, K.Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. Biometrika, 73*(1), 13-22.* *(link)*

Examples

```
require(lme4)
data(sch29, package = 'MLMusingR')
robust_mixed(lmer(math ~ male + minority + mses + mhmwk + (1|schid), data = sch29))
```

---

satdf   *Compute Satterthwaite degrees of freedom*

Description

Function to compute empirical degrees of freedom based on Bell and McCaffrey (2002).

Usage

```
satdf(m1, type = "none", Vinv2, Vm2, br2, Gname = NULL)
```

Arguments

m1	The lmerMod or lme model object.
type	The type of cluster robust correction used (i.e., CR2 or none).
Vinv2	Inverse of the variance matrix.
Vm2	The variance matrix.
br2	The bread component.
Gname	The group (clustering variable) name'

Author(s)

Francis Huang, <huangf@missouri.edu>

Bixi Zhang, <bixizhang@missouri.edu>

---

sch29   *Data from 29 schools (based on the NELS dataset) used for regression diagnostics*

Description

For examining the association between amount homework done per week and math outcome.

Usage

```
data(sch29)
```

Format

A data frame with 648 rows and 8 variables:

schid The school identifier (the grouping variable)

ses Student-level socioeconomic status

byhomewk Total amount of time the student spent on homework per week. 1 = None, 2 = Less than one hour, 3 = 1 hour, 4 = 2 hours, 5 = 3 hours, 6 = 4-6 hours, 7 = 7 - 9 hours, 8 = 10 or more

math Mathematics score.

male Dummy coded gender, 1 = male, 0 = female

minority Dummy coded minority status, 1 = yes, 0 = no

mses Aggregated socioeconomic status at the school level

mhmwk Aggregated time spent on homework at the school level

Source

<https://nces.ed.gov/pubs92/92030.pdf>

---

suspend	*Suspension data from Virginia*

Description

Data from 8465 students from 100 schools in Virginia

Usage

```
data(suspend)
```

Format

Dataset:

school School identifier

pminor Percent minority enrollment at school

male 1 = male, 0 = female

sus Whether the student was suspended (1 = yes) in the school year or not (0 = no). Self reported.

fight If the student got into one or more fights (1 = yes) in the school year

gpa Students self-reported GPA; 1 = D to 4 = A

`thai`   *Thai data from PISA*

Description

Example data to be used for centering

Usage

```
data(thai)
```

Format

A data frame with 6606 rows and 18 variables:

pv1math First plausible value in mathematics.

escs Index of economic, social, and cultural status.

hisei Highest parent occupational status.

sex Student gender. 1 = Female, 2 = Male.

intmat Mathematics interest.

matheff Mathematics self-efficacy.

schoolid School identifier

othl Spoke another language at home other than Thai. 1 = yes, 0 = no.

books How many books at home.

pared Highest parental education in years.

w_fstuwt Student weight.

pv1read Plausible value #1 for reading.

pv2read Plausible value #2 for reading.

pv3read Plausible value #3 for reading.

pv4read Plausible value #4 for reading.

pv5read Plausible value #5 for reading.

private Private school. 1 = yes, 0 = no.

schsize Total school enrolment.

Source

https://gpseducation.oecd.org/CountryProfile?primaryCountry=THA

---

thai.sm	*Thai data from PISA (reduced)*

Description

Example data to be used for centering

Usage

```
data(thai)
```

Format

A data frame with 4271 rows and 7 variables:

math First plausible value in mathematics.

escs Index of economic, social, and cultural status.

intmat Mathematics interest.

schoolid School identifier

othl Spoke another language at home other than Thai. 1 = yes, 0 = no.

private Private school. 1 = yes, 0 = no.

schsize Total school enrolment.

Source

<https://gpseducation.oecd.org/CountryProfile?primaryCountry=THA>

---

wide	*Wide dataset to be used for growth modeling*

Description

A dataset containing 30 observations with reading scores taken in the fall kindergarten, spring kindergarten, and spring first grade

Usage

```
wide
```

Format

A wide data frame of 30 observations:

studentid Factor indicating student identification

int treatment or control

female 1 = female, 0 = male

fall_k Reading scores in fall kindergarten

spring_k Reading scores in spring kindergarten

spring_g1 Reading scores in spring first grade

Index