Package ‘pbkrtest’

March 15, 2017

Version 0.4-7
Title Parametric Bootstrap and Kenward Roger Based Methods for Mixed Model Comparison
Author Ulrich Halekoh <uhalekoh@health.sdu.dk> Søren Højsgaard
<sorenh@math.aau.dk>
Maintainer Søren Højsgaard <sorenh@math.aau.dk>
Description Test in mixed effects models. Attention is on mixed effects models as implemented in the 'lme4' package. This package implements a parametric bootstrap test and a Kenward Roger modification of F-tests for linear mixed effects models and a parametric bootstrap test for generalized linear mixed models.
URL http://people.math.aau.dk/~sorenh/software/pbktest/
Depends R (>= 3.2.3), lme4 (>= 1.1.10)
Imports Matrix (>= 1.2.3), parallel, MASS, methods
Encoding UTF-8
ZipData no
License GPL (>= 2)
RoxygenNote 5.0.1
NeedsCompilation no
Repository CRAN
Date/Publication 2017-03-15 08:49:25

R topics documented:

data-beets ................................................................. 2
data-budworm ............................................................. 3
getkr ........................................................................... 4
get_ddf_Lb ................................................................... 5
kr-modcomp ................................................................. 7
kr-vcov ................................................................. 8
Description

Yield and sugar percentage in sugar beets from a split plot experiment. Data is obtained from a split plot experiment. There are 3 blocks and in each of these the harvest time defines the "whole plot" and the sowing time defines the "split plot". Each plot was $25m^2$ and the yield is recorded in kg. See 'details' for the experimental layout.

Usage

beets

Format

The format is: chr "beets"

Details

Experimental plan

Sowing times
1  4. april
2  12. april
3  21. april
4  29. april
5  18. may

Harvest times
1  2. october
2  21. october

Plot allocation:

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plot</td>
<td>1 1 1 1 1</td>
<td>2 2 2 2 2</td>
</tr>
<tr>
<td>1-15</td>
<td>3 4 5 2 1</td>
<td>3 2 4 5 1</td>
</tr>
<tr>
<td>Harvest time</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plot</td>
<td>2 2 2 2 2</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>16-30</td>
<td>2 1 5 4 3</td>
<td>4 1 3 2 5</td>
</tr>
<tr>
<td>Sowing time</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References

data-budworm

Examples

data(beets)

beets$bh <- with(beets, interaction(block, harvest))
summary(aov(yield ~ block + sow + harvest + Error(bh), beets))
summary(aov(sugpct ~ block + sow + harvest + Error(bh), beets))

---

data-budworm  budworm data

Description

Effect of Insecticide on survival of tobacco budworms number of killed budworms exposed to an insecticide pp mortality of the moth tobacco budworm 'Heliothis virescens' for 6 doses of the pyrethroid trans-cypermethrin differentiated with respect to sex

Usage

budworm

Format

This data frame contains 12 rows and 4 columns:

sex: sex of the budworm
dose: dose of the insecticide trans-cypermethrin in [µg]
dead: budworms killed in a trial
ntotal: total number of budworms exposed per trial

Source


References


Examples

data(budworm)

## function to calculate the empirical logits
empirical.logit<- function(nevent,ntotal) {
  y <- log((nevent + 0.5) / (ntotal - nevent + 0.5))
  y
getkr

# plot the empirical logits against log-dose

log.dose <- log(budworm$dose)
emp.logit <- empirical.logit(budworm$ndead, budworm$ntotal)
plot(log.dose, emp.logit, type='n', xlab='log-dose', ylab='empirical logit')
title('budworm: empirical logits of probability to die')

male <- budworm$sex=='male'
female <- budworm$sex=='female'
lines(log.dose[male], emp.logit[male], type='b', lty=1, col=1)
lines(log.dose[female], emp.logit[female], type='b', lty=2, col=2)
legend(0.5, 2, legend=c('male', 'female'), lty=c(1,2), col=c(1,2))

## Not run:
* SAS example;
data budworm;
infile 'budworm.txt' firstobs=2;
input sex dose ndead ntotal;
run;
## End(Not run)

---

getkr

Extract (or "get") components from a KRmodcomp object.

Description

Extract (or "get") components from a KRmodcomp object, which is the result of the KRmodcomp function.

Usage


Arguments

object A KRmodcomp object, which is the result of the KRmodcomp function
name The available slots. If name is missing or NULL then everything is returned.

Author(s)

Søren Højsgaard <sorenh@math.aau.dk>
get_ddf_Lb

References

See Also
KRmodcomp, PBmodcomp, vcovAdj

Examples

data(beets, package='pbkrtest')
lg <- lmer(sugpct ~ block + sow + harvest + (1|block:harvest), data=beets, REML=FALSE)
sm <- update(lg, .~. - harvest)
xx <- KRmodcomp(lg, sm)
getKR(xx, "ddf") # get denominator degrees of freedom.

---

get_ddf_Lb

Adjusted denominator degrees freedom for linear estimate for linear mixed model.

Description
Get adjusted denominator degrees freedom for testing Lb=0 in a linear mixed model where L is a restriction matrix.

Usage
get_Lb_ddf(object, L)

## S3 method for class 'lmerMod'
get_Lb_ddf(object, L)

Lb_ddf(L, V0, Vadj)

get_ddf_Lb(object, Lcoef)

## S3 method for class 'lmerMod'
get_ddf_Lb(object, Lcoef)

ddf_Lb(VVa, Lcoef, VV0 = VVa)
Arguments

object  A linear mixed model object.
L       A vector with the same length as fixef(object) or a matrix with the same number of columns as the length of fixef(object)
V0, Vadj Unadjusted and adjusted covariance matrix for the fixed effects parameters. Unadjusted covariance matrix is obtained with vcov() and adjusted with vcovAdj().
Lcoef   Linear contrast matrix
VVa     Adjusted covariance matrix
VV0     Unadjusted covariance matrix

Value

Adjusted degrees of freedom (adjustment made by a Kenward-Roger approximation).

Author(s)

Søren Højsgaard, <sorenh@math.au.dk>

References


See Also

KRmodcomp, vcovAdj, model2restrictionMatrix, restrictionMatrix2model

Examples

(fmLarge <- lmer(Reaction ~ Days + (Days|Subject), sleepstudy))
## removing Days
(fmSmall <- lmer(Reaction ~ 1 + (Days|Subject), sleepstudy))
anova(fmLarge, fmSmall)
KRmodcomp(fmLarge, fmSmall) ## 17 denominator df's
get_Lb_ddf(fmLarge, c(0,1)) ## 17 denominator df's

# Notice: The restriction matrix L corresponding to the test above
# can be found with
L <- model2restrictionMatrix(fmLarge, fmSmall)
L
Description

An approximate F-test based on the Kenward-Roger approach.

Usage

```r
KRmodcomp(largeModel, smallModel, betaH = 0, details = 0)
```

```r
## S3 method for class 'lmerMod'
KRmodcomp(largeModel, smallModel, betaH = 0, details = 0)

## S3 method for class 'mer'
KRmodcomp(largeModel, smallModel, betaH = 0, details = 0)
```

Arguments

- `largeModel`: An `lmer` model
- `smallModel`: An `lmer` model or a restriction matrix
- `betaH`: A number or a vector of the beta of the hypothesis, e.g. `L beta=L betaH`. `betaH=0` if `modelSmall` is a model not a restriction matrix.
- `details`: If larger than 0 some timing details are printed.
- `...`: Additional arguments to print function

Details

The model object must be fitted with restricted maximum likelihood (i.e. with `REML=TRUE`). If the object is fitted with maximum likelihood (i.e. with `REML=FALSE`) then the model is refitted with `REML=TRUE` before the p-values are calculated. Put differently, the user needs not worry about this issue.

An F test is calculated according to the approach of Kenward and Roger (1997). The function works for linear mixed models fitted with the `lmer` function of the `lme4` package. Only models where the covariance structure is a sum of known matrices can be compared.

The `largeModel` may be a model fitted with `lmer` either using `REML=TRUE` or `REML=FALSE`. The `smallModel` can be a model fitted with `lmer`. It must have the same covariance structure as `largeModel`. Furthermore, its linear space of expectation must be a subspace of the space for `largeModel`. The model `smallModel` can also be a restriction matrix `L` specifying the hypothesis $L\beta = L\beta_H$, where $L$ is a $k \times p$ matrix and $\beta$ is a $p$ column vector the same length as `fixef(largeModel)`. The $\beta_H$ is a $p$ column vector.

Notice: if you want to test a hypothesis $L\beta = c$ with a $k$ vector $c$, a suitable $\beta_H$ is obtained via $\beta_H = Lc$ where $L_n$ is a g-inverse of $L$.

Notice: It cannot be guaranteed that the results agree with other implementations of the Kenward-Roger approach!
Note
This functionality is not thoroughly tested and should be used with care. Please do report bugs etc.

Author(s)
Ulrich Halekoh <ulrich.halekoh@agrsci.dk>, Søren Højsgaard <sorenh@math.aau.dk>

References

See Also
getKR, lmer, vcovAdj, PBmodcomp

Examples

(fmLarge <- lmer(Reaction ~ Days + (Days|Subject), sleepstudy))
## removing Days
(fmSmall <- lmer(Reaction ~ 1 + (Days|Subject), sleepstudy))
anova(fmLarge,fmSmall)
KRmodcomp(fmLarge,fmSmall)

## The same test using a restriction matrix
L <- cbind(0,1)
KRmodcomp(fmLarge, L)

## Same example, but with independent intercept and slope effects:
m.large <- lmer(Reaction ~ Days + (1|Subject) + (0+Days|Subject), data = sleepstudy)
m.small <- lmer(Reaction ~ 1 + (1|Subject) + (0+Days|Subject), data = sleepstudy)
anova(m.large, m.small)
KRmodcomp(m.large, m.small)

---

kr-vcov

**Adjusted covariance matrix for linear mixed models according to Kenward and Roger**

Description

Kenward and Roger (1997) desribbe an improved small sample approximation to the covariance matrix estimate of the fixed parameters in a linear mixed model.
Usage

vcovAdj(object, details = 0)

## S3 method for class 'lmerMod'
vcovAdj(object, details = 0)

Arguments

object

An lmer model

details

If larger than 0 some timing details are printed.

Value

phiA

the estimated covariance matrix, this has attributed P, a list of matrices used in KR_adjust and the estimated matrix W of the variances of the covariance parameters of the random effects

SigmaG

list: Sigma: the covariance matrix of Y; G: the G matrices that sum up to Sigma; n.ggamma: the number (called M in the article) of G matrices

Note

If N is the number of observations, then the vcovAdj() function involves inversion of an N x N matrix, so the computations can be relatively slow.

Author(s)

Ulrich Halekoh <uhalekoh@health.sdu.dk>, Søren Højsgaard <sorenh@math.aau.dk>

References


See Also

getKR, KRmodcomp, lmer, PBmodcomp, vcovAdj

Examples

fm1 <- lmer(Reaction ~ Days + (Days|Subject), sleepstudy)

## Here the adjusted and unadjusted covariance matrices are identical,
## but that is not generally the case
v1 <- vcov(fm1)
v2 <- vcovAdj(fm1,detail=0)
model-coerce

Conversion between a model object and a restriction matrix

Description

Testing a small model under a large model corresponds imposing restrictions on the model matrix of the larger model and these restrictions come in the form of a restriction matrix. These functions converts a model to a restriction matrix and vice versa.

Usage

model2restrictionMatrix(largeModel, smallModel)

## S3 method for class 'merMod'
model2restrictionMatrix(largeModel, smallModel)

## S3 method for class 'lm'
model2restrictionMatrix(largeModel, smallModel)

restrictionMatrix2model(largeModel, LL)

## S3 method for class 'merMod'
restrictionMatrix2model(largeModel, LL)

## S3 method for class 'lm'
restrictionMatrix2model(largeModel, LL)
Arguments
largeModel, smallModel
   Model objects of the same "type". Possible types are linear mixed effects models and linear models (including generalized linear models)

LL
   A restriction matrix.

Value

Note
That these functions are visible is a recent addition; minor changes may occur.

Author(s)
Ulrich Halekoh <uhalekoh@health.sdu.dk>, Søren Højsgaard <sorenh@math.aau.dk>

References

See Also
PBmodcomp, PBrefdist, KRmodcomp

Examples

library(pbkrtest)
data("beets", package = "pbkrtest")
sug <- lm(sugpct ~ block + sow + harvest, data=beets)
sug.h <- update(sug, .~- harvest)
sug.s <- update(sug, .~- sow)

## Construct restriction matrices from models
L.h <- model2restrictionMatrix(sug, sug.h); L.h
L.s <- model2restrictionMatrix(sug, sug.s); L.s

## Construct submodels from restriction matrices
mod.h <- restrictionMatrix2model(sug, L.h); mod.h
mod.s <- restrictionMatrix2model(sug, L.s); mod.s

## The models have the same fitted values
plot(fitted(mod.h), fitted(sug.h))
plot(fitted(mod.s), fitted(sug.s))
## and the same log likelihood
logLik(mod.h)
logLik(sug.h)
**pb-modcomp**

Model comparison using parametric bootstrap methods.

**Description**

Model comparison of nested models using parametric bootstrap methods. Implemented for some commonly applied model types.

**Usage**

```r
pbmodcomp(largemodel, smallmodel, nsim = 1000, ref = NULL, seed = NULL, 
cl = NULL, details = 0)
```

```r
## S3 method for class 'merMod'
pbmodcomp(largemodel, smallmodel, nsim = 1000, ref = NULL, 
seed = NULL, cl = NULL, details = 0)
```

```r
## S3 method for class 'lm'
pbmodcomp(largemodel, smallmodel, nsim = 1000, ref = NULL, 
seed = NULL, cl = NULL, details = 0)
```

**Arguments**

- `largemodel`: A model object. Can be a linear mixed effects model or generalized linear mixed effects model (as fitted with `lmer()` and `glmer()` function in the `lme4` package) or a linear normal model or a generalized linear model. The `largemodel` must be larger than `smallmodel` (see below).

- `smallmodel`: A model of the same type as `largemodel` or a restriction matrix.

- `nsim`: The number of simulations to form the reference distribution.

- `ref`: Vector containing samples from the reference distribution. If NULL, this vector will be generated using `PBrefdist()`.

- `seed`: A seed that will be passed to the simulation of new datasets.

- `cl`: A vector identifying a cluster; used for calculating the reference distribution using several cores. See examples below.

- `details`: The amount of output produced. Mainly relevant for debugging purposes.
Details

The model object must be fitted with maximum likelihood (i.e. with REML=FALSE). If the object is fitted with restricted maximum likelihood (i.e. with REML=TRUE) then the model is refitted with REML=FALSE before the p-values are calculated. Put differently, the user needs not worry about this issue.

Under the fitted hypothesis (i.e. under the fitted small model) nsim samples of the likelihood ratio test statistic (LRT) are generated.

Then p-values are calculated as follows:

- **LRT**: Assuming that LRT has a chi-square distribution.
- **PBtest**: The fraction of simulated LRT-values that are larger or equal to the observed LRT value.
- **Bartlett**: A Bartlett correction is of LRT calculated from the mean of the simulated LRT-values
- **Gamma**: The reference distribution of LRT is assumed to be a gamma distribution with mean and variance determined as the sample mean and sample variance of the simulated LRT-values.
- **F**: The LRT divided by the number of degrees of freedom is assumed to be F-distributed, where the denominator degrees of freedom are determined by matching the first moment of the reference distribution.

Note

It can happen that some values of the LRT statistic in the reference distribution are negative. When this happens one will see that the number of used samples (those where the LRT is positive) are reported (this number is smaller than the requested number of samples).

In theory one can not have a negative value of the LRT statistic but in practice one can: We speculate that the reason is as follows: We simulate data under the small model and fit both the small and the large model to the simulated data. Therefore the large model represents - by definition - an overfit; the model has superfluous parameters in it. Therefore the fit of the two models will for some simulated datasets be very similar resulting in similar values of the log-likelihood. There is no guarantee that the the log-likelihood for the large model in practice always will be larger than for the small (convergence problems and other numerical issues can play a role here).

To look further into the problem, one can use the **PBrefdist()** function for simulating the reference distribution (this reference distribution can be provided as input to **PBmodcomp()**). Inspection sometimes reveals that while many values are negative, they are numerically very small. In this case one may try to replace the negative values by a small positive value and then invoke **PBmodcomp()** to get some idea about how strong influence there is on the resulting p-values. (The p-values get smaller this way compared to the case when only the originally positive values are used).

Author(s)

Søren Højsgaard <soren@math.aau.dk>

References

See Also

krmodcomp, PBrefdist

Examples

data(beets, package="pbkrtest")
head(beets)

## Linear mixed effects model:
sug <- lmer(sugpct ~ block + sow + harvest + (1|block:harvest), data=beets, REML=FALSE)
sug.h <- update(sug, .~ -harvest)
sug.s <- update(sug, .~ -sow)

anova(sug, sug.h)
PBmodcomp(sug, sug.h, nsim=50)
anova(sug, sug.h)
PBmodcomp(sug, sug.s, nsim=50)

## Linear normal model:
sug <- lm(sugpct ~ block + sow + harvest, data=beets)
sug.h <- update(sug, .~ -harvest)
sug.s <- update(sug, .~ -sow)

anova(sug, sug.h)
PBmodcomp(sug, sug.h, nsim=50)
anova(sug, sug.s)
PBmodcomp(sug, sug.s, nsim=50)

## Generalized linear model
counts <- c(18,17,15,20,10,20,25,13,12)
outcome <- gl(3,1,9)
treatment <- gl(3,3)
d.Ad <- data.frame(treatment, outcome, counts)
head(d.Ad)
glm.D93 <- glm(counts ~ outcome + treatment, family = poisson())
glm.D93.o <- update(glm.D93, .~ -outcome)
glm.D93.t <- update(glm.D93, .~ -treatment)

anova(glm.D93, glm.D93.o, test="Chisq")
PBmodcomp(glm.D93, glm.D93.o, nsim=50)
anova(glm.D93, glm.D93.t, test="Chisq")
PBmodcomp(glm.D93, glm.D93.t, nsim=50)

## Generalized linear mixed model (it takes a while to fit these)
## Not run:
(gm1 <- glmer(cbind(incidence, size - incidence) ~ period + (1 | herd),
data = cbpp, family = binomial))
(gm2 <- update(gm1, .~.-period))
anova(gm1, gm2)
PBmodcomp(gm1, gm2)
pb-refdist

Calculate reference distribution using parametric bootstrap

Description

Calculate reference distribution of likelihood ratio statistic in mixed effects models using parametric bootstrap.
Usage

```r
PBrefdist(largeModel, smallModel, nsim = 1000, seed = NULL, cl = NULL,
          details = 0)
```

### S3 method for class 'lm'

```r
PBrefdist(largeModel, smallModel, nsim = 1000, seed = NULL,
          cl = NULL, details = 0)
```

### S3 method for class 'merMod'

```r
PBrefdist(largeModel, smallModel, nsim = 1000, seed = NULL,
          cl = NULL, details = 0)
```

### S3 method for class 'mer'

```r
PBrefdist(largeModel, smallModel, nsim = 1000, seed = NULL,
          cl = NULL, details = 0)
```

Arguments

- **largeModel**: A linear mixed effects model as fitted with the `lmer()` function in the `lme4` package. This model must be larger than `smallModel` (see below).
- **smallModel**: A linear mixed effects model as fitted with the `lmer()` function in the `lme4` package. This model must be smaller than `largeModel` (see above).
- **nsim**: The number of simulations to form the reference distribution.
- **seed**: Seed for the random number generation.
- **cl**: A vector identifying a cluster; used for calculating the reference distribution using several cores. See examples below.
- **details**: The amount of output produced. Mainly relevant for debugging purposes.

Details

The model object must be fitted with maximum likelihood (i.e. with `REML=FALSE`). If the object is fitted with restricted maximum likelihood (i.e. with `REML=TRUE`) then the model is refitted with `REML=FALSE` before the p-values are calculated. Put differently, the user needs not worry about this issue.

Value

A numeric vector

Author(s)

Søren Højsgaard <sorenh@math.aau.dk>

References

See Also

PBmodcomp, KRmodcomp

Examples

data(beets)
head(beets)
beet0 <- lmer(sugpct ~ block + sow + harvest + (1|block:harvest), data=beets, REML=FALSE)
beet_no.harv <- update(beet0, -.harvest)
rr <- PBrefdist(beet0, beet_no.harv, nsim=20)
rr

## Note: Many more simulations must be made in practice.

## Computations can be made in parallel using several processors:
## Not run:
c1 <- makeSOCKcluster(rep("localhost", 4))
clusterEvalQ(c1, library(lme4))
clusterSetupSPRNG(c1)
rr <- PBrefdist(beet0, beet_no.harv, nsim=20)
stopCluster(c1)

## End(Not run)
## Above, 4 cpu's are used and 5 simulations are made on each cpu.
Index

*Topic datasets
  data-beets, 2
  data-budworm, 3

*Topic inference
  get_ddf_Lb, 5
  kr-modcomp, 7
  kr-vcov, 8
  pb-modcomp, 12
  pb-refdist, 15

*Topic models
  get_ddf_Lb, 5
  kr-modcomp, 7
  kr-vcov, 8
  pb-modcomp, 12
  pb-refdist, 15

*Topic utilities
  getkr, 4
  model-coerce, 10

beets (data-beets), 2
budworm (data-budworm), 3
data-beets, 2
data-budworm, 3
ddf_Lb (get_ddf_Lb), 5

get_ddf_Lb, 5
get_Lb_ddf (get_ddf_Lb), 5
get_SigmaG (kr-vcov), 8
getKR, 8, 9
getKR (getkr), 4
getkr, 4
getLRT (pb-modcomp), 12

kr-modcomp, 7
kr-vcov, 8
KRmodcomp, 5, 6, 9, 11, 14, 17
KRmodcomp (kr-modcomp), 7
KRmodcomp_internal (kr-modcomp), 7

Lb_ddf (get_ddf_Lb), 5

lmer, 8, 9
LMM_Sigma_G (kr-vcov), 8

model-coerce, 10
model2restrictionMatrix, 6
model2restrictionMatrix (model-coerce), 10

pb-modcomp, 12
pb-refdist, 15
PBmodcomp, 5, 8, 9, 11, 17
PBmodcomp (pb-modcomp), 12
PBrefdist, 11, 14
PBrefdist (pb-refdist), 15
plot.XXmodcomp (pb-modcomp), 12

restrictionMatrix2model, 6
restrictionMatrix2model (model-coerce), 10

vcovAdj, 5, 6, 8, 9
vcovAdj (kr-vcov), 8
vcovAdj@ (kr-vcov), 8
vcovAdj2 (kr-vcov), 8
vcovAdj_internal (kr-vcov), 8

18