Package ‘BHH2’

January 20, 2025

Version 2016.05.31

Date 2016-05-31

Title Useful Functions for Box, Hunter and Hunter II

Author Ernesto Barrios

Maintainer Kjetil B.Halvorsen <kjetil1001@gmail.com>

Description Functions and data sets reproducing some examples in Box, Hunter and Hunter II. Useful for statistical design of experiments, especially factorial experiments.

Depends R (>= 2.0.0)

Suggests FrF2

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2016-05-31 23:18:13

Contents

`anovaPlot` *Graphical Anova*

Description

Dots plot displaying the deviations of factor levels from the mean showing the residuals as reference distribution.

Usage

```
anovaPlot(obj, stacked = TRUE, base = TRUE, axes = TRUE,
faclab = TRUE, labels = FALSE, cex = par("cex"),
cex.lab = par("cex.lab"), ...)
```

Arguments

`obj`	Object of class `aov` or `lm` for which marginal deviations from the mean and the residuals distribution is displayed.
`stacked`	logical. If `TRUE` and if it is necessary the dots are stacked, otherwise all points are displayed at same level with possible overlapping.
`base`	logical. By default a base line is displayed for each factor. If `FALSE` this line is omitted.
`axes`	logical. By default a scaled axes is drawn for each factor. If `FALSE` the axes are omitted.
`faclab`	logical. By default factor effect names and 'Residuals' are used to label each dot plot. No axis is labelled otherwise.
`labels`	logical. By default, dots are used to the display. If `labels=TRUE` then factor levels are displayed for the factor dots plots and sequential enumeration is used for the residuals.
`cex`	numeric. Expansion factor of the character used for labelling the factor levels.
`cex.lab`	numeric. Expansion factor of the character used for labelling each factor.
`...`	additional parameters passed to the `dots` function.

Details

Dots plot are displayed for the scaled deviations of factor levels from the grand mean and the distribution of the residuals is shown at the bottom of the plot for graphical comparison. The scaled factor for the factor deviations is $\sqrt{n/k}$, where k and n are the factor and residuals degrees of freedom reported by `anova(obj)`. If `labels=TRUE` then the factor levels are used for as points instead of dots. This option is useful to post labelling the dot plots. See `dots` function. The Anova plot is built in a $(0, 1) \times (0, 1)$ plot area. The area plot is divided to accommodate each of the factors and the residual at the bottom of the plotting area. The function returns a list with the coordinates of all the dots displayed.

Value

The function is called for graphical display of factor levels mean and residuals as reference distribution. An invisible list with the actual (x,y) coordinates used for each of the factors and residuals.

warning

The function identifies as an interaction factor any factor with the colon character ":" in its name. Factors like "I(A:B)" will give you problems.

Note

The anova plot presented here is thought for graphical comparison of factor effects in one-layer balanced designed experiments. The function is not prepared for general situations. However, representation of some simple split-plot experiments is possible.

Author(s)

Ernesto Barrios

References

Box G. E. P. (2000). *Box on Quality*. Edited by G. C. Tiao et al. New York: Wiley.

Box G. E. P, Hunter, J. S. and Hunter, W. C. (2005). *Statistics for Experimenters II*. New York: Wiley.

See Also

dots, dotPlot

Examples

```
library(BHH2)
data(heads.data)
heads.data$periods <- factor(heads.data$periods)
heads.data$heads <- factor(heads.data$heads)

heads.aov <- aov(resp~periods+heads,data=heads.data)
anovaPlot(heads.aov)

anovaPlot(heads.aov,labels=TRUE,faclab=TRUE)
```

corrosion.data *Corrosion data*

Description

Corrosion resistance study data set.

Usage

```
data(corrosion.data)
```

Format

A data frame with 24 observations on the following 4 variables.

run factor with 6 levels. The casting order.

heats factor with 3 levels. The casting temperature.

coating factor with 4 levels. The coating treatment.

resistance numeric vector. Corrosion resistance response.

References

Box G. E. P, Hunter, J. S. and Hunter, W. C. (2005). *Statistics for Experimenters II*. New York: Wiley.

Examples

```
data(corrosion.data)
str(corrosion.data)
plot(corrosion.data)
```

dotPlot *Dot plot: scatter plot with stacked dots similar to the stem-and-leaf plot*

Description

Displays an one-dimensional scatter plot with stacking similar to stem-and-leaf plot or histograms.

Usage

```
dotPlot(x, y = 0, xlim = range(x,na.rm=TRUE), xlab = NULL,
    scatter = FALSE, hmax = 1, base = TRUE, axes = TRUE, frame = FALSE,
    pch = 21, pch.size = "x", labels = NULL, hcex = 1, cex =par("cex"),
    cex.axis = par("cex.axis"),...)
```

Arguments

`x`	numeric vector to be displayed.
`y`	numeric. Height of the basis of the plot.
`xlim`	numeric. Range of the x axis.
`xlab`	character string. label for the horizontal axis.
`scatter`	logical. If TRUE a one-dimensional scatter plot of `x`, similar to `rug`, is displayed at the base of the plot.
`hmax`	numeric. Height of the highest dot. `hmax=1` as default. See **Details**.
`base`	logical. If `TRUE` (default) a base line for the dots (characters) is displayed.
`axes`	logical. If `TRUE` labelled axis is displayed.
`frame`	logical. If `FALSE` the plot frame is omitted.
`pch`	numeric or character. Character number or character to be used for the display.
`pch.size`	numeric. Character to be used to distribute the "dots" (`pch`). See **Details**.
`labels`	character vector. If `NULL` (default) each point (dot) is displayed using character `pch`, otherwise vector `labels` is used for the display. See **Details**.
`hcex`	numeric. Expansion (shrink) factor for character height. See **Details**.
`cex`	numeric. Expansion factor used for character display. See `par`.
`cex.axis`	numeric. Expansion factor used in case of labelling the axis.
`...`	additional graphical parameters.

Details

Basically function `dotPlot` calls function dots to display a stacked one-dimensional scatter plot within vertical limits 0 and 1. See `dots` for more details.

Value

The function is called for its side effect which is to produce one-dimensional scatter plot with stacking as described, for example, in Chambers et al. (1983) It returns invisible a data frame with the actual coordinates (in users units).

Note

Since the dots are stacked vertically, their alignment is subject to rounding errors. Dots may be slightly moved in either side from their actual value.

Author(s)

Ernesto Barrios

References

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983) *Graphical Methods for Data Analysis.* New York: Chapman \& Hall

See Also

dots, stem,hist,dotchart

Examples

```
library(BHH2)
data(tab03B1)
attach(tab03B1)
stem(yield) #stem-leaf plot
plt <- dotPlot(yield) # equivalent dotPlot

# same dot plot with max and min observations labelled
plt <- dotPlot(yield,xlim=c(75,95),xlab="yield",pch.size="x",hcex=1)
text(c(min(yield),max(yield),80),rep(0.05,3),c("min","max",80))
segments(80,min(plt$y),80,max(plt$y),lty=2)
detach()
```

---

dots *Dots display*

Description

The function adds to the current plot an one-dimensional scatter plot with stacking similar to a stem-leaf plot or histograms but using characters. .

Usage

```
dots(x = , y = 0.1, xlim = range(x,na.rm=TRUE), stacked = FALSE, hmax= 0.5,
base = TRUE, axes = FALSE, pch = 21, pch.size = "x", labels = NULL,
hcex = 1, cex = par("cex"), cex.axis = par("cex.axis"))
```

Arguments

x	numeric vector to be displayed in the dot plot.
y	numeric. Height of the dots (characters) at the base level. By default y=0.1 thinking on a plot with ylim=c(0,1).
xlim	numeric vector with 2 entries: xmin and xmax. These values determine the width of the displayed dot plot not necessarily equal to the limits of the plot.
stacked	logical. If TRUE characters are stacked, otherwise a scatter plot of the data is displayed at y level using character pch.
hmax	numeric. The maximum height in user units. By default hmax=0.5 thinking on a plot with ylim=c(0,1). See y.
base	logical. If TRUE a horizontal line is displayed at the bottom of the plot.
axes	logical. If TRUE an labelled axis is shown.
pch	numeric or character. Character number or character to be used for the display.

`pch.size`	numeric. Character to be used to distribute the "dots" (`pch`). See **Details**.
`labels`	character vector. If `NULL` (default) each point (dot) is displayed using character `pch`, otherwise vector `labels` is used for the display. See **Details**.
`hcex`	numeric. Expansion (shrink) factor for character height. See details.
`cex`	numeric. Expansion factor used for character display. See par.
`cex.axis`	numeric. Expansion factor used in case of labelling the axis.

Details

Function `dots` adds to the current plot a dot plot similar to a stem-and-leaf plot using characters specified by `pch` and `labels=NULL`. If `labels` is not `NULL` then it is expected to be a character vector and will will be used to display each of the points. Its use is repeated or cut short if necessary. The function computes the width and height size using character `pch.size` calling `strwidth` and `strheight`, but displays `pch` instead. Mainly this is used when `pch` is not given by a quoted character, for example, `pch=21`. Also, currently the par("mkh") is ignored so `hcex` is used to compute the "working" height of the characters: `hcex*strheight(pch.size,units="user")`. If `stacked=TRUE`, the base line is divided in subintervals of size `strwidth(pch.size)` and computed the number of points in each subinterval. If maximum number of stacked characters exceed `hmax` then the characters are overlapped to adjust their total height to `hmax`.

Value

Invisible data frame with columns (x,y,labels). 'x' and 'y' are the coordinates in user units of each point and 'labels' the corresponding character displayed.

Author(s)

Ernesto Barrios

See Also

`dotPlot`, `anovaPlot`

Examples

```
library(BHH2)
set.seed(4)
# Defines the height of the plot area between c(0,1)
dotPlot(rnorm(100),xlab="x")

x <- rnorm(100)

# plots (possibly) overlapping points at y=0.3
dots(x,y=0.3)
# plots (possibly) overlapping points at y=0.4
dots(x,y=0.4,stacked=TRUE,base=FALSE)
# plots (hopefully) stacked points at y=0.5 allowing the dots to as high as 0.9
dots(x,y=0.5,stacked=TRUE,base=FALSE,hmax=.9)
```

ffDesMatrix — *Full or fractional factorial design matrix generation*

Description

The function generates the design matrix provided the number of 2-levels design factors and defining relations.

Usage

```
ffDesMatrix(k, gen = NULL)
```

Arguments

k	numeric. The number of 2-levels design factors in the designs.
gen	list. If NULL (default) a full factorial design is generated. Otherwise, each component of the list is a numeric vector of corresponding to each of the defining relations used to compose the design. See **Details**.

Details

A defining relation is declared by a vector where the first entry corresponds to the left hand side (LHS) of the defining equation. For example, if k=5, and gen=list(c(-5,1,2,3,4)), then the defining equation is $-5 = 1 * 2 * 3 * 4$. A full 2-levels (-1,1) factorial design is generated. For each defining relation the LHS column is replaced by the corresponding columns product. At the end repeated runs are removed from the matrix.

Value

The function returns a 2-levels design matrix with k columns.

Author(s)

Ernesto Barrios

See Also

conf.design of the **conf.design** package, FrF2 from the **FrF2** package.

Examples

```
ffDesMatrix(5) # Full 2^5 factorial design
ffDesMatrix(5,gen=list(c(5,1,2,3,4))) # 2^(5-1) factorial design
ffDesMatrix(5,gen=list(c(4,1,2),c(-5,1,3))) # 2^(5-2) factorial design
```

ffFullMatrix *Full model matrix from a design matrix*

Description

The function builds the full matrix with the constant term, main effects and interactions from a design matrix.

Usage

```
ffFullMatrix(X, x, maxInt, blk = NULL)
```

Arguments

`X`	numeric matrix. Design matrix.
`x`	numeric vector. Design matrix entries to use to construct the full model matrix.
`maxInt`	numeric. Highest interaction order.
`blk`	numeric matrix. Each column correspond to a blocking factor.

Details

Columns `x` of matrix `X` are used for main effects. All the 2, ..., `maxInt` order interaction are constructed. The first columns of the final matrix correspond to the constant term (1's) and block factors.

Value

The function returns list with the following components:

`Xa`	matrix

. Augmented matrix with columns for the constant terms, blocking factors, main effects, second order interactions, ..., etc.

`x`	numeric vector. Design matrix `X` factor (column) numbers used to build the complete model matrix.
`maxInt`	numeric. The highest interaction order.
`nTerms`	numeric vector. Contains the number of blocking factors, main effects, 2nd order interaction effects, ..., etc.

Author(s)

Ernesto Barrios

See Also

`ffDesMatrix`

Examples

```
print(X <- ffDesMatrix(5,gen=list(c(5,1,2,3,4))))
ffFullMatrix(X[,1:4],x=c(1,2,3,4),maxInt=2,blk=X[,5])
ffFullMatrix(X[,1:5],x=c(1,3,5),maxInt=3)
```

---

heads.data *Machine heads data*

Description

Data set of the variability of machine heads in a quality improvement experiment.

Usage

```
data(heads.data)
```

Format

A data frame with 30 observations on the following 6 variables.

obs numeric. Observation number.

periods factor. Periods factor (P1, . . . , P6).

heads factor. Type of head factor (H1, . . . , H5).

days factor. Day factor (D1 and D2).

shifts factor. Shift factor (S1, S2, and S3).

resp numeric. Response.

Source

Box, G. E. P. (1993). "How to Get Lucky". *Quality Engineering*, Vol. 5, No. 3, pp 517-524.

References

Box G. E. P, Hunter, J. S. and Hunter, W. C. (2005). *Statistics for Experimenters II*. New York: Wiley.

Examples

```
data(heads.data)
str(heads.data)
plot(heads.data)
```

`lambdaPlot` *Lambda plot: traces the t and F statistics in Box-Cox transformation of the response*

Description

Trace regression coefficients' t-values or F-ratios for different values of λ in the Box-Cox transformation.

Usage

```
lambdaPlot(mod, lambda = seq(-1, 1, by = 0.1), stat = "F", global = TRUE,
    cex = par("cex"), ...)
```

Arguments

`mod`	list. A list of class `lm`.
`lambda`	numeric. The values of λ in the Box-Cox transformation. See **Details**.
`stat`	character. Either `"t"` of `"F"`, corresponding to the coefficients' t-values or F-ratios to display.
`global`	logical. Applied only for `stat="F"`, if `TRUE`, the model's F-ratio is traced, otherwise the coefficients' F-statistics.
`cex`	numeric. Expansion factor used to label the trace lines.`par("cex")` bu default.
`...`	additional graphical parameters passed to `plot` function.

Details

The response is transformed as $Y = (y^\lambda - 1)/\lambda$ for each value of λ (`lambda`) and the model refitted. The t-values or F-ratios of the coefficients are saved for the display. If `global=TRUE`, then the F-ratio of the whole model is plotted instead.

Value

The function returns an invisible list with components:

`lambda`	numeric. Vector of length `m` with the different values of λ.
`t.lambda`	matrix (`k x m`), where `m` is the number of coefficients in model `mod` without the intercept, with the coefficient's t-values.
`f.lambda`	matrix (`k x m`) with the coefficient's F-values. if `global = FALSE`, otherwise the matrix is (`1 x m`), with the corresponding model F-ratio.

Note

For each value of λ the model is refitted. Computations can be done more efficiently and will be incorporated in future versions.

Author(s)

Ernesto Barrios

References

Box, G. E. P. and C. Fung (1995) "The Importance of Data Transformation in Designed Experiments for Life Testing". *Quality Engineering*, Vol. 7, No. 3, pp. 625-68.

Box G. E. P, Hunter, J. S. and Hunter, W. C. (2005). *Statistics for Experimenters II*. New York: Wiley.

Examples

```
library(BHH2)
# Lambda Plot tracing t values.
data(woolen.data)
woolen.lm <- lm(y~x1+x2+x3+I(x1^2)+I(x2^2)+I(x3^2)+
                    I(x1*x2)+I(x1*x3)+I(x2*x3)+I(x1*x2*x3),data=woolen.data)
lambdaPlot(woolen.lm,cex=.8,stat="t")

# Lambda Plot tracing F values.
woolen2.lm <- lm(y~x1+x2+x3,data=woolen.data)
lambdaPlot(woolen2.lm,lambda=seq(-1,1,length=41),stat="F",global=TRUE)

# Lambda Plot tracing F values.
data(poison.data)
poison.lm <- lm(y~treat*poison,data=poison.data)
lambdaPlot(poison.lm,lambda=seq(-3,1,by=.1),stat="F",global=FALSE)
```

---

penicillin.data *Penicillin data*

Description

Penicillin yield example data set.

Usage

```
data(penicillin.data)
```

Format

A data frame with 20 observations on the following 4 variables.

blend factor with 5 levels: `B1 B2 B3 B4 B5`. Blend factor used to block the experiment.

run numeric vector. Run order within the blocking (Blend) factor.

treat factor with levels: `A B C D`. The process variants called treatment.

yield numeric vector. Experiment yield response.

Source

Box G. E. P, Hunter, W. C. and Hunter, J. S. (1978). *Statistics for Experimenters*. New York: Wiley.

References

Box G. E. P, Hunter, J. S. and Hunter, W. C. (2005). *Statistics for Experimenters II*. New York: Wiley.

Examples

```
data(penicillin.data)
str(penicillin.data)
plot(penicillin.data)
```

---

permtest — *Permutation test: randomization test for small size samples*

Description

Permutation test for means and variance comparisons.

Usage

```
permtest(x, y = NULL)
```

Arguments

`x`	numeric vector. Sample group X.
`y`	numeric vector. Sample group Y.

Details

In the one–sample problem, the function builds all 2^n possible $\pm x_i$ combinations. For the two–sample problem, all possible $B(n+m, n)$ samples size n (=`length(x)`) and m (=`length(y)`) are generated and the permutation distributions for the *t*-statistics and *F*-ratios. *p*-values are computed based on these distributions.

Value

The function returns the number N of different samples generated for the permutation distribution, the observed *t*-statistic, its *p*-value, based on both, the parametric and permutation distributions as well as the observed *F*-ratio and its corresponding *p*-values. The test may take a long time to generate all the possible combinations. It has been tested for $n + m = 22$ and $n < 12$.

WARNING

The test may take a long time to generate all the possible combinations.

Author(s)

Ernesto Barrios

References

Box G. E. P, Hunter, J. S. and Hunter, W. C. (2005). *Statistics for Experimenters II*. New York: Wiley.

See Also

`onet.permutation` and `towt.permutation` of **DAAG** package, and `perm.test` of the **exactRankTests**.

Examples

```
library(BHH2)

# Permutation test for Tomato Data
data(tomato.data)
cat("Tomato Data (not paired):\n")
attach(tomato.data)
a <- pounds[fertilizer=="A"]
b <- pounds[fertilizer=="B"]
print(round(test <- permtest(b,a),3))
detach()

# Permutation test for Boy's Shoes Example
data(shoes.data)
cat("Shoes Data (paired):\n")
attach(shoes.data)
x <- matB-matA
print(round(test <- permtest(x),3))
detach()
```

---

poison.data *Poison example data set*

Description

Poison data from Biological Experiment

Usage

```
data(poison.data)
```

Format

This data frame contains the following columns:

poison factor with 3 levels: P1, P2 and P3.

treat factor with 4 levels: trA, trB, trC and trD.

y numeric. Survival time as response.

Source

Box, G. E. P. and D. R. Cox, An Analysis of Transformations (with discussion), *Journal of the Royal Statistical Society*, Series B, Vol. 26, No. 2, pp. 211–254.

References

Box G. E. P, Hunter, J. S. and Hunter, W. C. (2005). *Statistics for Experimenters II*. New York: Wiley.

Examples

```
data(poison.data)
str(poison.data)
plot(poison.data)
```

---

shoes.data *Boys' shoes data set*

Description

Data for the Boys' Shoes Example.

Usage

```
data(shoes.data)
```

Format

A data frame with 10 observations on the following 5 variables.

boy numeric. Boy number.

matA numeric. Amount of wear of shoe made from material A.

sideA factor. Foot side which shoe of material A is used.

matB numeric. Amount of wear of shoe made from material B.

sideB factor. Foot side which shoe of material B is used.

Source

Box G. E. P, Hunter, W. C. and Hunter, J. S. (1978). *Statistics for Experimenters*. New York: Wiley.

References

Box G. E. P, Hunter, J. S. and Hunter, W. C. (2005). *Statistics for Experimenters II*. New York: Wiley.

Examples

```
data(shoes.data)
str(shoes.data)
plot(shoes.data)
```

`subsets` *Generation of all the combinations of k elements from n possible*

Description

Generates all different subsets of size r chosen from n different elements.

Usage

```
subsets(n, r, v = 1:n)
```

Arguments

`n`	numeric. Number of elements to choose from.
`r`	numeric. Size of the subsets.
`v`	vector. Numeric or character vector of size `n` with the labels of the n elements to choose from.

Value

A matrix of dimension $(N \times r)$, where N is the total number of different combinations of r elements chosen from n possible.

Note

This particular version of the function was taken from a message from Bill Venables to '`r-help`' list on Sun, 17 Dec 2000.

Author(s)

Bill Venables <`Bill.Venables@cmis.csiro.au`>

References

Venables, Bill. "Programmers Note", R-News, Vol 1/1, Jan. 2001. http://cran.r-project.org/doc/Rnews

See Also

`combinations` of the **gtools** package.

Examples

```
library(BHH2)
subsets(5,3)
subsets(5,3,letters)
subsets(5,3,c(10,20,30,50,80))
```

---

tab03B1 *Table 3.2*

Description

Production record of 210 consecutive batch yield values

Usage

```
data(tab03B1)
```

Format

This data frame contains the following columns:

yield a numeric vector

ave10 a numeric vector. Moving average of last 10 observations. First 9 entries NA

Details

The `tab03B1` data frame has 210 rows and 2 columns.

Source

Box G. E. P, Hunter, W. C. and Hunter, J. S. (1978). *Statistics for Experimenters*. New York: Wiley.

References

Box G. E. P, Hunter, J. S. and Hunter, W. C. (2005). *Statistics for Experimenters II*. New York: Wiley.

Examples

```
library(BHH2)
data(tab03B1)
attach(tab03B1)
stem(yield)
stem(ave10)
plot(yield,xlab="time order",ylab="yield")
detach()
```

`tab03B2` *Table 3.3*

Description

Reference set of differences between averages of two adjacent sets of 10 successive batches.

Usage

```
data(tab03B2)
```

Format

This data frame contains the following columns:

diff10 a numeric vector

Details

The `tab03B2` data frame has 200 rows and 1 column. First 9 entries are `NA`.

Source

Box G. E. P, Hunter, W. C. and Hunter, J. S. (1978). *Statistics for Experimenters*. New York: Wiley.

References

Box G. E. P, Hunter, J. S. and Hunter, W. C. (2005). *Statistics for Experimenters II*. New York: Wiley.

Examples

```
library(BHH2)
data(tab03B2)
attach(tab03B2)
# displays the differences as dot plot (similar to histograms)
plt <- dotPlot(diff10,xlim=2.55*c(-1,+1),xlab="differences")
segments(1.3,0,1.3,max(plt$y))  #vertical line at x=1.3
detach()
```

tomato.data — *Tomato plants data set*

Description

Yield of tomato plants under two different fertilizers.

Usage

```
data(tomato.data)
```

Format

This data frame contains the following columns:

pos numeric. Row position

pounds numeric. Plant's yield in pounds.

fertilizer factor. Type of fertilizer (A or B).

Source

Box G. E. P, Hunter, W. C. and Hunter, J. S. (1978). *Statistics for Experimenters*. New York: Wiley.

References

Box G. E. P, Hunter, J. S. and Hunter, W. C. (2005). *Statistics for Experimenters II*. New York: Wiley.

Examples

```
data(tomato.data)
str(tomato.data)
plot(tomato.data)
```

woolen.data — *Textile experiment data set*

Description

Woolen thread experiment data set.

Usage

```
data(woolen.data)
```

Format

This data frame with 27 observations contains the following columns:

x1 numeric. Length of test specimens factor.

x2 numeric. Amplitude of loading cycle factor.

x3 numeric. Load factor.

y numeric. Cycles to failure response.

Source

Box, G. E. P. and D. R. Cox (1964). "An Analysis of Transformations (with discussion)", *Journal of the Royal Statistical Society*, Series B, Vol. 26, No. 2, pp. 211–254.

References

Box G. E. P, Hunter, J. S. and Hunter, W. C. (2005). *Statistics for Experimenters II*. New York: Wiley.

Examples

```
data(woolen.data)
str(woolen.data)
plot(woolen.data)
```

Index