

Package ‘ANOM’

April 12, 2017

Title Analysis of Means

Description Analysis of means (ANOM) as used in technometrical computing. The package takes results from multiple comparisons with the grand mean (obtained with 'multcomp', 'SimComp', 'nparcomp', or 'MCPAN') or corresponding simultaneous confidence intervals as input and produces ANOM decision charts that illustrate which group means deviate significantly from the grand mean.

Version 0.5

Author Philip Pallmann

Maintainer Philip Pallmann <p.pallmann@lancaster.ac.uk>

Depends R (>= 3.0.0)

Imports ggplot2, MCPAN, multcomp, nparcomp, SimComp

Suggests knitr, lme4, nlme, sandwich

VignetteBuilder knitr

License GPL (>= 2)

LazyData true

Date 2017-04-10

URL <https://github.com/PhilipPallmann/ANOM>

NeedsCompilation no

Repository CRAN

Date/Publication 2017-04-12 13:32:33 UTC

R topics documented:

ANOM	2
ANOMgen	7
ANOMintern	9
hemoglobin	10
math	11
spring	12
waterfilter	12

Index	14
--------------	-----------

Description

Compute an analysis of means (i.e., a multiple contrast test involving comparisons of each group versus the grand mean) and draw a decision chart as commonly used in technometrics.

Usage

```
ANOM(mc, xlabel=NULL, ylabel=NULL, printn=TRUE, printp=TRUE,
      stdep=NULL, stind=NULL, pst=NULL, pbin=NULL, bg="white",
      bgrid=TRUE, axlsize=18, axtsize=25, npsize=5, psize=5,
      lwidth=1, dlstyle="dashed", fillcol="darkgray")
```

Arguments

<code>mc</code>	An object of class <code>glht</code> , <code>SimCi</code> , <code>mctp</code> , or <code>binomRDci</code> involving group comparisons with the grand mean of all groups. See details.
<code>xlabel</code>	An optional character string specifying the label of the horizontal axis.
<code>ylabel</code>	An optional character string specifying the label of the vertical axis.
<code>printn</code>	A logical. Should per-group sample sizes be included in the chart? Default is TRUE.
<code>printp</code>	A logical. Should simultaneous p-values be included in the chart? Default is TRUE.
<code>stdep</code>	A numerical vector giving the values of the dependent variable. Only required if an object of class <code>SimCi</code> is inserted for <code>mc</code> , otherwise ignored. Default is NULL.
<code>stind</code>	A factor specifying the levels of the independent variable. Only required if an object of class <code>SimCi</code> is inserted for <code>mc</code> , otherwise ignored. Default is NULL.
<code>pst</code>	An object of class <code>SimTest</code> . Only required if an object of class <code>SimCi</code> is inserted for <code>mc</code> and simultaneous p-values are to be printed (<code>printp=TRUE</code>), otherwise ignored. Default is NULL.
<code>pbin</code>	An object of class <code>binomRDtest</code> . Only required if an object of class <code>binomRDci</code> is inserted for <code>mc</code> and simultaneous p-values are to be printed (<code>printp=TRUE</code>), otherwise ignored. Default is NULL.
<code>bg</code>	A character string. Should the plot's background be "white" (default) or "gray" (or "grey")?
<code>bgrid</code>	A logical. Should background grid lines be plotted? Default is TRUE.
<code>axlsize</code>	A numerical value specifying the font size of the axis labels. Default is 18.
<code>axtsize</code>	A numerical value specifying the font size of the axis titles. Default is 25.
<code>npsize</code>	A numerical value specifying the font size of the sample sizes and p-values (if printed). Default is 5.
<code>psize</code>	A numerical value specifying the size of the points (group means). Default is 5.

<code>lwidth</code>	A numerical value specifying the width of the lines (grand mean, decision limits, vertical connections). Default is 1.
<code>dlstyle</code>	A character string specifying the style of the decision limits. Default is "dashed".
<code>fillcol</code>	A character string specifying the color of the area of no significant deviation from the grand mean. Default is "darkgray".

Details

The 'standard' version of ANOM is invoked by inserting a `glht` object (created with function `glht` from package `multcomp` using a contrast matrix of type `GrandMean`) for `mc`. The `glht` object must be based on one of the model types `aov`, `lm`, `glm`, `gls`, `lme`, or `lmer`. That is, ANOM is feasible not only for simple linear (ANOVA) models with Gaussian data, but indeed for a broad range of (semi-)parametric models e.g., with Poisson or binomial data, hierarchical and clustered structures, and more (Hothorn et al. 2008). If the underlying model is a Poisson or binomial GLM (involving log and logit link functions, respectively), the effects are automatically transformed back to counts (Poisson) or proportions (binomial).

When analyzing binomial data, we need to distinguish between two data formats. As long as replicated observations of proportions are available for each group (i.e., a within-group variance can be estimated), we can fit a binomial GLM. However, if there is only one proportion per group (i.e., we have a 2-by-k data table), we need to fall back on some simpler procedure than a GLM. In the case of ANOM, we insert a `binomRDci` object (created with function `binomRDci` from package `MCPAN`) for `mc` with the contrasts being of type `GrandMean`. If printing out simultaneous p-values with the ANOM decision chart is desired (`printp=TRUE`), a `binomRDtest` object (generated with function `binomRDtest`) must be inserted for `pbin` (see examples). Notice that `printn=TRUE` prints the total sample size per group when inserting an object of class `binomRDci` for `mc`; by contrast, inserting a `glht` object based on a binomial GLM makes `printn=TRUE` print the number of independent observations of proportions per group.

There are two options for coping with heterogeneous variances. The 'standard' version using `glht` objects is basically designed for homoscedastic data but can also cope with unequal variances by employing sandwich covariance estimates (Herberich et al. 2010). All you have to do is set the option `vcov=vcovHC` in the `glht` call (see examples).

A different heterogeneity adjustment using multiple Satterthwaite degrees of freedom (Hasler and Hothorn 2008) is invoked by inserting a `SimCi` object (created with function `SimCiDiff` from package `SimComp` with the setting `covar.equal=FALSE`) for `mc`. Notice that `covar.equal=TRUE` would compute a 'standard' multiple contrast test assuming equal variances, which is just the opposite of what we want.

Another instance when the `SimComp` package proves useful is for ANOM with ratios i.e., we assess each group's percentage change in comparison to the grand mean (which is always 100%). Again, we insert an object of class `SimCi` for `mc` but now we generated it with function `SimCiRatio` instead of `SimCiDiff`. Setting `covar.equal` to `TRUE` gives the homoscedastic version whereas `FALSE` calls the heteroscedastic variant using multiple degrees of freedom.

All ANOM functionality based on objects from `SimComp` requires you to submit the data (dependent and independent variable via `stdep` and `stind`, respectively) separately (see examples). Moreover, if you wish to print simultaneous p-values with your ANOM decision chart, you may compute them using function `SimTestDiff` or `SimTestRat` and insert the resulting object for `pst`. Again, don't forget to set the option `covar.equal=FALSE` if you want to account for heteroscedasticity. Make sure in all cases to perform comparisons of type `GrandMean`.

Nonparametric ANOM is performed by inserting an object of class `mctp` (created with function `mctp` from package `nparcomp`) for `mc`. It is based on estimation of relative effects via global pseudo-rankings as proposed by Konietschke et al. (2012). This procedure can naturally cope with heteroscedasticity in the data.

The relative effect of two independent random variables X_1 and X_2 following some distributions F_1 and F_2 , respectively, is generally defined as

$$p = P(X_1 < X_2) + 0.5P(X_1 = X_2).$$

Loosely speaking, p is the probability that X_1 takes smaller values than X_2 (plus half the probability of taking equal values). Hence when $p < 0.5$, X_1 is stochastically more likely to take larger values than X_2 , and vice versa for $p > 0.5$.

Creating the `mctp` object requires to specify a grand-mean-type contrast matrix by hand (see examples) as it is not among the options provided by the package `nparcomp`. For the test statistics you may choose one out of three asymptotic approximation methods:

- a multivariate t approximation with Satterthwaite degrees of freedom (`asy.method="mult.t"`), which works reasonably well most of the time,
- a multivariate normal approximation (`asy.method="normal"`), which is unfavorable with small sample sizes,
- a Fisher transform (`asy.method="fisher"`), which ensures that the decision limits preserve the range of $[-1, 1]$.

Make sure that the argument `correlation` in the function `mctp` is set to `TRUE`. Do not use the function `nparcomp` instead of `mctp` since it does not involve global ranking and is inoperative with ANOM-type contrast matrices.

Value

An ANOM decision chart.

Note

Notice that some of the more sophisticated ANOM variants (ratios, nonparametric, heteroscedastic with multiple degrees of freedom) are limited to one-way layouts without covariates.

Author(s)

Philip Pallmann <p.pallmann@lancaster.ac.uk>

References

- Djira, G. D., Hothorn, L. A. (2009) Detecting relative changes in multiple comparisons with an overall mean. *Journal of Quality Technology* 41(1), 60-65.
- Hasler, M. and Hothorn, L. A. (2008) Multiple contrast tests in the presence of heteroscedasticity. *Biometrical Journal* 50(5), 793–800.
- Herberich, E., Sikorski, J., Hothorn, T. (2010) A robust procedure for comparing multiple means under heteroscedasticity in unbalanced designs. *PLoS One* 5(3), e9788.

Hothorn, T., Bretz, F., Westfall, P. (2008) Simultaneous inference in general parametric models. *Biometrical Journal* 50(3), 346–363.

Konietschke, F., Hothorn, L. A., Brunner, F. (2012) Rank-based multiple test procedures and simultaneous confidence intervals. *Electronic Journal of Statistics* 6, 738–759.

Pallmann, P. and Hothorn, L. A. (2016) Analysis of means (ANOM): A generalized approach using R. *Journal of Applied Statistics*, 43(8), 1541–1560.

Examples

```
#####
### Standard ANOM (Gaussian, homoscedastic) ###
#####

### Devices of which brand filter bacteria significantly worse?
head(waterfilter)
str(waterfilter)

library(multcomp)
model <- lm(colonies ~ brand, waterfilter)
hom <- glht(model, mcp(brand="GrandMean"), alternative="less")
ANOM(hom)

#####
### Heteroscedastic ANOM ###
#####

## With sandwich covariance matrix estimate (Herberich et al. 2010)

library(multcomp)
library(sandwich)
het1 <- glht(model, mcp(brand="GrandMean"), alternative="less", vcov=vcovHC)
ANOM(het1)

## With multiple degrees of freedom (Hasler and Hothorn 2008)

library(SimComp)
het2 <- SimCiDiff(data=waterfilter, grp="brand", resp="colonies",
                  type="GrandMean", alternative="less", covar.equal=FALSE)
het2p <- SimTestDiff(data=waterfilter, grp="brand", resp="colonies",
                    type="GrandMean", alternative="less", covar.equal=FALSE)
ANOM(het2, stdep=waterfilter$colonies, stind=waterfilter$brand, pst=het2p)

#####
### ANOM for ratios ###
#####

## Homoscedastic

library(SimComp)
rel <- SimCiRat(data=waterfilter, grp="brand", resp="colonies",
               type="GrandMean", alternative="less", covar.equal=TRUE)
```

```

relp <- SimTestRat(data=waterfilter, grp="brand", resp="colonies",
                  type="GrandMean", alternative="less", covar.equal=TRUE)
ANOM(rel, stdep=waterfilter$colonies, stind=waterfilter$brand, pst=relp)

## Heteroscedastic (with multiple degrees of freedom)

library(SimComp)
relh <- SimCiRat(data=waterfilter, grp="brand", resp="colonies",
                type="GrandMean", alternative="less", covar.equal=FALSE)
relhp <- SimTestRat(data=waterfilter, grp="brand", resp="colonies",
                   type="GrandMean", alternative="less", covar.equal=FALSE)
ANOM(relh, stdep=waterfilter$colonies, stind=waterfilter$brand, pst=relhp)

#####
### Nonparametric ANOM ###
#####

# Compute sample sizes per group
ss <- tapply(waterfilter$colonies, waterfilter$brand, length)
# Build a grand-mean-type contrast matrix
library(multcomp)
Mat <- contrMat(ss, "GrandMean")

## Using a multivariate t approximation

library(nparcomp)
mult <- mctp(colonies ~ brand, data=waterfilter, type="UserDefined",
            contrast.matrix=Mat, alternative="less", info=FALSE,
            correlation=TRUE, asy.method="mult.t")
ANOM(mult)

## Using a range-preserving Fisher transform

library(nparcomp)
fish <- mctp(colonies ~ brand, data=waterfilter, type="UserDefined",
            contrast.matrix=Mat, alternative="less", info=FALSE,
            correlation=TRUE, asy.method="fisher")
ANOM(fish)

#####
### ANOM for binomial proportions ###
#####

### Which schools' math achievements differ from the grand mean?

head(math)
str(math)

## Based on Wald-type confidence intervals

library(MCPAN)
wald <- binomRDci(n=math$enrolled, x=math$proficient, names=math$school,
                 alternative="two.sided", method="Wald", type="GrandMean")

```

```

waldp <- binomRDtest(n=math$enrolled, x=math$proficient, names=math$school,
                    alternative="two.sided", method="Wald", type="GrandMean")
ANOM(wald, pbin=waldp)

## Based on add-2 confidence intervals

library(MCPAN)
add2 <- binomRDci(n=math$enrolled, x=math$proficient, names=math$school,
                alternative="two.sided", method="ADD2", type="GrandMean")
add2p <- binomRDtest(n=math$enrolled, x=math$proficient, names=math$school,
                    alternative="two.sided", method="ADD2", type="GrandMean")
ANOM(add2, pbin=add2p)

#####
### ANOM for variances ###
#####

### Springs of which brand are significantly more variable?

head(spring)
str(spring)

# Compute the median weight per brand
spring$median <- tapply(spring$weight, spring$brand, median)[spring$brand]
# Compute the absolute deviations from the median (robust Levene residuals)
spring$absdev <- with(spring, abs(weight - median))

library(multcomp)
mod <- lm(absdev ~ brand, spring)
test <- glht(mod, mcp(brand="GrandMean"))
ANOM(test)

```

ANOMgen

Generic Function for Drawing ANOM Decision Charts

Description

Graphical representation of the analysis of means: convert simultaneous confidence intervals (that were computed with ANY method) into ANOM decision limits and draw a decision chart as commonly used in technometrics.

Usage

```

ANOMgen(mu, n=NULL, gm=NULL, lo, up, names, alternative="two.sided",
        xlabel="Group", ylabel="Endpoint", printn=TRUE, p=NULL, bg="white",
        bgrid=TRUE, axlsize=18, axtsize=25, npsize=5, psize=5, lwidth=1,
        dlstyle="dashed", fillcol="darkgray")

```

Arguments

<code>mu</code>	A numeric vector of group means.
<code>n</code>	A numeric vector of sample sizes per group. Either <code>n</code> or <code>gm</code> must be provided.
<code>gm</code>	A single numeric value giving the grand mean of all groups. Either <code>n</code> or <code>gm</code> must be provided.
<code>lo</code>	A numeric vector of lower (simultaneous) confidence interval bounds for comparisons to the grand mean.
<code>up</code>	A numeric vector of upper (simultaneous) confidence interval bounds for comparisons to the grand mean.
<code>names</code>	An (optional) vector of characters specifying the groups' names.
<code>alternative</code>	A character string indicating the direction of the alternative hypothesis. Default is "two.sided", but may be changed to one-sided alternatives (either "greater" or "less").
<code>xlabel</code>	A character string specifying the label of the horizontal axis.
<code>ylabel</code>	A character string specifying the label of the vertical axis.
<code>printn</code>	A logical. Should per-group sample sizes be included in the chart? Default is TRUE. If <code>n</code> is left at NULL, the function automatically sets <code>printn</code> to FALSE.
<code>p</code>	An (optional) numeric vector of (simultaneous) p-values to be printed.
<code>bg</code>	A character string. Should the plot's background be "white" (default) or "gray" (or "grey")?
<code>bgrid</code>	A logical. Should background grid lines be plotted? Default is TRUE.
<code>axlsize</code>	A numerical value specifying the font size of the axis labels. Default is 18.
<code>axtsize</code>	A numerical value specifying the font size of the axis titles. Default is 25.
<code>npsize</code>	A numerical value specifying the font size of the sample sizes and p-values (if printed). Default is 5.
<code>psize</code>	A numerical value specifying the size of the points (group means). Default is 5.
<code>lwidth</code>	A numerical value specifying the width of the lines (grand mean, decision limits, vertical connections). Default is 1.
<code>dlstyle</code>	A character string specifying the style of the decision limits. Default is "dashed".
<code>fillcol</code>	A character string specifying the color of the area of no significant deviation from the grand mean. Default is "darkgray".

Details

This is a generic tool that translates (simultaneous) confidence intervals into ANOM decision limits.

Value

An ANOM decision chart.

Note

The confidence intervals must arise from comparisons to the grand mean; otherwise the ANOM chart is meaningless!

Author(s)

Philip Pallmann <p.pallmann@lancaster.ac.uk>

References

Pallmann, P. and Hothorn, L. A. (2016) Analysis of means (ANOM): A generalized approach using R. *Journal of Applied Statistics*, 43(8), 1541–1560.

Examples

```
### A toy example (n given, two-sided)
groupmeans <- c(2.8, 2.3, 3.4, 5.6)
samplesizes <- c(5, 5, 10, 5)
low <- c(-1.2, -1.7, -0.4, 1.6)
upp <- c(-0.2, -0.7, 0.2, 2.6)
names <- c("1st", "2nd", "3rd", "4th")
ANOMgen(mu=groupmeans, n=samplesizes, lo=low, up=upp, names=names, alternative="two.sided")

### Another toy example (gm given, one-sided, with p-values)
groupmeans <- c(2.8, 2.3, 3.4, 5.6)
gm <- 3.5
low <- rep(-Inf, 4)
upp <- c(-0.2, -0.7, 0.2, 2.6)
names <- c("1st", "2nd", "3rd", "4th")
pvalues <- c(0.01, 0.003, 0.8, 1)
ANOMgen(mu=groupmeans, gm=gm, lo=low, up=upp, names=names, alternative="less", p=pvalues)
```

ANOMintern

Internal Function

Description

Only for internal use.

Author(s)

Philip Pallmann <p.pallmann@lancaster.ac.uk>

hemoglobin

Hemoglobin Levels

Description

Hemoglobin levels of 30 male cancer patients treated with radiation or chemotherapy and one of three drugs.

Usage

```
data(hemoglobin)
```

Format

A data frame with 30 observations on the following 3 variables.

therapy A factor with 2 levels giving the types of therapy.

drug A factor with 3 levels giving the drugs administered.

level1 A numeric vector giving the patients' hemoglobin levels.

Details

This is a complete balanced two-way layout. 15 male cancer patients were radiated, and another 15 underwent chemotherapy. In addition, the patients were treated with either drug 1, 2, or 3. The endpoint of interest was the level of hemoglobin (in grams per deciliter blood).

Source

Nelson, P. R., Wludyka, P. S., Copeland, K. A. F. (2005) *The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, and American Statistical Association (ASA), Alexandria, VA, pp. 71 ff.

Examples

```
data(hemoglobin)
str(hemoglobin)
```

math

Math Proficiency Scores

Description

Proportion of fifth graders with proficient math test scores in 10 elementary schools.

Usage

```
data(math)
```

Format

A data frame with 10 observations on the following 3 variables.

`school` A factor with 10 levels giving the ID of the school.

`enrolled` A numeric vector giving the number of students taking part in the math test.

`proficient` A numeric vector giving the number of students with proficient math test scores.

Details

A study compared math achievements of students from 10 elementary schools in a U.S. district; 6 of them were conventional neighborhood schools (N1–N6) and 4 alternative schools (A1–A4). 563 fifth graders took standardized math tests, and each school’s proportion of students who scored proficient was recorded.

Source

Nelson, P. R., Wludyka, P. S., Copeland, K. A. F. (2005) *The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, and American Statistical Association (ASA), Alexandria, VA, pp. 42–43.

Examples

```
data(math)  
str(math)
```

spring

Stiffness of Springs

Description

Weights required to stretch springs of four brands by 0.1 inches.

Usage

```
data(spring)
```

Format

A data frame with 24 observations on the following 2 variables.

brand A factor with 4 levels giving the brands of springs.

weight A numeric vector giving the weight required to extend the spring by 0.1 inches.

Source

Nelson, P. R., Wludyka, P. S., Copeland, K. A. F. (2005) *The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, and American Statistical Association (ASA), Alexandria, VA, p. 53.

Examples

```
data(spring)
str(spring)
```

waterfilter

Comparison of Water Filters

Description

Filtering performances of seven brands of water filters, measured as the number of bacterial colonies growing on each device.

Usage

```
data(waterfilter)
```

Format

A data frame with 20 observations on the following 2 variables.

brand A factor with 7 levels giving the brands of water filters.

colonies A numeric vector giving the number of bacterial colonies found on each filter.

Details

A high number of bacterial colonies on a filter corresponds to good performance of this particular device. Note that the dataset is unbalanced (n=2 for brands 4 and 7, n=3 for all other brands).

Source

Hsu, J. C. (1984) Ranking and selection and multiple comparisons with the best. In: Santner, T. J. and Tamhane, A. C. (Editors) Design of Experiments: Ranking and Selection (Essays in Honor of Robert E. Bechhofer). Marcel Dekker, New York, NY, pp. 23–33.

References

Westfall, P. H., Tobias, R. D., Wolfinger, R. D. (2011) Multiple Comparisons and Multiple Tests Using SAS, Second Edition. SAS Institute Inc., Cary, NC, pp. 592–593.

Examples

```
data(waterfilter)
str(waterfilter)
```

Index

*Topic **\textasciitildedplot**

ANOM, [2](#)

ANOMgen, [7](#)

ANOMintern, [9](#)

*Topic **datasets**

hemoglobin, [10](#)

math, [11](#)

spring, [12](#)

waterfilter, [12](#)

ANOM, [2](#)

ANOMgen, [7](#)

ANOMintern, [9](#)

hemoglobin, [10](#)

math, [11](#)

spring, [12](#)

waterfilter, [12](#)