

# Package ‘dfphase1’

January 14, 2017

**Type** Package

**Title** Phase I Control Charts (with Emphasis on Distribution-Free Methods)

**Version** 1.1.1

**Author** Giovanna Capizzi and Guido Masarotto

**Maintainer** Giovanna Capizzi <giovanna.capizzi@unipd.it>

**Description** Statistical methods for retrospectively detecting changes in location and/or dispersion of univariate and multivariate variables. Data values are assumed to be independent, can be individual (one observation at each instant of time) or sub-grouped (more than one observation at each instant of time). Control limits are computed, often using a permutation approach, so that a prescribed false alarm probability is guaranteed without making any parametric assumptions on the stable (in-control) distribution.

**License** LGPL (>= 2)

**Imports** Rcpp (>= 0.11.1), lattice, robustbase

**LinkingTo** Rcpp

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2017-01-14 01:04:15

## R topics documented:

dfphase1-package . . . . .	2
changept . . . . .	3
colonscopy . . . . .	6
fe . . . . .	6
gravel . . . . .	7
mphase1 . . . . .	7
mphase1.methods . . . . .	9
mshewhart . . . . .	10
phase1Plot . . . . .	13
rsp . . . . .	14
ryan . . . . .	15

shewhart . . . . .	16
Student . . . . .	18

<b>Index</b>	<b>20</b>
--------------	-----------

---

dfphase1-package	<i>Phase I Control Charts (with Emphasis on Distribution-Free Methods)</i>
------------------	--

---

## Description

Statistical methods for retrospectively detecting changes in location and/or dispersion of univariate and multivariate variables. Data values are assumed to be independent, can be individual (one observation at each instant of time) or subgrouped (more than one observation at each instant of time). Control limits are computed, often using a permutation approach, so that a prescribed false alarm probability is guaranteed without making any parametric assumptions on the stable (in-control) distribution.

## Details

The main functions are:

- `shewhart` and `mshewhart`: univariate and multivariate Shewhart-type control charts based either on the original observations or on a rank transformation. These functions are particularly useful for detecting isolated shifts in the mean and/or variance of subgrouped observations. Functions `shewhart` and `mshewhart` also allow the simultaneous use of two control charts originally designed to detect separately location and scale shifts. In particular, note that when more than one critical values are needed, the false alarm probability is “balanced” between the separate control charts as discussed by Capizzi (2015).
- `changeoint` and `mchangeoint`: univariate or multivariate control charts useful for detecting sustained (and other patterned) mean and/or variance shifts. The control statistic is based on a generalized likelihood ratio test computed under a Gaussian assumption. However, the control limits are computed by permutation. An optional preliminary rank transformation can be used to improve the performance in the case of nonnormal data.
- `rsp` and `mphase1`: the univariate and multivariate methods introduced by Capizzi and Masarotto (2013) and (2016) to detect multiple isolated or step shifts in individual or subgrouped data.

The use of distribution-free control limits is emphasized. However, the package also includes some functions for computing normal-based control limits. As noted in the individual help pages, these limits can also be suitable for some non-normal distributions (e.g., applying a multivariate rank-transformation, normal-based control limits maintain the desired false alarm probability in the class of the multivariate elliptical distributions). Nevertheless, their use is not generally recommended.

The data should be organized as follows:

- *Univariate control charts*: an  $n \times m$  matrix, where  $n$  and  $m$  are the size of each subgroup and the number of subgroups, respectively. A vector of length  $m$  is accepted in the case of individual data, i.e., when  $n=1$ .
- *Multivariate control charts*: a  $p \times n \times m$  array, where  $p$  denotes the number of monitored variables. A  $p \times m$  matrix is accepted in the case of individual data.

Functions `phase1Plot` and `mphase1Plot` can be used for plotting the data.

**Author(s)**

Giovanna Capizzi and Guido Masarotto (maintainer: Giovanna Capizzi <giovanna.capizzi@unipd.it>).

**References**

G. Capizzi (2015) “Recent advances in process monitoring: Nonparametric and variable-selection methods for Phase I and Phase II (with discussion)”. *Quality Engineering*, **27**, pp. 44–80.

G. Capizzi and G. Masarotto (2013), “Phase I Distribution-Free Analysis of Univariate Data”. *Journal of Quality Technology*, **45**, pp. 273-284.

G. Capizzi and G. Masarotto (2016), “Phase I Distribution-Free Analysis with the R Package dfphase1”. To appear on *Frontiers in Statistical Quality Control 12*, eds. S. Knoth and W. Schmid, Springer.

G. Capizzi and G. Masarotto (2016), Phase I Distribution-Free Analysis of Multivariate Data, *Technometrics*, to appear, <http://dx.doi.org/10.1080/00401706.2016.1272494>.

**See Also**

[shewhart](#), [shewhart.normal.limits](#), [mshewhart](#), [mshewhart.normal.limits](#), [changepoint](#), [changepoint.normal.limits](#), [mchangepoint](#), [mchangepoint.normal.limits](#), [rsp](#), [phase1Plot](#), [mphase1Plot](#).

---

changepoint

*Detection of a sustained change-point in univariate and multivariate data*

---

**Description**

changepoint (univariate data) and mchangepoint (multivariate data) test for the presence of a sustained location and/or dispersion shift. Both functions can be applied to individual and subgrouped observations.

changepoint.normal.limits and mchangepoint.normal.limits precompute the corresponding control limits when the in-control distribution is normal.

**Usage**

```
changepoint(x, subset, score = c("Identity", "Ranks"), only.mean = FALSE,
  plot = TRUE, FAP = 0.05, seed = 11642257, L = 10000, limits = NA)
```

```
mchangepoint(x, subset, score = c("Identity", "Signed Ranks", "Spatial Signs",
  "Spatial Ranks", "Marginal Ranks"), only.mean = FALSE,
  plot = TRUE, FAP = 0.05, seed = 11642257, L = 10000, limits = NA)
```

```
changepoint.normal.limits(n, m, score = c("Identity", "Ranks"),
  only.mean = FALSE, FAP = 0.05, seed = 11642257, L = 100000)
```

```
mchangepoint.normal.limits(p, n, m, score = c("Identity", "Signed Ranks", "Spatial Signs",
```

"Spatial Ranks", "Marginal Ranks"), only.mean = FALSE,  
FAP = 0.05, seed = 11642257, L = 100000)

### Arguments

x	changepoint: a nxm numeric matrix or a numeric vector of length m. mchangeoint: a pxnxm data numeric array or a pxm numeric vector. See below, for the meaning of p, n and m.
p	integer: number of monitored variables.
n	integer: size of each subgroup (number of observations gathered at each time point).
m	integer: number of subgroups (time points).
subset	an optional vector specifying a subset of subgroups/time points to be used
score	character: the transformation to use; see <a href="#">mshewhart</a> .
only.mean	logical; if TRUE only a location change-point is searched.
plot	logical; if TRUE, the control statistic is displayed.
FAP	numeric (between 0 and 1): the desired false alarm probability.
seed	positive integer; if not NA, the RNG's state is resetted using seed. The current .Random.seed will be preserved. Unused by <a href="#">mshewhart</a> when limits is not NA.
L	positive integer: the number of Monte Carlo replications used to compute the control limits. Unused by <a href="#">changeoint</a> and <a href="#">mchangeoint</a> when limits is not NA.
limits	numeric: a precomputed vector of length m containing the control limits.

### Details

After an optional rank transformation (argument `score`), `changeoint` and `mchangeoint` compute, for  $\tau = 2, \dots, m$ , the normal likelihood ratio test statistics for verifying whether the mean and dispersion (or only the mean when `only.mean=TRUE`) are the same before and after  $\tau$ . See Sullivan and Woodall (1999, 2000) and Qiu (2013), Chapter 6 and Section 7.5.

Note that the control statistic is equivalent to that proposed by Lung-Yut-Fong *et al.* (2011) when `score="Marginal Ranks"` and `only.mean=TRUE`.

As suggested by Sullivan and Woodall (1999, 2000), control limits proportional to the in-control mean of the likelihood ratio test statistics are used. Further, when `plot=TRUE`, the control statistics divided by the time-varying control limits are plotted with a "pseudo-limit" equal to one.

When `only.mean=FALSE`, the decomposition of the likelihood ratio test statistic suggested by Sullivan and Woodall (1999, 2000) for diagnostic purposes is also computed, and optionally plotted.

### Value

`changeoint` and `mchangeoint` return an invisible list with elements

`glr` control statistics.

mean, dispersion

decomposition of the control statistics in the two parts due to changes in the mean and dispersion, respectively. These elements are present only when `only.mean=FALSE`.

limits control limits.

score, `only.mean`, FAP, L, seed  
input arguments.

`changepoint.normal.limits` and `mchangepoint.normal.limits` return a numeric vector containing the control limits.

### Note

1. When `limits` is NA, `changepoint` and `mchangepoint` compute the control limits by permutation. The resulting control charts are distribution-free.
2. Pre-computed limits, like those computed using `changepoint.normal.limits` and `mchangepoint.normal.limits`, are recommended *only* for univariate data when `score=Ranks`. Indeed, in all the other cases, the resulting control chart will not be distribution-free.
3. However, note that, when `score` is Signed Ranks, Spatial Signs, Spatial Ranks the normal-based control limits are distribution-free in the class of all multivariate elliptical distributions.

### Author(s)

Giovanna Capizzi and Guido Masarotto.

### References

- A. Lung-Yut-Fong, C. Lévy-Leduc, O. Cappé O (2011) “Homogeneity and change-point detection tests for multivariate data using rank statistics”. arXiv preprint arXiv:11071971.
- P. Qiu (2013) *Introduction to Statistical Process Control*. Chapman & Hall/CRC Press.
- J. H. Sullivan, W. H. Woodall (1996) “A control chart for preliminary analysis of individual observations”. *Journal of Quality Technology*, **28**, pp. 265–278.
- J. H. Sullivan, W. H. Woodall (2000) “Change-point detection of mean vector or covariance matrix shifts using multivariate individual observations”. *IIE Transactions* **32**, pp. 537–549.

### Examples

```
data(gravel)
changepoint(gravel[1,,])
mchangepoint(gravel)
mchangepoint(gravel,score="Signed Ranks")
```

---

colonscopy

*Colonscopy Times*

---

### Description

This data set contains the colonscopy times (minutes) for 30 subgroups of 5 patients given in Allison Jones-Farmer *et al.* (2009).

### Usage

```
data(colonscopy)
```

### Format

A 5x30 matrix.

### References

L. A. Jones-Farmer, V. Jordan, C. W. Champs (2009) “Distribution-free Phase I control charts for subgroup location”, *Journal of Quality Technology*, **41**, pp. 304–316.

### Examples

```
data(colonscopy)
phase1Plot(colonscopy)
```

---

fe

*Ferric Oxide data*

---

### Description

This data set contains 189 ferric-oxide individual measurement collected in an aluminum smelter.

### Usage

```
data(fe)
```

### Format

A vector of length 189.

### References

M. D. Holland, D. M. Hawkins (2014) “A Control Chart Based on a Nonparametric Multivariate Change-Point Model”, *Journal of Quality Technology*, **46**, pp 63–77.

**Examples**

```
data(fe)
phase1Plot(fe)
```

---

gravel	<i>Gravel data</i>
--------	--------------------

---

**Description**

This data set contains 56 individual bivariate observations from a gravel-producing plant given by Holmes and Mergen (1993). There are two variables measuring the percentage of the particles (by weight) that are large or medium in size, respectively.

**Usage**

```
data(gravel)
```

**Format**

A 2x56 matrix.

**References**

D. S. Holmes, A. Mergen (1993) “Improving the Performance of the  $T^2$  Control Chart”, *Quality Engineering*, **5**, pp. 619–625.

**Examples**

```
data(gravel)
mphase1Plot(gravel)
```

---

mphase1	<i>Distribution-free Phase I analysis of multivariate data</i>
---------	--

---

**Description**

Retrospective change point detection using the method described by Capizzi and Masarotto (2016).

**Usage**

```
mphase1(x, plot = TRUE, post.signal = TRUE, isolated = dim(x)[2] > 1, step = TRUE,
alpha = 0.05, gamma = 0.5, K = min(50, round(sqrt(dim(x)[3]))),
lmin = 5, L = 1000, seed = 11642257)
```

**Arguments**

x	a pxnxm array containing the observations; $x[r, j, i]$ is the $j$ th observation on the $r$ th variable of the $i$ th subgroup.
plot	logical; if FALSE the diagnostic plot is not displayed.
post.signal	logical; if FALSE the diagnostic LASSO-based analysis is not performed.
isolated	logical; if FALSE isolated shifts are not detected.
step	logical; if FALSE step shifts are not detected.
alpha	real; the acceptable false alarm probability; if the observed p-value is greater than alpha, then the estimated mean function is a constant.
gamma	real; the extra penalization for the extended BIC criteria.
K	integer; the maximum number of shifts which the procedure tries to detect.
lmin	integer; the minimum length of a step shift.
L	integer; the number of random permutations used to compute the p-values.
seed	integer; if not NA, the RNG's state is re-setted using seed. The current .Random.seed will be preserved.

**Value**

Functions mphase1 returns an object of class mphase1 containing

p.value	The p-value.
Wobs	The overall test statistic.
alasso	A data-frame containing the result of the post-signal diagnosis analysis, i.e., the times and types of shifts and the involved variables identified using the adaptive LASSO.
forward	A data frame containing the result of the forward search analysis, i.e., the times and types of the possible shifts as well as the elementary test statistics and the estimates of their (conditional) means and standard deviations.
center, scatter	The location vector and dispersion matrix used to standardize the original data.
signed.ranks	A pxnxm array containing the signed ranks.
fitted, residuals	Two pxnxm arrays containing the fitted means and the residuals, i.e., the difference between the observations and the fitted values.

**Author(s)**

Giovanna Capizzi and Guido Masarotto.

**References**

G. Capizzi and G. Masarotto (2016), Phase I Distribution-Free Analysis of Multivariate Data, *Technometrics*, to appear, <http://dx.doi.org/10.1080/00401706.2016.1272494>.



**See Also**[postsignal.](#)**Examples**

```
# A simulated in-control data from a Student's t distribution
# with 3 degrees of freedom
set.seed(123)
x <- sweep(array(rnorm(5*5*50),c(5,5,50)),c(2,3),sqrt(rchisq(5*50,3)/3),"/")
mphase1(x)
# Reproduction of the two examples given in Capizzi and Masarotto (2016)
data(ryan)
mphase1(ryan)
data(gravel)
mphase1(gravel)
```

mphase1.methods

*Methods for objects of class mphase1***Description**

Methods `print` and `plot` allow to write to the console and plot (optionally changing the layout) the result of the Phase I analysis performed with function `mphase1`.

Method `postsignal` implements the post-signal Phase I analysis based on the adaptive LASSO described in Capizzi and Masarotto (2016). It uses the p-value and the results on the forward search contained in its first argument. Hence, it is useful for re-running the analysis with different values of  $\alpha$  and/or  $\gamma$ .

**Usage**

```
## S3 method for class 'mphase1'
print(x,...)
## S3 method for class 'mphase1'
plot(x,layout=c(1,p),...)
## S3 method for class 'mphase1'
postsignal(x, plot = TRUE, alpha = 0.05, gamma = 0.5,...)
```

**Arguments**

<code>x</code>	an object returned by function <a href="#">mphase1</a> .
<code>layout</code>	an integer vector describing the multi-panel (and possible multi-page) layout.
<code>plot</code>	logical; if TRUE the diagnostic plot is displayed.
<code>alpha</code>	real; the acceptable false alarm probability; if the observed p-value is greater than $\alpha$ , then the estimated mean function is a constant.
<code>gamma</code>	real; the extra penalization for the extended BIC criteria.
<code>...</code>	ignored.

**Value**

An object of class `mphase1`. See `mphase1` for the description.

**Author(s)**

Giovanna Capizzi and Guido Masarotto.

**References**

G. Capizzi and G. Masarotto (2016), Phase I Distribution-Free Analysis of Multivariate Data, *Technometrics*, to appear, <http://dx.doi.org/10.1080/00401706.2016.1272494>.

**Examples**

```
data(gravel)
u <- mphase1(gravel,plot=FALSE)
print(u)
plot(u,layout=c(2,1))
postsignal(u,plot=FALSE,gamma=1)
```

---

mshewhart

*Multivariate Shewhart-type control charts*


---

**Description**

`mshewhart` computes, and, optionally, plots, several Shewhart-type Phase I control charts for detecting location and scale changes in multivariate subgrouped data.

`mshewhart.normal.limits` pre-computes the corresponding control limits when the in-control distribution is multivariate normal.

**Usage**

```
mshewhart(x, subset, stat = c("T2Var", "T2", "Var", "Depth Ranks"), score = c("Identity",
  "Signed Ranks", "Spatial Signs", "Spatial Ranks", "Marginal Ranks"),
  loc.scatter = c("Classic", "MCD"), plot = TRUE, FAP = 0.05,
  seed = 11642257, L = 1000, limits = NA)
```

```
mshewhart.normal.limits(p, n, m, stat = c("T2Var", "T2", "Var", "Depth Ranks"),
  score = c("Identity", "Signed Ranks", "Spatial Signs", "Spatial Ranks",
  "Marginal Ranks"), loc.scatter = c("Classic", "MCD"),
  FAP = 0.05, seed = 11642257, L = 100000)
```

**Arguments**

<code>x</code>	a $p \times n \times m$ data numeric array (n observations gathered at m time points on p variables).
<code>p</code>	integer: number of monitored variables.
<code>n</code>	integer: size of each subgroup (number of observations gathered at each time point).
<code>m</code>	integer: number of subgroups (time points).
<code>subset</code>	an optional vector specifying a subset of subgroups/time points to be used
<code>stat</code>	character: control statistic[s] to use; see Details.
<code>score</code>	character: transformation to use; unused when <code>stat=Depth Ranks</code> ; see Details.
<code>loc.scatter</code>	character: estimates of the multivariate location and scatter to use when no preliminary rank transformation is applied. Unused when <code>stat</code> is equal to <code>Depth Ranks</code> or <code>score</code> is different from <code>Identity</code> . See Details.
<code>plot</code>	logical; if TRUE, control statistic[s] is[are] displayed.
<code>FAP</code>	numeric (between 0 and 1): desired false alarm probability.
<code>seed</code>	positive integer; if not NA, the RNG's state is resetted using <code>seed</code> . The current <code>.Random.seed</code> will be preserved. Unused by <code>mshewhart</code> when <code>limits</code> is not NA.
<code>L</code>	positive integer: number of Monte Carlo replications used to compute the control limits. Unused by <code>mshewhart</code> when <code>limits</code> is not NA.
<code>limits</code>	numeric: pre-computed vector of control limits. This vector should contain $(A, B)$ when <code>stat=T2Var</code> , $(A)$ when <code>stat=T2</code> , $(B)$ when <code>stat=Var</code> and $(C)$ when <code>stat=Depth Ranks</code> . See Details for the definition of the critical values $A$ , $B$ and $C$ .

**Details**

The implemented control statistics are

- `T2Var`: combination of the `T2` and `Var` statistics described below.
- `T2`: Hotelling's  $T^2$  control statistics (see Montgomery, 2009, equation 11.19, or Qiu, 2013, equation 7.7) with control limit equal to  $A$ .
- `Var`: normal likelihood ratio control statistics for detecting changes in the multivariate dispersion (see Montgomery, 2009, equation 11.34), with control limit equal to  $B$ .
- `Depth Ranks`: control statistics based on the rank of the Mahalanobis depths, proposed by Bell *et al.*. As suggested Bell *et al.*, the Mahalanobis depths are computed using the BACON estimates of the multivariate mean vector and the mean of the subgroups sample covariance matrices. An alarm is signalled if any of the statistics is greater than a positive control limit  $C$ .

The `T2` and `Var` control statistics are computed

- `score=Identical`: from the original data standardized using either the classical pooled estimates of the mean vector and dispersion matrix (Montgomery, 2009, equations 11.14–11.18; Qiu, 2013, equations at page 269) or the highly robust minimum covariance determinant (MCD) estimate when argument `loc.scatter` is equal to `Classic` or `MCD`, respectively.

- score=Signed Ranks, Spatial Signs, Spatial Ranks, Marginal Ranks: from a “rank” transformation of the original data. In particular, see Hallin and Paindaveine (2005) for the definition of the multivariate signed ranks and Oja (2010) for those of the spatial signs, spatial ranks, and marginal ranks. Multivariate signed ranks, spatial signs and ranks are “inner” standardized while marginal ranks are “outer” standardized (see Oja (2010) for the definition of “inner” and “outer” standardization).

Note that the  $T^2$  control statistics based on the spatial signs corresponds to the control charts suggested by Cheng and Shiau (2015).

### Value

mshewhart returns an invisible list with elements:

T2	$T^2$ control statistic; this element is present only if stat is T2Var or T2.
Var	Var control statistic; this element is present only if stat is T2Var or Var.
DepthRanks	control statistic based on the rank of the Mahalanobis depths; this element is present only if stat is Depth Ranks.
center, scatter	estimates of the multivariate location and scatter used to standardized the observations.
limits	control limits.
stat, score, loc.scatter, FAP, L, seed	input arguments.

mshewhart.normal.limits returns a numeric vector containing the control limits.

### Note

1. When limits is NA, mshewhart computes the control limits by permutation. Then, the resulting control chart is distribution-free.
2. Pre-computed limits, such as those computed by using mshewhart.normal.limits, are not recommended. Indeed, the resulting control chart will not be distribution-free.
3. However, when score is Signed Ranks, Spatial Signs, Spatial Ranks or stat is Depth Ranks, the computed control limits are distribution-free in the class of all multivariate elliptical distributions.

### Author(s)

Giovanna Capizzi and Guido Masarotto.

### References

- R. C. Bell, L. A. Jones-Farmer, N. Billor (2014) “A distribution-free multivariate Phase I location control chart for subgrouped data from elliptical distributions”. *Technometrics*, **56**, pp. 528–538.
- C. R. Cheng, J. J. H. Shiau JJH (2015) “A distribution-free multivariate control chart for Phase I applications”. *Quality and Reliability Engineering International*, **31**, pp. 97–111.
- M. Hallin and D. Paindaveine (2005) “Affine-Invariant Aligned Rank Tests for the Multivariate General Linear Model with VARMA Errors”. *Journal of Multivariate Analysis*, **93**, pp. 122–163.

D. C. Montgomery (2009) *Introduction to Statistical Quality Control*, 6th edn. Wiley.

H. Oja (2010) *Multivariate Nonparametric Methods with R. An Approach Based on Spatial Signs and Ranks*. Springer.

P. Qiu (2013) *Introduction to Statistical Process Control*. Chapman & Hall/CRC Press.

## Examples

```
data(ryan)
mshewhart(ryan)
mshewhart(ryan, subset=-10)
mshewhart(ryan, subset=-c(10, 20))
mshewhart(ryan, score="Signed Ranks")
mshewhart(ryan, subset=-10, score="Signed Ranks")
mshewhart(ryan, subset=-c(10, 20), score="Signed Ranks")
```

---

phase1Plot

*Plot of Phase 1 data*

---

## Description

phase1Plot and mphase1Plot plot univariate or multivariate Phase 1 observations, organized as required by the dfphase1 package.

## Usage

```
phase1Plot(x)

mphase1Plot(x, layout = c(1, p))
```

## Arguments

**x** phase1Plot: a  $n \times m$  numeric matrix or a numeric vector of length  $m$ .  
 mphase1Plot: a  $p \times n \times m$  data numeric array or a  $p \times m$  numeric matrix.  
 Here,  $p$  denotes the number of variables,  $n$  the size of each subgroup and  $m$  the number of subgroups.

**layout** an integer vector describing the multi-panel (and possible multi-page) layout. See the third example below.

## Author(s)

Giovanna Capizzi and Guido Masarotto.

**Examples**

```
x <- matrix(rt(5*20,5),5)
x[,10] <- x[,10]+3
phase1Plot(x)
# a data set with many variables
x <- array(rnorm(20*5*50),c(20,5,50))+10*(1:20)
mphase1Plot(x)
# it is better to organize the plot on two pages
if (interactive()) old <- grDevices::devAskNewPage(TRUE)
mphase1Plot(x,c(2,5,2))
if (interactive()) grDevices::devAskNewPage(old)
```

rsp

*Distribution-Free Phase I Analysis of Univariate Data based on Recursive Segmentation and Permutation*

**Description**

rsp implements the Phase I method described in Capizzi and Masarotto (2013).

**Usage**

```
rsp(y, plot = TRUE, L = 1000, seed = 11642257, alpha = 0.05,
    maxsteps = min(50, round(NROW(y)/15)), lmin = max(5, min(10, round(NROW(y)/10))))
```

**Arguments**

y	Phase I data; y can be either (i) a vector or a 1xm matrix in the case of individual observations or (ii) a nxm matrix for subgrouped data (n observations gathered at m time points).
plot	logical; if TRUE, the diagnostic plot is displayed.
L	integer; the number of random permutations used to compute the p-values.
seed	positive integer; if not NA, the RNG's state is resetted using seed. The current .Random.seed will be preserved.
alpha	real; the significance level used to compute the level and scale estimates; if one of the p-values is greater than alpha, the corresponding estimate is a constant.
maxsteps	integer; the maximum number of step shifts which the procedure tries to detect.
lmin	integer; the minimum length of a step.

**Value**

A list with elements

p	the adjusted p-values
stat	the summary statistics (a mx2 matrix)
fitted	the (possibly time-variant) estimates of the process level and scale (a mx2 matrix).

**Author(s)**

Giovanna Capizzi and Guido Masarotto.

**References**

G. Capizzi, G. Masarotto (2013), “Phase I Distribution-Free Analysis of Univariate Data”. *Journal of Quality Technology*, **45**, pp. 273-284.

**Examples**

```
# Individual observations with a transient level change
set.seed(112233)
level <- c(rep(0,20),rep(3,10),rep(0,20))
x <- level+rt(50,4)
rsp(x)
# Individual observations with a scale step change
scale <- c(rep(1,25),rep(3,25))
x <- scale*rt(50,4)
rsp(x)
data(fe)
rsp(fe)
data(colonscopy)
rsp(colonscopy)
```

---

ryan

*Ryan data*

---

**Description**

This data set contains the data given in Table 9.2 by Ryan (2011, p. 323). The sample comprises 20 subgroups, each with 4 observations, on two quality characteristics  $X_1$  and  $X_2$ . According to Ryan (2011), the 10th and 20th subgroups are out-of-control.

**Usage**

```
data(ryan)
```

**Format**

A 2x4x20 array.

**References**

T. P. Ryan (2011), *Statistical Methods for Quality Improvement*, 3rd ed., Wiley.

**Examples**

```
data(ryan)
mphase1Plot(ryan)
```

shewhart

*Univariate Shewhart-type control charts***Description**

shewhart computes, and, optionally, plots, Shewhart-type Phase I control charts for detecting changes in location and scale of univariate subgrouped data.

shewhart.normal.limits pre-computes the corresponding control limits when the in-control distribution is normal.

**Usage**

```
shewhart(x, subset, stat = c("XbarS", "Xbar", "S", "Rank", "lRank", "sRank"),
  aggregation = c("median", "mean"), plot = TRUE, FAP = 0.05,
  seed = 11642257, L = if (stat %in% c("XbarS", "Xbar", "S")) 1000 else 100000,
  limits = NA)
```

```
shewhart.normal.limits(n, m, stat = c("XbarS", "Xbar", "S", "Rank", "lRank", "sRank"),
  aggregation = c("median", "mean"), FAP = 0.05,
  seed = 11642257, L = 100000)
```

**Arguments**

x	a nxm data numeric matrix (n observations gathered at m time points).
n	integer: size of each subgroup (number of observations gathered at each time point).
m	integer: number of subgroups (time points).
subset	an optional vector specifying a subset of subgroups/time points to be used
stat	character: the control statistic[s] to use; see Details.
aggregation	character: it specify how to aggregate the subgroup means and standard deviations. Used only when stat is XbarS, Xbar or S.
plot	logical; if TRUE, control statistic[s] is[are] displayed.
FAP	numeric (between 0 and 1): desired false alarm probability.
seed	positive integer; if not NA, the RNG's state is resetted using seed. The current .Random.seed will be preserved. Unused by shewhart when limits is not NA.
L	positive integer: number of Monte Carlo replications used to compute the control limits. Unused by shewhart when limits is not NA.
limits	numeric: a precomputed vector of control limits. The vector should contain $(A, B_1, B_2)$ when $stat=XbarS$ , $(A)$ when $stat=Xbar$ , $(B_1, B_2)$ when $stat=S$ , $(C, D)$ when $stat=Rank$ , $(C)$ when $stat=lRank$ and $(D)$ when $stat=sRank$ . See Details for the definition of the critical values $A, B_1, B_2, C$ and $D$ .



## Details

The implemented control charts are:

- XbarS: combination of the Xbar and S control charts described in the following.
- Xbar: chart based on plotting the subgroup means with control limits

$$\hat{\mu} \pm A \frac{\hat{\sigma}}{\sqrt{n}}$$

where  $\hat{\mu}$  ( $\hat{\sigma}$ ) denotes the estimate of the in-control mean (standard deviation) computed as the mean or median of the subgroup means (standard deviations).

- S: chart based on plotting the (unbiased) subgroup standard deviations with lower control limit  $B_1 \hat{\sigma}$  and upper control limit  $B_2 \hat{\sigma}$ .
- Rank: combination of the lRank and sRank control charts described in the following.
- lRank: control chart based on the standardized rank-sum control statistic suggested by Jones-Farmer *et al.* (2009) for detecting changes in the location parameter. Control limits are of the type  $\pm C$ .
- sRank: control chart based on the standardized rank-sum control statistic suggested by Jones-Farmer and Champ (2010) for detecting changes in the scale parameter. Control limits are of the type  $\pm D$ .

## Value

shewhart returns an invisible list with elements

Xbar	subgroup means; this element is present only if stat is XbarS or Xbar.
S	subgroup standard deviation; this element is present only if stat is XbarS or S.
lRank	rank-based control statistics for detecting changes in location; this element is present only if stat is Rank or lRank.
sRank	rank-based control-statistics for detecting changes in scale; this element is present only if stat is Rank or sRank.
limits	control limits.
center, scale	estimates $\hat{\mu}$ and $\hat{\sigma}$ of the in-control mean and standard deviation; these elements are present only if stat is XbarS, Xbar and S.
stat, L, aggregation, FAP, seed	input arguments.

shewhart.normal.limits returns a numeric vector containing the limits.

## Note

1. If argument limits is NA, shewhart computes the control limits using (i) the permutation approach if stat is Xbars, Xbar, S and (ii) the distribution-free unconditional distribution when stat is Rank, lRank and sRank. In both cases, the resulting control chart is distribution-free.

2. Pre-computed limits, such as those computed using `shewhart.normal.limits`, are not recommended when `stat` is `XbarS`, `Xbar` or `S`. Indeed, the resulting control chart will not be distribution-free.
3. When `stat` is `lRank`, `lRank` or `sRank`, the control limits computed by `shewhart.normal.limits` are distribution-free in the class of all univariate continuous distributions. So, if user plan to apply rank-based control charts on a repeated number of samples of the same size, pre-computing the control limits using `mshewhart.normal.limits` can reduce the overall computing time.

### Author(s)

Giovanna Capizzi and Guido Masarotto.

### References

- L. A. Jones-Farmer, V. Jordan, C. W. Champs (2009) “Distribution-free Phase I control charts for subgroup location”, *Journal of Quality Technology*, **41**, pp. 304–316.
- L. A. Jones-Farmer, C. W. Champ (2010) “A distribution-free Phase I control chart for subgroup scale”. *Journal of Quality Technology*, **42**, pp. 373–387.

### Examples

```
# A simulated example
set.seed(12345)
y <- matrix(rt(100,3),5)
y[,20] <- y[,20]+3
shewhart(y)
# Reproduction of the control charts shown
# by Jones-Farmer et. al. (2009,2010)
data(colonscopy)
u <- shewhart.normal.limits(NROW(colonscopy),NCOL(colonscopy),stat="lRank",FAP=0.1)
u
# control limits based on a limited number of replications
# to avoid a (relatively) long execution time
shewhart(colonscopy,stat="lRank",limits=u,L=10000)
shewhart(colonscopy,stat="sRank",FAP=0.1,L=10000)
```

---

Student

*A simulated dataset*

---

### Description

This simulated data set consists in 50 subgroups, each with 5 observations, on 4 variables.

There is an isolated location shift involving only the first variable at time  $t = 10$  and a step shift, involving the third and fourth variables, starting from  $t = 31$ . The in-control distribution is Student's  $t$  with 3 degrees of freedom, zero mean and such that  $cov(X_i, X_j) = 0.8^{|i-j|}$ .

See the example for the exact code used to simulate the data.

**Usage**

```
data(Student)
```

**Format**

A 4x5x50 array.

**Examples**

```
data(Student)
mphase1(Student)
#
# Replication of the simulation
#
# Generation of the in-control observations
set.seed(1)
m <- 50
n <- 5
p <- 4
df <- 3
Sigma <- outer(1:p,1:p,function(i,j) 0.8^abs(i-j))
Sigma
xnorm <- crossprod(chol(Sigma),matrix(rnorm(p*n*m),p))
xchisq <- sqrt(rchisq(n*m,df)/(df-2))
x <- array(sweep(xnorm,2,xchisq,"/"),c(p,n,m))
# Then, we add an isolated shift at time 10
# (only for the first variable)
x[1,,10] <- x[1,,10]+1
# and, a step shift starting at time 31
# (only for the third and fourth variable)
x[3:4,,31:50] <- x[3:4,,31:50] + c(0.50,-0.25)
dimnames(x)<-list(paste("X",1:4,sep=""),NULL,NULL)
identical(x,Student)
```

# Index

## \*Topic **datasets**

colonscopy, 6  
fe, 6  
gravel, 7  
ryan, 15  
Student, 18

## \*Topic **distribution-free methods**

changepoint, 3  
mphase1, 7  
mshewhart, 10  
rsp, 14  
shewhart, 16

## \*Topic **hplot**

phase1Plot, 13

## \*Topic **htest**

changepoint, 3  
mphase1, 7  
mshewhart, 10  
rsp, 14  
shewhart, 16

## \*Topic **multivariate**

changepoint, 3  
mphase1, 7  
mshewhart, 10

## \*Topic **package**

dfphase1-package, 2

## \*Topic **statistical process control**

changepoint, 3  
mphase1, 7  
mshewhart, 10  
rsp, 14  
shewhart, 16

## \*Topic **statistical process monitoring**

changepoint, 3  
mphase1, 7  
mshewhart, 10  
rsp, 14  
shewhart, 16

changepoint, 2, 3, 3

changepoint.normal.limits, 3  
colonscopy, 6

dfphase1 (dfphase1-package), 2  
dfphase1-package, 2

fe, 6

gravel, 7

mchangepoint, 2, 3  
mchangepoint (changepoint), 3  
mchangepoint.normal.limits, 3  
mphase1, 2, 7, 9, 10  
mphase1.methods, 9  
mphase1Plot, 2, 3  
mphase1Plot (phase1Plot), 13  
mshewhart, 2-4, 10  
mshewhart.normal.limits, 3

phase1Plot, 2, 3, 13  
plot.mphase1 (mphase1.methods), 9  
postsignal, 9  
postsignal (mphase1.methods), 9  
print.mphase1 (mphase1.methods), 9

rsp, 2, 3, 14  
ryan, 15

shewhart, 2, 3, 16  
shewhart.normal.limits, 3  
Student, 18