

Package ‘gofMC’

December 8, 2016

Title Goodness of Fit Noise Analysis Using Monte Carlo Techniques

Version 1.1.2

Date 2016-12-08

Author Joseph G Kreke, PhD; Harris, Inc.
Sangeet Khemlani, PhD; Naval Research Laboratory.
Greg Trafton, PhD; Naval Research Laboratory.

Maintainer Joseph G Kreke <jkreke2@gmail.com>

Description Goodness-of-fit metrics, such as R-Squared, RMSE, etc., share a sensitivity to noise, dependent on the degrees of freedom. Some metrics, such as R-Squared, decrease with increasing dof and some, such as RMSE, increase with increasing dof. This package calculates the noise baseline (ceiling) by random sampling, calculating the metric’s value for each sample and counting the number of samples below a desired level, 95% by default. If one’s measure is above (below) the calculation corresponding to the desired level, then the measurement is distinguishable from noise. In addition, the ratio of the measurement to the calculated level provides a way to compare measurements of different degrees of freedom.

Depends R (>= 3.3.1), ggplot2, grDevices, scales, stats

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 5.0.1.9000

NeedsCompilation no

Repository CRAN

Date/Publication 2016-12-08 15:29:56

R topics documented:

difn	2
fit	3
fitEquiv	3
fitmetric_check	4
fitNoise	5

KS_D	5
KS_Dm	6
KS_Dp	7
pcdfs	7
plotcdf	8
plotConstNoise	9
plotConstValue	9
plotNoiseLevel	10
plotpdf	11
R2	11
rmse	12
Table_dofbypct	13
Table_pctbyfuncs	13
utrend	14

Index	15
--------------	-----------

difn	<i>Find differences</i>
------	-------------------------

Description

Finds the difference between two vector components, n places apart. Like diff(), but faster and pads missing values with 0.

Usage

```
difn(x, n = 1)
```

Arguments

x	a numeric vector
n	an integer, pos or neg

Value

numeric vector

Examples

```
difn(c(1,2,4,7,12),n=2)
```

fit	<i>Print boilerplate stats</i>
-----	--------------------------------

Description

Prints a standard set of stats for a given degree of freedom, fitmetric and noise distribution

Usage

```
fit(measured_value, dof, pct = 0.95, fitmetric = R2, order = 5,
    ndecimals = 2, dist = rnorm, table = TRUE, ...)
```

Arguments

measured_value	a real number within the range of fitmetric
dof	an integer
pct	a real number between 0 and 1
fitmetric	a character string
order	a real number
ndecimals	an integer
dist	a random number distribution function
table	a logical value
...	any argument that functions within this routine might use

Value

data frame

Examples

```
fit(6, dof=10, pct=0.99)
```

fitEquiv	<i>Fit Equivalent</i>
----------	-----------------------

Description

Rescales the fit value to distance to the threshold

Usage

```
fitEquiv(measured_value, dof, pct = 0.95, ndecimals = 2, fitmetric = R2,
    trend = NULL, ...)
```

Arguments

measured_value	a real number
dof	an integer
pct	a real number between 0 and 1
ndecimals	an integer
fitmetric	a character string naming a standard fit metric (R2, rmse, or user-defined function)
trend	a character string "Increasing", "Decreasing", "Flat", "Uncertain" which describes the general slope of the fitmetric function.
...	any argument that functions within this routine might use

Value

a real number

Examples

```
fitEquiv(0.8, 6)
fitEquiv(0.1, dof=8, fitmetric=rmse)
```

fitmetric_check	<i>Convert fitmetric arguments to matrices.</i>
-----------------	---

Description

Checks classes of fitmetric arguments and converts them to matrices if they are not already.

Usage

```
fitmetric_check(x, y)
```

Arguments

x	a vector of class numeric or integer or a numeric matrix
y	a vector of class numeric or integer or a numeric matrix

Value

a list containing x, y, N(number of rows), and dof(number of columns)

Examples

```
vlist <- fitmetric_check(runif(100000),runif(100000))
mlist <- fitmetric_check(matrix(rnorm(10000),ncol=10),matrix(rnorm(10000),ncol=10))
```

fitNoise	<i>Find The Threshold Noise Level</i>
----------	---------------------------------------

Description

Finds the threshold noise level for a certain fit metric

Usage

```
fitNoise(dof, pct = 0.95, ndecimals = 2, fitmetric = R2, dist = rnorm,
trend = NULL, ...)
```

Arguments

dof	an integer
pct	a real number between 0 and 1
ndecimals	an integer
fitmetric	a character string naming a standard fit metric ("R2", "rmse", or "user")
dist	a random number distribution function
trend	a character string "Increasing", "Decreasing", "Flat", "Uncertain" which describes the general slope of the fitmetric function.
...	any argument that functions within this routine might use

Value

a real number

Examples

```
fitNoise(7)
fitNoise(4, fitmetric=rmse)
```

KS_D	<i>KS_D</i>
------	-------------

Description

Calculates the two-sided KS statistic given vectors for x and y

Usage

```
KS_D(x, y)
```

Arguments

x a vector or matrix of real numbers
y a vector or matrix of real numbers

Value

a vector of numbers with length equal to the number of trials (N)

Examples

```
KS_D(c(1,2,3,4,5),c(1,2,3,4,4))  
KS_D(matrix(runif(100),ncol=7),matrix(runif(100),ncol=7))
```

KS_Dm

KS_Dm

Description

Calculates the one-sided KS statistic given vectors for x and y

Usage

```
KS_Dm(x, y)
```

Arguments

x a vector or matrix of real numbers
y a vector or matrix of real numbers

Value

a vector of numbers with length equal to the number of trials (N)

Examples

```
KS_Dm(c(1,2,3,4,5),c(1,2,3,4,4))  
KS_Dm(matrix(runif(100),ncol=7),matrix(runif(100),ncol=7))
```

KS_Dp	<i>KS_Dp</i>
-------	--------------

Description

Calculates the one-sided KS statistic given vectors for x and y

Usage

```
KS_Dp(x, y)
```

Arguments

x	a vector or matrix of real numbers
y	a vector or matrix of real numbers

Value

a vector of numbers with length equal to the number of trials (N)

Examples

```
KS_Dp(c(1,2,3,4,5),c(1,2,3,4,4))
KS_Dp(matrix(runif(100),ncol=7),matrix(runif(100),ncol=7))
```

pcdfs	<i>Construct pdf and cdf for one of several distributions and one of several possible noise distributions</i>
-------	---

Description

Generates the vectors of random numbers for both observation and model and calls pcdfs

Usage

```
pcdfs(dof, order = 5, ndecimals = 2, dist = rnorm, fitmetric = R2, ...)
```

Arguments

dof	an integer
order	a real number
ndecimals	an integer
dist	a random number distribution function
fitmetric	a character string naming a standard fit metric (R2, rmse, or user-defined)
...	any argument that functions within this routine might use

Value

a data frame

Examples

```
pcdfs(5, order=5)
pcdfs(10, order=4, fitmetric=rmse)
```

plotcdf

Plot CDF

Description

Plots the Cumulative Probability Density Function

Usage

```
plotcdf(dof, order = 4, dist = rnorm, fitmetric = R2, ...)
```

Arguments

dof	an integer
order	a real number
dist	a random number distribution function
fitmetric	a character string naming a standard fit metric (R2, rmse, or user)
...	any argument that functions within this routine might use

Value

ggplot object

Examples

```
plotcdf(5, dist=rnorm, fitmetric=rmse)
```

plotConstNoise	<i>Plot Measured Value with Constant Noise</i>
----------------	--

Description

Plots the Fit Equivalent

Usage

```
plotConstNoise(measured_value, dof, doflist = c(2:30), pct = 0.95,
  order = 4, add_mvp = F, fitmetric = R2, ...)
```

Arguments

measured_value	a real number within the range of fitmetric
dof	an integer, degrees of freedom
doflist	a vector of integers identifying the range of degrees of freedom to plot over
pct	a real number between 0 and 1, minimum acceptable level of noise
order	a real number
add_mvp	a logical value indicating whether to include or not the fitEquive placed at each dof.
fitmetric	a character string naming a standard fit metric (R2, rmse, or user)
...	any argument that functions within this routine might use

Value

ggplot object

Examples

```
plotConstNoise(0.8, 5, doflist=c(2:10),order=3)
plotConstNoise(0.1, 5, doflist=c(2:10),order=3,fitmetric=rmse)
```

plotConstValue	<i>Plot Fit Equivalent and Constant Value</i>
----------------	---

Description

Plots the Fit Equivalent with Constant Measured Value

Usage

```
plotConstValue(measured_value, doflist = c(2:30), pct = 0.95, order = 4,
  ndecimals = 2, fitmetric = R2, ...)
```

Arguments

<code>measured_value</code>	a real number, the actual measured value
<code>doflist</code>	a vector of integers for the degrees of freedom
<code>pct</code>	a real number between 0 and 1, acceptable minimum level of noise
<code>order</code>	a real number
<code>ndecimals</code>	an integer
<code>fitmetric</code>	a character string naming a standard fit metric (R2, rmse, or user)
<code>...</code>	any argument that functions within this routine might use

Value

ggplot object

Examples

```
plotConstValue(0.8, doflist=c(2:10),order=3)
plotConstValue(0.1, doflist=c(2:10),order=3,fitmetric=rmse)
```

<code>plotNoiseLevel</code>	<i>Plot Noise threshold</i>
-----------------------------	-----------------------------

Description

Plots the Noise threshold for each degree of freedom

Usage

```
plotNoiseLevel(doflist = c(2:30), pctlist = c(0.95), order = 4,
  ndecimals = 2, fitmetric = R2, ...)
```

Arguments

<code>doflist</code>	a vector of degrees of freedom
<code>pctlist</code>	a vector of percentiles (expressed as fractions)
<code>order</code>	a real number greater than 0 and less than 7
<code>ndecimals</code>	an integer
<code>fitmetric</code>	a character string naming a standard fit metric (R2, rmse, or user)
<code>...</code>	any argument that functions within this routine might use

Value

ggplot object

Examples

```
plotNoiseLevel(doflist=c(2:10))
```

plotpdf

Plot PDF

Description

Plots the Probability Density Function

Usage

```
plotpdf(dof, order = 5, dist = rnorm, fitmetric = R2, ...)
```

Arguments

dof	an integer
order	a real number
dist	a random number distribution function
fitmetric	a character string naming a standard fit metric (R2, rmse, or user)
...	any argument that functions within this routine might use

Value

ggplot object

Examples

```
plotpdf(5, dist=runif, fitmetric=rmse)
```

R2

R-squared

Description

Calculates R-squared given vectors for x and y

Usage

```
R2(x1, y1)
```

Arguments

x1 a vector or matrix of real numbers
y1 a vector or matrix of real numbers

Value

a vector of numbers with length equal to the number of trials (N)

Examples

```
R2(c(1,2,3,4,5),c(1,2,3,4,4))  
R2(matrix(runif(10000),ncol=5),matrix(runif(10000),ncol=5))
```

rmse

rmse

Description

Calculates rmse given vectors for y and y_pred

Usage

```
rmse(obs, mdl)
```

Arguments

obs a vector or matrix of real numbers
mdl a vector or matrix of real numbers

Value

a vector of numbers with length equal to the number of trials (N)

Examples

```
rmse(c(2:20),c(2:20 + c(rep(0.1,5),rep(-0.2,10),rep(0.3,4))))  
mata <- matrix(runif(10000),ncol=5)  
matb <- matrix(runif(10000),ncol=5)  
rmse(mata, matb)
```

Table_dofbypct	<i>Noise Threshold Table</i>
----------------	------------------------------

Description

Constructs a table in the form of a data frame of threshold noise values for a given fit metric, a given set of possible degrees of freedom and a given set of percentiles

Usage

```
Table_dofbypct(doflist = NULL, pctlist = NULL, order = 4, ndecimals = 2,
  fitmetric = R2, trend = NULL, ...)
```

Arguments

doflist	a vector of integers indicating the degrees of freedom
pctlist	a vector of numbers between 0 and 1 standing for the percentile of noise
order	a real number
ndecimals	an integer
fitmetric	a character string naming a standard fit metric (R2, rmse, or user)
trend	a character string "Increasing", "Decreasing", "Flat", "Uncertain" which describes the general slope of the fitmetric function.
...	any argument that functions within this routine might use

Value

a data frame

Examples

```
Table_dofbypct(fitmetric=rmse)
```

Table_pctbyfuncs	<i>Print boilerplate stats</i>
------------------	--------------------------------

Description

Prints a standard set of stats for a given degree of freedom

Usage

```
Table_pctbyfuncs(dof, pctlist = c(0.9, 0.95, 0.99), ndecimals = 2,
  dist = rnorm, order = 5, fitmetriclist = c(R2, rmse), ...)
```

Arguments

dof	an integer
pctlist	a vector of real numbers between 0 and 1, minimum acceptable noise levels.
ndecimals	an integer
dist	a random number distribution function
order	a real number less than 7
fitmetriclist	a vector of goodness-of-fit functions
...	any argument that functions within this routine might use

Value

data frame

Examples

```
Table_pctbyfuncs(6, dist=rnorm, sd=0.3)
```

utrend	<i>Determine the general trend (general, overall slope) of the fitmetric function</i>
--------	---

Description

Calculates the fitmetric function at two distant points and calculates the slope using lineal regression.

Usage

```
utrend(fun)
```

Arguments

fun	as function R2, rmse, or user (user-defined function)
-----	---

Value

a character string ("Increasing", "Decreasing", "Flat", "Uncertain")

Examples

```
utrend(R2)
utrend(rmse)
```

Index

difn, [2](#)

fit, [3](#)

fitEquiv, [3](#)

fitmetric_check, [4](#)

fitNoise, [5](#)

KS_D, [5](#)

KS_Dm, [6](#)

KS_Dp, [7](#)

pcdfs, [7](#)

plotcdf, [8](#)

plotConstNoise, [9](#)

plotConstValue, [9](#)

plotNoiseLevel, [10](#)

plotpdf, [11](#)

R2, [11](#)

rmse, [12](#)

Table_dofbypct, [13](#)

Table_pctbyfuncs, [13](#)

utrend, [14](#)