Package 'stats4teaching'

July 23, 2025

Type Package

Title Simulate Pedagogical Statistical Data

Version 0.1.0

Description Univariate and multivariate normal data simulation. They also supply a brief summary of the analysis for each experiment/design:
- Independent samples.
- One-way and two-way Anova.
- Paired samples (T-Test & Regression).
- Repeated measures (Anova & Multiple Regression).
- Clinical Assay.

License GPL-3

Encoding UTF-8

RoxygenNote 7.1.2

Imports asbio, car, clusterGeneration, knitr, MASS, MVN, nortest, psych, pwr, rstatix, stats

NeedsCompilation no

Author Cabello Esteban [aut, cre],
Femia Pedro [aut]

Maintainer Cabello Esteban <estebancabellogarcia@gmail.com>

Repository CRAN

Date/Publication 2022-10-04 05:30:07 UTC

Contents

---

anova1way   *One-Way ANOVA*

Description

anova1way is used to generate multivariate data in order to compute analysis of variance with 1 factor. It provides balanced and unbalanced ANOVA (as long as homogeneity of variances is satisfied. In other case it is provided Welch test).

Usage

```
anova1way(k = 3,n , mean = 0, sigma = 1,
      coefvar = NULL, method = c("Tukey", "LSD", "Dunnett", "Bonferroni", "Scheffe"),
         conf.level = 0.95, dec = 2)
```

Arguments

k	number of levels. By default k = 3.
n	size of samples.
mean	vector of means.
sigma	vector of standard deviations.
coefvar	an optional vector of coefficients of variation.
method	post-hoc method applied. There are five possible choices: "Tukey", "LSD", "Dunnett", "Bonferroni", "Scheffe". Can be specified just the initial letter.
conf.level	confidence level of the interval.
dec	number of decimals for observations.

Details

If mean or sigma are not specified it is assumed the default values of 0 and 1.

If coefvar (= sigma/mean) is specified, function omits sigma.

Number of samples is choosen by k (by default k = 3). Therefore, if the others parameters (n, mean, sigma, coefvar) have not same length, function rep will be used. Pay attention if vectors dont have same length.

Moreover, not only gives samples for each level, but also the ANOVA table and post-hoc test (in case of significance). By default conf.level = 0.95 and Tukey method is used. If the homogeneity of variances is not verified (using Bartlett test), the Welch test is performed.

Value

List containing the following components:

- `Data`: a data frame containing the samples created.
- `Anova`: anova fitted model.
- `Significance`: significance of the factor.
- `Size.effect`: size effect of the factor.
- `Test Post-Hoc`: test Post-Hoc.

Examples

```
anova1way(k=4,n=c(40,31,50),mean=c(55,52,48,59),coefvar=c(0.12,0.15,0.13),conf.level = 0.99)

anova1way(k=3,n=15,mean=c(10,15,20),sigma =c(1,1.25,1.1),method ="B")
```

---

anova2way   *Two-Way ANOVA*

Description

anova2way returns multivariate data in order to compute analysis of variance with 2 factors.

Usage

```
anova2way(k =2 , j = 2, n,  mean = 0, sigma = 1,
      coefvar = NULL, method = c("Tukey", "LSD", "Dunnett", "Bonferroni", "Scheffe"),
         conf.level = 0.95, dec = 2)
```

Arguments

k	number of levels Factor I. By default k=2.
j	number of levels Factor II. By default j=2.
n	number of elements in each group (k,j).
mean	vector of means.
sigma	vector of standard deviations.
coefvar	an optional vector of coefficients of variation.
method	post-hoc method applied. There are five possible choices: "Tukey", "LSD", "Dunnett", "Bonferroni", "Scheffe". Can be specified just the initial letter.
conf.level	confidence level of the interval.
dec	number of decimals for observations.

Value

A list containing the following components:

- `Data`: a data frame containing the samples created.
- `Size.effect`: size effect for each factor and interaction.
- `Significance/Test Post-Hoc`: significance for each factor and interaction and test Post-Hoc for each factor.

Examples

```
anova2way(k=3, j=2, n=c(3,4,4,5,5,3), mean = c(1,4,2.5,5,6,3.75), sigma = c(1,1.5))
```

---

cassay   *Clinical Assay*

Description

Simulates a clinical Assay with 2 groups (control and treatment) before and after intervention.

Usage

```
cassay(n, mean = 0, sigma = 1, coefvar = NULL,
        d.cohen = NULL, dec = 2)
```

Arguments

n	size of samples.
mean	sample mean. Same for both groups before intervention (Pre-test).
sigma	sample standard error.
coefvar	sample coefficient of variation.
d.cohen	size effect (d-Cohen). If not given, randomly generated.
dec	number of decimals for observations.

Value

List containing the following components:

- `Data`: a data frame containing the samples created (Columns: Group, PreTest & PostTest).
- `Model`: linear regression model.

Examples

```
cassay(c(10,12), mean = 115, sigma = 7.5, d.cohen= 1.5)
cassay(24, mean = 100, sigma = 5.1)
```

---

generator   *Generation of multivariate normal data.*

Description

This function generates univariate and multivariate normal data. It allows simulating correlated and independent samples. Moreover, normality tests and numeric informations are provided.

Usage

```
generator(n , mean = 0, sigma = 1, coefvar = NULL,
    sigmaSup = NULL, dec = 2)
```

Arguments

n	vector size of samples.
mean	vector of means.
sigma	vector of standard deviations or covariance/correlation matrix.
coefvar	an optional vector of coefficients of variation.
sigmaSup	an optional vector of standard deviations if sigma is a correlation matrix.
dec	number of decimals for observations.

Details

If `mean` or `sigma` are not specified it's assumed the default values of 0 and 1.

If `coefvar` (= sigma/mean) is specified, function omits `sigma` and `sigmaSup`. It's assumed that independent samples are desired.

Number of samples are choosen by taken the longest parameter (`n`, `mean`, `sigma`, `coefvar`). Therefore, function `rep` is used. Pay attention if vectors don't have same length!

If `sigma` is a vector, samples are independent. In other case (`sigma` is a matrix), samples are dependent (following information meanst be taken into account: if `sigma` is a correlation matrix, `sigmaSup` is required).

Value

List containing the following components for independent (with the same length) and dependent samples:

- `Samples`: a data frame containing the samples created.
- `Test` normality test for the data (`shapiro.test()` for n <= 50 and `lillie.test()` in other case).

List containing the following components for independent samples with different lengths:

- `X_i` sample number i.

Examples

```
generator(4,0,2)

sigma <- matrix(c(1,0.8,0.8,1),nrow = 2, byrow = 2)
d <- generator(4,mean = c(1,2),sigma, sigmaSup = 1)

generator(10,1,coefvar = c(0.3,0.5))

generator(c(10,11,10),c(1,2),coefvar = c(0.3,0.5))
```

---

is.corrmatrix   *Correlation matrix*

Description

Checks if a given matrix is a correlation matrix for non-degenerate distributions.

Usage

```
is.corrmatrix(matrix)
```

Arguments

matrix   a (non-empty) numeric matrix of data values.

Value

A logical value: True/False.

Examples

```
m1<-matrix(c(1,2,2,1),nrow = 2,byrow = TRUE)
is.corrmatrix(m1)

m2<-matrix(c(1,0.8,0.8,1),nrow = 2,byrow = TRUE)
is.corrmatrix(m2)

m3<-matrix(c(1,0.7,0.8,1),nrow = 2,byrow = TRUE)
is.corrmatrix(m3)
```

---

is.covmatrix   *Covariance matrix*

Description

Checks if a given matrix is a covariance matrix for non-degenerate distributions.

Usage

```
is.covmatrix(matrix)
```

Arguments

matrix   a (non-empty) numeric matrix of data values.

Value

A logical value: True/False.

Examples

```
m1 <- matrix(c(2,1.5,1.5,1), nrow = 2, byrow = TRUE)
is.covmatrix(m1)

m2 <- matrix(c(1,0.8,0.8,1), nrow = 2, byrow = TRUE)
is.covmatrix(m2)

m3 <- matrix(c(1,0.7,0.8,1), nrow = 2, byrow = TRUE)
is.covmatrix(m3)
```

---

is.posDef   *Positive definited matrices*

Description

Checks if a given matrix is positive definited

Usage

```
is.posDef(matrix)
```

Arguments

matrix   a (non-empty) numeric matrix of data values.

Value

A logical value: True/False.

Examples

```
A <- matrix(c(1,2,2,1), nrow = 2, byrow = TRUE)
is.posDef(A)

B <- matrix(c(1,2,3,3,1,2,1,2,1), nrow = 3, byrow = TRUE)
is.posDef(B)
```

---

is.semiposDef   *Semi-Positive definited matrices*

Description

Checks if a given matrix is semi-positive definited.

Usage

```
is.semiposDef(matrix)
```

Arguments

matrix   a (non-empty) numeric matrix of data values.

Value

A logical value: True/False.

Examples

```
A<-matrix(c(2.2,1,1,3), nrow = 2, byrow = TRUE)
is.semiposDef(A)

B<-matrix(c(1,2,3,3,1,2,1,2,1), nrow = 3, byrow = TRUE)
is.semiposDef(B)
```

---

mCorrCov   *Correlation & Covariance matrices.*

Description

Given a correlation matrix and vector of standard deviations (or vector of means and vector of variation coefficients) returns a covariance matrix.

Usage

```
mCorrCov(mcorr, sigma = 1, mu = NULL, coefvar = NULL)
```

Arguments

mcorr	a (non-empty) numeric correlation matrix.
sigma	an optional vector of standard deviations.
mu	an optional vector of means.
coefvar	an optional vector of coefficients of variation.

Details

`coefvar = sigma/mu`.

If `sigma`, `mu` or `coefvar` are not specified, it´s assumed that default values for standard error's are 1. Length of standard error's is created using number of rows of correlation matrix. It's necessary to provide `sigma` or `mu` and `coefvar` (both) in order to obtain a desired covariance matrix.

Length of vectors is taken using `rep`. Pay attention if vectors don't have same length!

Value

`mCorrCov` gives the covariance matrix for a specified correlation matrix.

Examples

```
A <- matrix(c(1,2,2,1), nrow = 2, byrow = TRUE)
mCorrCov(A)

B <- matrix(c(1,0.8,0.7,0.8,1,0.55,0.7,0.55,1), nrow = 3, byrow = TRUE)
mCorrCov(B,mu = c(2,3.5,1), coefvar = c(0.3,0.5,0.7))
```

---

pairedm   *Paired measures (T-Test & Regression)*

Description

Generates two paired measures. It provides T-test and a simple linear regression model for generated data.

Usage

```
pairedm(n, mean = 0, sigma = 1, coefvar = NULL,
        rho = NULL, alternative = c("two.sided", "less", "greater"),
        delta = 0, conf.level = 0.95, dec = 2,
        random = FALSE)
```

Arguments

n	size of each sample.
mean	vector of means.
sigma	vector of standard deviations.
coefvar	an optional vector of coefficients of variation.
rho	Pearson correlation coefficient (optional). If rho = NULL a random covariance matrix is generated by genPositiveDefMat().
alternative	a character string specifying the alternative hypothesis for T-Test. Must be one of "two.sided" (default), "greater" or "less". Can be specified just the initial letter.
delta	true value of the difference in means.
conf.level	confidence level for interval in T-Test.
dec	number of decimals for observations.
random	a logical a logical indicating whether you want a random covariance/variance matrix.

Details

If random = TRUE, rho is omitted and sigma is taken as range for variances of the covariance matrix.

Value

List containing the following components :

- Data: a data frame containing the samples created.
- Model: linear regression model.
- T.Test: a t-test for the samples.

See Also

[clusterGeneration::genpositiveDefMat()]

Examples

```
pairedm(10, mean = c(10,2), sigma = c(1.2,0.7), rho = 0.5, alternative = "g")
pairedm(15, mean =c(1,2), coefvar = 0.1, random = TRUE)
```

---

repeatedm	*Repeated Measures (ANOVA & Multiple Regression)*

Description

Repeated Measures (ANOVA & Multiple Regression)

Usage

```
repeatedm(k, n, mean = 0, sigma = 1, coefvar = NULL,
          sigmaSup = NULL, conf.level = 0.95,
          random = FALSE, dec = 2)
```

Arguments

k	number of variables.
n	number of observations.
mean	vector of means.
sigma	vector of standard deviations/covariance-correlation matrix.
coefvar	vector (optional) of coefficients of variation.
sigmaSup	vector (optional) of standard deviations if sigma is a correlation matrix.
conf.level	confidence level for interval in T-Test.
random	a logical indicating whether you want a random covariance/variance matrix.
dec	number of decimals for observations.

Details

Number of variables must be greater than 3, in order to ensure an ANOVA of repeated measures or a multiple Linear Regression.

sigma can represent a vector or a covariance/correlation matrix. In case sigma is a vector, independent samples are created. By other hand, if it's a correlation matrix parameter sigmaSup is required. For covariance matrices, the function does not require any other parameter or special treatment.

If random = TRUE, a random covariance matrix is generated by using *genpositiveDefMat()*.

Value

A data frame.

See Also

[clusterGeneration::genpositiveDefMat()]

Examples

```
randm <- clusterGeneration::genPositiveDefMat(8, covMethod = "unifcorrmat")
mcov <- randm$Sigma
Sigma <- cov2cor(mcov)
is.corrmatrix(Sigma)
repeatedm(k = 8, n = 8, mean = c(20,5, 30, 15),sigma = Sigma, sigmaSup = 2,  dec = 2)

repeatedm(k = 5, n = 5, mean = c(8,10,5,14,22.5), random = TRUE)
repeatedm(k = 3, n = 8, mean = c(10,5,22.5), sigma = c(3.3,1.5,5), dec = 2)
```

---

  sample2indp   *Independent normal data*

Description

Generates two normal independent samples. It also provides Cohen's effect and T-Test.

Usage

```
sample2indp(n , mean = 0, sigma = 1, coefvar = NULL,
            alternative = c("two.sided", "less", "greater"), delta = 0,
            conf.level = 0.95, dec = 2)
```

Arguments

n	vector of size of samples.
mean	vector of means.
sigma	vector of standard deviations.
coefvar	an optional vector of coefficients of variation.
alternative	a character string specifying the alternative hypothesis for T-Test. meanst be one of "two.sided" (default), "greater" or "less". Can be specified just the initial letter.
delta	true value of the difference in means.
conf.level	confidence level of the interval. It determines level of significance for comparing variances.
dec	number of decimals for observations.

Details

If `mean` or `sigma` are not specified it's assumed the default values of `0` and `1`.

`n` is a vector, so it's possible to generate samples with same or different sizes.

If `coefvar` is given, `sigma` is omitted. Vector of means cannot have any 0.

Value

A list containing the following components:

- `Data`: a data frame containing the samples created.
- `T.Test`: a t-test of the samples.
- `Power`: power of the test.

Examples

```
sample2indp(c(10,12),mean = c(2,3),coefvar = c(0.3,0.5), alternative = "less", delta = -1)

sample2indp(8,sigma = c(1,1.5), dec = 3)
```

---

sample2indp.pow   *Independent normal data*

Description

Generates two normal independent samples with desired power and cohen's effect.

Usage

```
sample2indp.pow(n1, mean = 0, s1= 1, d.cohen, power,
    alternative = c("two.sided", "less", "greater"), delta = 1,
    conf.level = 0.95, dec = 2)
```

Arguments

n1	first sample size.
mean	vector of sample means.
s1	standard deviation for first sample.
d.cohen	Cohen's effect.
power	power of the test.
alternative	a character string specifying the alternative hypothesis for T-Test. Must be one of "two.sided" (default), "greater" or "less". Can be specified just the initial letter.
delta	true value of the difference in means.
conf.level	confidence level of the interval.
dec	number of decimals for observations.

Details

Pooled standard deviation= sp = sqrt((n1 - 1) sigma1^2 +(n2 - 1) sigma2^2) / (n1 + n2 - 2)

`d.cohen` = |mean1 - mean2| / sqrt(sp)

Value

A list containing the following components:

- `Data`: a data frame containing the samples created.
- `Size`: size of each sample.
- `T.test`: a t-test of the samples.

Examples

```
sample2indp.pow(n1 = 30, mean = c(2,3), s1= 0.5, d.cohen = 0.8, power = 0.85, delta = 1)
sample2indp.pow(n1 = 50, mean = c(15.5,16), s1=2 , d.cohen = 0.3, power = 0.33, delta = 0.5)
```

---

stats4teaching *Teaching Statistics Data Simulation*

Description

Univariate and multivariate normal data simulation. They also supply a brief summary of the analysis for each experiment/design.

- Independent samples.
- One-way and two-way ANOVA.
- Paired samples (T-Test & Regression).
- Repeated measures (ANOVA & Multiple Regression).
- Clinical Assay.

Author(s)

Esteban Cabello García and Pedro Jesús Femia Marzo.

Index