Package: statistics (via r-universe)

August 25, 2024

Version 0.7

Date 2023-01-10

Title Miscellanea Functions For an Intro Stats Class

Depends R (>= 3.5.0)

License GPL (>= 3)

Imports admisc (>= 0.25), declared (> 0.15)

Description The majority of functions in this package are designed to facilitate undestanding the statistical concepts taught in the class (such as the functions to create graphics or areas under the normal curve), while some are designed to ease calculations for the lab exercises (for instance the degrees of freedom or the pooled variation for two independent samples).

Repository https://dusadrian.r-universe.dev

RemoteUrl https://github.com/dusadrian/statistics

RemoteRef HEAD

RemoteSha 70e929d406cbebab9b7925af72c45e6ac0501a48

Contents

About the statistics package

A package containing useful functions to teach introductory statistics.

Description

For discreet random variables, draws and calculates the probability of a certain number of favourable outcomes out of a number of repetition of an experiment. For a continous random variable, the graphics represents a normal curve with the area to the left or to the right of a certain z or t value, or between two such values. The package also contains functions to calculate the degrees of freedom and the pooled standard deviation using the t distribution etc.

Details

Package:	statistics
Type:	Package
Version:	0.7
Date:	2023-01-10
License:	GPL-v3

Author(s)

Adrian Dusa

Maintainer: Adrian Dusa (dusa.adrian@unibuc.ro)

See Also

dnorm, pnorm, dbinom

---

anovaFK *ANOVA including the homogeneity of variance test*

Description

The function 'anovaFK' - contains two separate tests: in a first state, the Fligner-Killeen test for the homogeneity of variances is run, and function of this test, the Welch approximation is applied if the groups are not homogeneous.

Usage

```
anovaFK(x, y = NULL, data)
```

Arguments

x	A vector of values or a formula object as in 'lhs ~ rhs', unde 'lhs' contains the values and the 'rhs' contains the groups. Both can be vectors or variables from a dataset.
y	An optional vector of values, when the two variables are not specified using a formula object.
data	A dataset containing the variables specified in the formula object, in case they don't exist as separate objects.

Details

When the variances are not equal, the output differs from the one presented by oneway.test, but the table is similar.

If the degrees of freedom are not what they should be ($k - 1$, and $n - k$ respectively) something must be wrong. Specifically, the grouping variable should be declared as a factor (in case it is not already character), otherwise it is considered metric and a regression model is applied instead of ANOVA.

Declaring a variable as a factor is done using the command: `as.factor`

Author(s)

Adrian Dusa

See Also

aov, anova, oneway.test, fligner.test

Examples

```
values <- c(15, 8, 17, 7, 26, 12, 8, 11, 16, 9, 16, 24, 20, 19, 9, 17, 11, 8, 15, 6, 14)
groups <- rep(1:3, each = 7)

anovaFK(values ~ groups)

# same thing with:
anovaFK(values, groups)

# using a dataset
vgdf <- data.frame(values, groups)
using(vgdf, anovaFK(values ~ groups))
```

daria *Calculate and draw the area under the normal curve z*

Description

The function "daria" - 'd'raws the 'area' under the normal curve for certain values of z.

Usage

```
daria(area, z1, z2, draw = FALSE)
```

Arguments

area	The required area
z1	First z value, in the interval +/- 4
z2	Second z value, in the interval +/- 4
draw	Logical; if TRUE, draw the area

Details

In the argument area, the function accepts:

"l", "u", "left" and "under" for the area to the left of z,

"r", "o", "a", "right" "over" and "above" for the area to the right of z

"b" and "between" for the area between two z values.

z values smaller than -4 and greater than +4 are truncated to these values, since the area to the left and to the right of these values is practically equal to zero.

Author(s)

Adrian Dusa

See Also

pnorm, qnorm

Examples

```
daria("between", -1.96, 1.96)

daria("over", -1)

daria("under", -1)

daria("over", 2, draw = TRUE)
```

dariat *Calculate and draw the area under the t distribution*

Description

A function similar to "daria", with the only difference it uses the t instead of the z distribution. In addition, the function expects an additional parameter for the degrees of freedom.

Usage

```
dariat(area, t1, t2, df, draw = FALSE)
```

Arguments

area	The required area
t1	First t value, in the interval +/- 4
t2	Second t value, in the interval +/- 4
df	Degrees of freedom
draw	Logical; if TRUE, draw the area

Details

In the argument area, the function accepts:

"l", "u", "left" and "under" for the area to the left of z,

"r", "o", "a", "right" "over" and "above" for the area to the right of z

"b" and "between" for the area between two z values.

z values smaller than -4 and greater than +4 are truncated to these values, since the area to the left and to the right of these values is practically equal to zero.

Author(s)

Adrian Dusa

See Also

pt, qt

Examples

```
# for 100 degrees of freedom
dariat("between", -1.96, 1.96, df = 100)

dariat("over", -1, df = 100)

dariat("under", -1, df = 100)

dariat("over", 2, df = 100, draw = TRUE)
```

dbinoms *Calculate probabilities and draw graphics for a binomial distribution*

Description

This function draws graphics for a certain number of repetitions of an experiment, at a certain probability of success, and calculates the probability of obtaining one or more values from a random variable.

Usage

```
dbinoms(x, size, prob, log = FALSE, draw = FALSE,
        zoom = FALSE, new = FALSE, text = FALSE)
```

Arguments

x	Number of favourable outcomes: a value or a vector of values
size	Number of repetitions
prob	Probability of success
log	Logical; if TRUE, the probability is returned as log(p)
draw	Logical; if TRUE, draws the binomial distribution
zoom	Logical; if TRUE, eliminates from the graphic all numbers with probability equal to zero
new	Logical; if TRUE, a new window will be created for each graphic
text	Logical; if TRUE, display the probability above each bar

Author(s)

Adrian Dusa

See Also

dbinom

Examples

```
# 8 repetitions, with a 0.5 probability of success, calculate the
# probability of obtaining between 2 and 4 favourable outcomes
dbinoms(2:4, 8, 0.5)

# less than 7 favourable outcomes
dbinoms(0:6, 8, 0.5)

#at most 7 favourable outcomes
dbinoms(0:7, 8, 0.5)
```

```
# above 5 favourable outcomes
dbinoms(6:8, 8, 0.5)

# at least 5 favourable outcomes
dbinoms(5:8, 8, 0.5)

# exactly 6 favourable outcomes
dbinoms(6, 8, 0.5)

# 1, 3 or 6 favourable outcomes
dbinoms(c(1, 3, 6), 8, 0.5)

# same, drawing the graphic
dbinoms(c(1, 3, 6), 8, 0.5, draw = TRUE)

# same, drawing the probabilities in the graphic
dbinoms(c(1, 3, 6), 8, 0.5, draw = TRUE, text = TRUE)
```

`dfcalc and spooled` *Calculates the degrees of freedom and the pooled variation for a t test*

Description

The function `dfcalc` is used only for two samples t test, when the group variations are NOT equal. For small and independent samples, and unknown but equal population variances, the variances of the two samples are used. As the sample variances are never equal, this function calculates their pooled variance based on the two standard deviations and their respective sample sizes.

Usage

```
dfcalc(x, y, n1, n2)
spooled(x, y, n1, n2)
```

Arguments

`x`	The values of the standard deviation for the first group
`y`	The values of the standard deviation for the second group
`n1`	Size of the first group
`n2`	Size of the second group

Author(s)

Adrian Dusa

Examples

```
group1 <- c(13, 14,  9, 12,  8, 10,  5, 10,  9, 12, 16)
group2 <- c(16, 18, 11, 19, 14, 17, 13, 16, 17, 18, 22, 12)
sd1 <- sd(group1)
sd2 <- sd(group2)
n1 <- length(group1)
n2 <- length(group2)

# more direct
dfcalc(group1, group2)

# if the standard deviations and group sizes are known
dfcalc(sd1, sd2, n1, n2)

# the pooled standard deviation
spooled(sd1, sd2, n1, n2)

# more direct
spooled(group1, group2)
```

---

histc *Histogram with a superimposed normal curve*

Description

Draws a histogram with a normal curve that approximates the distribution.

Usage

```
histc(x, from, to, size = 15, ...)
```

Arguments

x	Numeric vector
from	Starting point on the horizontal axis.
to	End point on the horizontal axis.
size	Size of the graphic, in centimeters.
...	Other parameters, specific to the base hist() function.

Author(s)

Adrian Dusa

Examples

```
x <- sample(18:93, 150, replace = TRUE)

histc(x)

histc(x, 10, 100)

histc(x, 10, 100, xlab = "Age", ylab = "Frequency",
      main = "Histogram for age in years")
```

`mbinom and sbinom` *Calculates the mean and the standard deviation of a discreet random variable*

Description

The function expects a tabel (a data frame or a matrix) with just two columns: the first containing the values of a random variable, and the associated probabilities in the second column.

Usage

```
mbinom(x)
sbinom(x)
```

Arguments

`x`	The data table.

Details

If the sum of the probabilities on the second columns is not equal to 1, the function interprets them as absolute values and recalculates the relative frequencies.

Author(s)

Adrian Dusa

Examples

```
data <- matrix(c(0:4, 0.015, 0.235, 0.425, 0.245, 0.080), ncol = 2)
mbinom(data)
sbinom(data)

data <- data.frame(X = 0:4, P_X = c(12, 188, 340, 196, 64))
mbinom(data)
sbinom(data)
```

t_testAB *Student's t test with a preliminary testing for the homogeneity of variances*

Description

This function executes the t test for one or two groups. In case of two independent groups, the function verifies if the group variances are equal, using the Ansari-Bradley test.

Usage

```
t_testAB(
  x, y = NULL,
  alternative = c("two.sided", "less", "greater"), var.equal = FALSE,
  mu = 0, paired = FALSE, conf.level = 0.95, data = NULL
)
```

Arguments

x	A numeric vector.
y	An optional numeric vector, corresponding to the second group.
alternative	Character, for the alternative hypothesis. See details below.
var.equal	Logical argument indicating whether to treat the two variances as being equal
mu	A number indicating the true value of the mean (or difference in means if performing a two sample test).
paired	Logical indicating whether to perform a paired t-test.
conf.level	Confidence level of the interval
data	An optional matrix or a set of data containing the variables from a formula

Details

The argument `alternative` follows the standard in the base function `t.test()`, and it can be `"two.sided"`, `"less"` or `"greater"`. In addition to those options, this function also allows for `"!="` and `"two.tailed"` for the bidirectional alternative hypothesis, as well as `"<"` and `"lower"` for the one tailed test on the left tail, and `">"` and `"higher"` for the right tailed test, respectively.

Author(s)

Adrian Dusa

Examples

```
group1 <- c(13, 14,  9, 12,  8, 10,  5, 10,  9, 12, 16)
group2 <- c(16, 18, 11, 19, 14, 17, 13, 16, 17, 18, 22, 12)

t_testAB(group1, group2)


# or, if the variables are inside a dataset
dataset <- data.frame(
  values = c(group1, group2),
  group = c(rep(1,11), rep(2,12))
)

t_testAB(values ~ group, data = dataset)
```

Index