Package 'Trendtwosub'

October 12, 2022

Title Two Sample Order Free Trend Nonparametric Inference

Version 0.0.2

Description Non-parametric trend comparison of two independent samples with sequential subsamples. For more details, please refer to Wang, Stapleton, and Chen (2018) <doi:10.1080/00949655.2018.1482492>.

Depends R (>= 3.5.0)

License GPL (>= 2)

Imports magrittr,stats,usethis

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

NeedsCompilation no

Author Yishi Wang Developer [aut, cre],
Matthew Villanueva Developer [aut],
Ann Stapleton Developer [aut]

Maintainer Yishi Wang Developer <wangy@uncw.edu>

Repository CRAN

Date/Publication 2021-02-08 18:40:02 UTC

R topics documented:

`chi.stat` *chi.stat function*

Description

This function calculates the M statistics value as defined in the reference paper.

Usage

```
chi.stat(ftab)
```

Arguments

`ftab`	it is a matrix with dimension 2 by K.

Details

The M statistics is defined as:

$$M = \sum_{l=1}^{K} \left(\frac{(O_{x,l} - E_{x,l})^2}{E_{x,l}} + \frac{(O_{x,l} - E_{x,l})^2}{(n_l n_{l+1} - E_{x,l})} \right) + \sum_{l=1}^{K} \left(\frac{(O_{y,l} - E_{y,l})^2}{E_{y,l}} + \frac{(O_{y,l} - E_{y,l})^2}{(m_l m_{l+1} - E_{y,l})} \right).$$

Value

chi.val, a chisuqre type of statistics value

References

Wang, Y., Stapleton, A. E., & Chen, C. (2018). Two-sample nonparametric stochastic order inference with an application in plant physiology. Journal of Statistical Computation and Simulation, 88(14), 2668-2683.

Examples

```
chi.stat(ftab=rbind(c(20,10,20),c(15,15,20)))
```

`freq.less` *List of functions freq.less function*

Description

This function finds the sum of counts that the x-sample observations is greater than or less than the ones from the y-sample.

Usage

```
freq.less(x, y)
```

Arguments

`x, y`	x and y are numerical vectors of different subsamples. The length of the two vectors can vary.

Details

When there is a tie between any pair of observations, 0.5 is added to the count. Missing value is allowed. Missing value is only added to the calculation when it is compared with another missing value from the other subsample.

Value

Two values are returned: less.count and more.count. The first one is the total count that the observations in x-sample is less than the ones from the y-sample, and the second output is the total count that the observations in x-sample is more than the ones from the y-sample. When there is a tie, 0.5 is added to the count, instead of 1 or 0.

Examples

```
freq.less(x=c(1,2,4,9,0,0,NA),y=c(1,4,9,NA))
```

`gen.decision` *gen.decision function This function compares frequency comparison counts of subsamples from two independent samples, and calculates a simulated p-value with a novel bootstrap method proposed in the reference paper.*

Description

gen.decision function This function compares frequency comparison counts of subsamples from two independent samples, and calculates a simulated p-value with a novel bootstrap method proposed in the reference paper.

Usage

```
gen.decision(est.prob, effn.subsam1, effn.subsam2, fn.rep = 10^3,
  alpha = 0.05)
```

Arguments

est.prob	a matrix of two rows, with each row represents the the sequential comparison results of subsamples from a sample.
effn.subsam1	the subsample sizes from sample 1.
effn.subsam2	the subsample sizes from sample 2.
fn.rep	the total number of replications.
alpha	the size of type I error.

Details

The dimensions of est.prob, effn.subsam1 and effn.subsam2 need to match. For example, the first two entries of the first two rows from est.prob are pf comparison results from subsample1 and subsample2 of sample1. Thus the sum of the two entries is the product of the two subsample sizes.

Value

critical.value the critical value of the test based on the alpha level provided

chi-stat the chisqure type test statistics value from the sample provided.

pvalue the simulated p-value.

References

Wang, Y., Stapleton, A. E., & Chen, C. (2018). Two-sample nonparametric stochastic order inference with an application in plant physiology. Journal of Statistical Computation and Simulation, 88(14), 2668-2683.

Examples

```
freq.mat<-rbind(c(20,5,10,15,20,5),c(15,10,15,10,20,5));
n.sam1<-rep(5,4);n.sam2<-rep(5,4); n.rep=1000;
gen.decision(freq.mat,n.sam1,n.sam2,n.rep);
### This command will replicate the first p-value in Table 4 of the reference paper.
freq.mat<-rbind(c(40,10,20,30,40,10),c(30,20,30,20,40,10));
n.sam1<-c(5,10,5,10);n.sam2<-c(10,5,10,5); n.rep=1000;
gen.decision(freq.mat,n.sam1,n.sam2,n.rep)
### This command will replicate the second p-value in Table 4 of the reference paper.
```

multi.freq *multi.freq function*

Description

This function find trend in a sample by comparing neighboring subsamples. The subsamples are stored in a list in R.

Usage

```
multi.freq(fsam)
```

Arguments

fsam	a list in R. The order of the vectors in the list follows the order of the subsamples.

Details

The first vector of data in the list will be compared with the second vector in the list by using function freq.less. Then the second vector will be compared with the 3rd vector if there is one. The statistics collected are based on computing:

$$\frac{1}{n_l n_{l+1}} \sum_{i=1}^{n_l} \sum_{j=1}^{n_{l+1}} 1(x_{li} < x_{(l+1)j})$$

Value

count.vec it is a collection of a sequence less.count, more.count based on freq.less function.

References

Wang, Y., Stapleton, A. E., & Chen, C. (2018). Two-sample nonparametric stochastic order inference with an application in plant physiology. Journal of Statistical Computation and Simulation, 88(14), 2668-2683.

Examples

```
x1=c(1,2,4,9,0,0,NA);x2=c(1,4,9,NA);x3=c(2,5,10);
sam=list(x1,x2,x3); #
multi.freq(sam);
```

`pow.ana.gen.decision` *pow.ana.gen.decision function*

Description

This function evaluates the type I error of the proposed test.

Usage

```
pow.ana.gen.decision(mean.prob1, mean.prob2, effn.subsam1, effn.subsam2,
  N.rep = 10^1, boot.rep = 10^1, rseed = 1234, alpha.level = 0.05)
```

Arguments

`mean.prob1,`	the probability that observations of a subsample is less than the ones from another subsample, in sample #1.
`mean.prob2,`	the probability that observations of a subsample is less than the ones from another subsample, in sample #2.
`effn.subsam1`	the subsample sizes from sample 1.
`effn.subsam2`	the subsample sizes from sample 2.
`N.rep`	the total number of bootstrap repetitions needed for calculating type I errors.
`boot.rep`	the number of repetitions needed to calculated simulated p-value,
`rseed`	a random seed.
`alpha.level`	the type I error level that will be assessed.

Value

the simulated type I error.

References

Wang, Y., Stapleton, A. E., & Chen, C. (2018). Two-sample nonparametric stochastic order inference with an application in plant physiology. Journal of Statistical Computation and Simulation, 88(14), 2668-2683.

Examples

```
prob.vec<-c(.4,.2,.3,.6);
sub.sizes1<-c(2,4,3,5,3);sub.sizes2<-c(6,3,2,4,2)
pow.ana.gen.decision(prob.vec,prob.vec,sub.sizes1, sub.sizes1)
pow.ana.gen.decision(prob.vec,prob.vec,sub.sizes1, sub.sizes1,alpha.level=0.1)
```

seedwt.multi.subsample *seedwt.multi.subsample dataset*

Description

seedwt.multi.subsample dataset

Usage

```
seedwt.multi.subsample
```

Format

An object of class `data.frame` with 2916 rows and 10 columns.

Details

multiple maize inbreds were exposed to all combinations of the following stressors: drought, nitrogen, and density stress. Plants were grown in an experimental plot divided into eight sections, and each of the sections received a combination of between zero and three of the stresses previously mentioned, so that all possible stress combinations were included. More details about the experiment can be found in the references

References

Wang, Y., Stapleton, A. E., & Chen, C. (2018). Two-sample nonparametric stochastic order inference with an application in plant physiology. Journal of Statistical Computation and Simulation, 88(14), 2668-2683.

Stutts, L., Wang, Y., & Stapleton, A. E. (2018). Plant growth regulators ameliorate or exacerbate abiotic, biotic and combined stress interaction effects on Zea mays kernel weight with inbred-specific patterns. Environmental and experimental botany, 147, 179-188.

simu.ustat.pattern *simu.ustat.pattern function*

Description

This function create two independent subsamples of various subsample sizes, with a given probability vector.

Usage

```
simu.ustat.pattern(mean.prob.vec, effn.subs, n.rep = 10^2)
```

Arguments

mean.prob.vec	a vector of length 2. Its first element represents the probability that a random observation from one subsample is less than the the one from another subsample..
effn.subs	a vector contains two subsample sizes.
n.rep	the total number of repetition.

Details

each subsample is generated from a normal distribution, with an average generated from the mean.prob.vec.

Value

simu.tab a list of length n.rep. Each element of the list is a 2 by 2 matrix, showing the comparison results from function multi.freq.

References

Wang, Y., Stapleton, A. E., & Chen, C. (2018). Two-sample nonparametric stochastic order inference with an application in plant physiology. Journal of Statistical Computation and Simulation, 88(14), 2668-2683.

Examples

```
simu.ustat.pattern(c(0.8,0.2),c(5,8),n.rep=100)
```

---

sub.test *sub.test function This function calculates the simulated p-value of comparing the trend in subsamples from two independent samples.*

Description

sub.test function This function calculates the simulated p-value of comparing the trend in subsamples from two independent samples.

Usage

```
sub.test(sam1, sam2, fn.rep2)
```

Arguments

sam1,	the first sample.
sam2,	the second sample
fn.rep2	the total number of bootstrap repetitions needed for calculating the simulated p-value.

Value

critical.value the critical value of the test based on the alpha level provided

chi-stat the chisqure type test statistics value from the sample provided.

pvalue the simulated p-value.

References

Wang, Y., Stapleton, A. E., & Chen, C. (2018). Two-sample nonparametric stochastic order inference with an application in plant physiology. Journal of Statistical Computation and Simulation, 88(14), 2668-2683.

Examples

```
attach(seedwt.multi.subsample)
Lev.TN<-levels(TreatmentName);
Lev.Line<-levels(Line);
n<-dim(seedwt.multi.subsample)[1];
level.show=c(1:8);fn.rep3=10^2;
line.name<-Lev.Line[1]; t1.name<-Lev.TN[1];t2.name<-Lev.TN[3];
### To compare the GA treatment and the PACGA treatment from line B73
par(mfrow=c(1,2))
idx<-subset((TreatmentName==t1.name)*(Line==line.name)*(1:n),Env %in% level.show)
idx2<-subset((TreatmentName==t2.name)*(Line==line.name)*(1:n),Env %in% level.show)
boxplot(seedwt[idx]~Env[idx],xlab="ENV levels",ylab=paste('seedwt from',t1.name),
        ylim=c(0,12),cex.lab=1.5,cex.axis=1.8);
boxplot(seedwt[idx2]~Env[idx2], xlab="ENV levels",ylab=paste('seedwt from',t2.name),
        cex.lab=1.5,cex.axis=1.8);
mtext( paste ("Line Name:",line.name), side = 3,outer = TRUE, cex = 2.2,line = -3)
temp.sw1<-seedwt[idx];lab<-Env[idx]; uni.lab<-unique(lab)
sam.1<-lapply(1:length(uni.lab), function(x) temp.sw1[lab==uni.lab[x]])
temp.sw2<-seedwt[idx2];lab2<-Env[idx2]; uni.lab2<-unique(lab2)
sam.2<-lapply(1:length(uni.lab2), function(x) temp.sw2[lab2==uni.lab2[x]])
print(paste("working with line ",line.name,'and treatment',t1.name ,'vs',t2.name ))
resu<-sub.test(sam.1,sam.2,fn.rep2=fn.rep3);
dev.off()
## This will show a similar result as the first experiment of section 5 in the paper.
```

Index