Package ‘DiagTest3Grp’

February 19, 2015

Type Package
Title Diagnostic test summary measures for three ordinal groups
Version 1.6
Date 2014-02-20
Author Jingqin Luo, Chengjie Xiong
Maintainer Jingqin Luo<rosy@wubios.wustl.edu>
Description The R package provides utilities to estimate and make inference on two summary measures for diagnostic tests when there are three ordinal groups ---- volume under ROC surface (VUS) and the extended Youden index, assuming that diagnostic markers increase monotonically and stochastically with disease severity. Point estimates, variance, confidence interval and optimal cut-points both under the normal assumption and also the non-parametric method(s) will be provided for the summary measures. Statistical tests are implemented to compare multiple diagnostic tests and two diagnostic test. Sample size is calculated to estimate the summary measures within a user-specified margin of error for future study planning.
License GPL (>= 2.0)
Depends car,KernSmooth,stats, gplots
LazyLoad yes
NeedsCompilation no
Repository CRAN
Date/Publication 2014-02-20 22:40:48

R topics documented:

  DiagTest3Grp-package ................................................. 2
  AL ................................................................. 4
  bootSample ......................................................... 5
  BW.ref ............................................................. 6
  CI.normal ........................................................... 7
  DiagTest3Grp.Test .................................................... 8
  empirical.cdf ....................................................... 11
  FisherZ ............................................................... 12
Description

Assuming diagnostic marker increase monotonically and stochastically with disease severity, the R package provides utilities to estimate two diagnostic test summary measures for three ordinal groups — volume under ROC surface (VUS) and the extended Youden index. Variance, confidence interval and optimal cut-points both under the normal assumption and also the non-parametric method(s) will be provided for the summary measures. Statistical tests are implemented to compare multiple diagnostic tests and two diagnostic test. Sample size is calculated to estimate the summary measure for a diagnostic test within user-specified margin of error for future study planning.

Details

<table>
<thead>
<tr>
<th>Package</th>
<th>DiagTest3Grp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Package</td>
</tr>
<tr>
<td>Version</td>
<td>1.0</td>
</tr>
<tr>
<td>Date</td>
<td>2010-10-21</td>
</tr>
<tr>
<td>License</td>
<td>GPL-3</td>
</tr>
<tr>
<td>LazyLoad</td>
<td>yes</td>
</tr>
</tbody>
</table>

DiagTest3Grp-package  Diagnostic test summary measures for three ordinal groups
DiagTest3Grp-package

Youden3Grp
Youden3Grp.Variance.Bootstrap
SampleSize.VUS
SampleSize.Youden3Grp

Author(s)

Jingqin Luo, Chengjie Xiong Maintainer: Jingqin Luo <rosy@wubios.wustl.edu>

References


See Also

Youden3Grp VUS

Examples

data(AL)
group <- AL$group
table(group)

###take the negated FACTOR1 marker measurements
factor1 <- ~AL$FACTOR1

x <- factor1[group=="D-"]
y <- factor1[group=="D0"]
z <- factor1[group=="D+"]

### VUS under normality assumption
normal.res <- VUS(x,y,z,method="Normal",p=0,q=0,alpha=0.05)
###S3 method
print(normal.res)
###S3 method
plot(normal.res)

###Youden index using normal method
youden.res <- Youden3Grp(x=x,y=y,z=z,method="Normal")
print(youden.res)
plot(youden.res)
Description

The dataset is a subset of the longitudinal cohort of Washington University (WU) Alzheimer’s Disease Research Center (ADRC). In the AL dataset, measurements of 14 neuropsychological markers were collected on 118 independent individuals of age 75. The individuals were classified into 3 groups based on published clinical demential rating (CDR).

Usage

data(AL)

Format

A data frame with 118 observations on the following 15 variables.

- **group**: a factor with levels $D^-$, $D^0$, $D^+$, the $D^-$ group has CDR (clinical demential rating)=0, i.e., no dementia group; the $D^0$ group has CDR=0.5, i.e., mild dementia and the $D^+$ group has CDR=1, Alzheimer’s disease.
- **factor1**: a numeric vector, measurements on the neuropsychometric test for “global factor”.
- **ktemp**: a numeric vector, measurements on the neuropsychometric test for “temporal factor”.
- **kpar**: a numeric vector, measurements on the neuropsychometric test for “parietal factor”.
- **kfront**: a numeric vector, measurements on the neuropsychometric test for “frontal factor”.
- **zpsyPPT**: a numeric vector, measurements on the neuropsychometric test for “longical memory”.
- **zpsyPPU**: a numeric vector, measurements on the neuropsychometric test for “digital span forward”.
- **zpsyPPV**: a numeric vector, measurements on the neuropsychometric test for “digital span backward”.
- **zinfo**: a numeric vector, measurements on the neuropsychometric test for “information”.
- **zbentc**: a numeric vector, measurements on the neuropsychometric test for “visual retention (10s)”.
- **zbentd**: a numeric vector, measurements on the neuropsychometric test for “visual retention (copy)”.
- **zboston**: a numeric vector, measurements on the neuropsychometric test for “boston naming”.
- **zmentcon**: a numeric vector, measurements on the neuropsychometric test for “mental control”.
- **zwarflu**: a numeric vector, measurements on the neuropsychometric test for “word fluency”.
- **zassc**: a numeric vector, measurements on the neuropsychometric test for “associate learning”.

Details

See references for more details on the explanation of the 14 markers and scoring details.
Source


Examples

data(AL)

```r
## maybe str(AL); plot(AL); ...
```

---

**bootSample**  
*Draw a bootstrap sample from a given vector of data values*

**Description**

Draw one bootstrap sample of a given vector of sample values

**Usage**

`bootSample(x, seed0)`

**Arguments**

- `x`: a numeric vector, sample data.
- `seed0`: a random seed for bootstrap sampling.

**Value**

Returns a numeric vector which is a bootstrap sample drawn randomly with replacement from and with the same vector length as the sample data `x`.

**Author(s)**

Jingqin Luo

**Examples**

```r
## generate random data
x <- rnorm(100)

## draw a bootstrap sample
boot.x <- bootSample(100, seed0=12345)
```
Calculate the bandwidth (normal reference rule and Sheather-Jones plug-in rule) in kernel smoothing for density estimation

Description

Calculate the bandwidth in kernel smoothing for density estimation by use of normal reference rule and by Sheather-Jones plug-in method.

Usage

```r
BW.ref(x, method = "KS-SJ")
```

Arguments

- `x`: a numeric vector, e.g., a diagnostic test’s measurements.
- `method`: specify which method to be used for bandwidth calculation for kernel smoothing density estimation in the extended Youden index estimation. The method choices include: “KS”—uses the simple normal reference rule; “KS-SJ”—uses the Sheather-Jones plug-in method which used the 2-stage solve-the-equation rules calling the R package `KernSmooth` (see references).

Details

Kernel smoothing estimation of density function requires specification of kernel function and bandwidth. Normal density is conveniently used as the kernel. Bandwidth selection is important in density estimation.

The normal reference rule calculates the bandwidth as $1.06 \min (sd, \frac{iqr}{1.34}) n^{-0.2}$ with $sd$ and $iqr$ indicating the standard deviation and inter-quartile range. The Sheather and Jones direct plug-in bandwidth estimator (JRSSB, 1991) is obtained by using the R package `KernSmooth`.

Value

Return a numeric value, the calculated bandwidth for kernel smoothing estimate on sample data $x$.

Note

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)

Jingqin Luo
References

density estimation. *Journal of Royal Statistical Society (Serie B)* **53** 683–690


See Also

`Youden3Grp KernelSmoothing.cdf`

Examples

```r
# generate data
x <- rnorm(100, 10, 1.5)

# calculate bandwidth by normal reference rule
bw1 <- BW.ref(x=x, method="KS")

# calculate bandwidth by Sheather-Jones plug-in method
bw2 <- BW.ref(x=x, method="KS-SJ")
```

---

**CI.normal**

*Provide the lower and upper limit of a normal confidence interval given an estimate and its associated variance*

**Description**

Provide the lower and upper limit of a 100*(1-alpha)% normal confidence interval (CI) given an
estimate and its associated variance

**Usage**

`CI.normal(est, var0, alpha = 0.05)`

**Arguments**

- `est` A numeric value, a estimate on which the CI to be calculated.
- `var0` A numeric value, the variance of the estimate.
- `alpha` A numeric value, for a (1-alpha)*100% confidence interval. Default alpha=0.05,
i.e., 95% confidence interval.
Details

The lower and upper limit of the confidence interval for an estimate \( \hat{\theta} \) is calculated as \( \hat{\theta} - Z_{\frac{\alpha}{2}} * \sqrt{(var0)} \). \( Z_{\frac{\alpha}{2}} \) is the normal \( \frac{\alpha}{2} \) quantile.

Value

Returns a list including,

- `est` The estimate provided by user.
- `var0` The variance of the estimate provided by user.
- `lower` the lower limit of a 100*(1-alpha)% normal confidence interval.
- `upper` the upper limit of a 100*(1-alpha)% normal confidence interval.

Author(s)

Jingqin Luo

See Also

Youden3Grp Youden3Grp.Variance.Bootstrap

Examples

```r
est <- 5
var0 <- 1.5
res <- CI.normal(est, var0, alpha=0.05)
```

Description

Statistical test to compare the summary measure (VUS or Youden) estimates between two markers or multiple markers assuming that the marker follows normal distribution in each diagnosis group, for both independent samples and paired samples.

Usage

```r
DiagTest3Grp.Test (dat, paired=FALSE, type=c("VUS","Youden"), p=0, q=0, mu=0, conf.level=0.95, alternative=c("two.sided","less","greater"))
```
Arguments

dat For unpaired data (data measured on different independent sets of samples), dat is a list of K (=number of markers) components, each corresponding to a marker. Each component of the list should be a data frame of number of observation by two columns: (1) group (note: must be at the 1st column) with the three levels of \(D^-\), \(D^0\) and \(D^+\) (corresponding to increasing disease severity respectively); (2) numeric marker measurements at the 2nd column. For paired data, dat should be a data frame with observations at row and group/markers at column (group must be at the 1st column) with the three levels \(D^-\), \(D^0\) and \(D^+\) (corresponding to increasing disease severity respectively), followed by markers at the 2nd to the last column.

paired A logic value TRUE/FALSE, indicating whether all markers are measured on the same set (paired=TRUE) or not (paired=FALSE).

type A character value. type="VUS" or "Youden", indicates which summary measure (VUS or Youden) will be used to indicate the discriminative ability of the three-group diagnostic tests. Allows unique argument matching, i.e., can use unique abbreviations such as "V" or "Y".

p A numeric value. To test on partial VUS satisfying specificity \(\geq p\) and sensitivity \(\geq q\), set p, q to desired value. default=0 for full VUS.

q A numeric value. To test on partial VUS satisfying specificity \(\geq p\) and sensitivity \(\geq q\), set p, q to desired value. default=0 for full VUS.

mu A numeric value. The hypothesized true difference between two markers. Default mu=0 but can be changed to a nonzero value.

conf.level a numeric value. The confidence interval. default=95%. The confidence interval on point estimate of abs diff between two markers’s summary measures will be returned. For \(\geq 3\) markers, this will be NA.

alternative A character value. Now is fixed as "two-sided".

Details

Let \(\theta_j\) indicate the summary measure (either VUS or Youden) on a marker j. When testing Q markers (\(Q \geq 3\)) markers, the omnibus test similar to one-way ANOVA is performed with the hypothesis \(H_0: \theta_1 = \theta_2 = \cdots = \theta_Q\) against the alternative \(H_1:\) at least a pair of markers have unequivalent summary measures. The resulting statistic is a Chi-square statistic with Q-1 degree of freedom. When testing 2 markers, the normal test is performed to test the hypothesis \(H_0: \theta_1 - \theta_2 = \theta_0\) against \(H_1: \theta_1 - \theta_2 \neq 0\) for alternative = “two.sided”, \(H_1: \theta_1 - \theta_2 < 0\) for alternative = “less” and \(H_1: \theta_1 - \theta_2 > 0\) for alternative = “greater". The z-statistic is formed (difference in summary measures divided by the standard error on the difference) for normal test.

Value

An object of the classic S3 class htest (slightly modified) with the list of components.

statistic A numeric value. In situations of testing \(\geq 3\) markers, this is a Chi-Square statistic; in testing 2 markers, this is a Z-statistic.

p.value A numeric value. The p-value testing the hypothesis (see details) from Chi-square test (for \(\geq 3\) markers) or normal test (2 markers).
parameter A numeric value. For the Chi-square test on >=3 markers, it’s the degree of freedom. In the normal test on 2 markers, it’s mean 0 and sd=1.
estimate A numeric vector. The summary measures estimated on each marker.
conf.int A numeric vector of length 2. The confidence interval on the difference in the summary measures of two marker(marker1 at the 1st column of dat input-marker2 at the 2nd column at dat input). For testing >=3 markers, NA.
method A character value. “Chi-square test” for >=3 markers or “Normal-test” for 2 markers.
type A character value. Either VUS or Youden, which type of summary measure is in use.
null.value A numeric value. The inputted value for mu which is now fixed as 0 for >=3 marker comparison but can be changed to any numeric value for two marker comparison.
alternative A character value which is always “two.sided” for >=3 marker comparison but can be “two.sided”, “less” or “greater”.
data.name Usually prints Test of the name of the inputted data frame on the type of a summary measure.
Sigma A numeric matrix. The covariance matrix on the summary measures estimated on markers.

Warning

The bootstrapping to obtain the variance on the nonparametric VUS estimate may take a while.

Note

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)

Jingqin Luo

References


Empirical cumulative density function estimation

Description

Provide empirical cumulative density function (CDF) estimation of a sample vector xx at a point \( c_0 \).

Usage

```r
empirical.cdf(xx, c0)
```

Arguments

- **xx**: A numeric vector, sample data.
- **c0**: A numerical value, the value at which the CDF will be estimated.

Details

The CDF of xx at \( c_0 \) is estimated as \( \sum I(\xi \leq c_0) / n \) where \( n \) is the length of xx and \( I(.) \) is the indicator function giving 1 if the criterion in the bracket is satisfied and 0 otherwise.

Value

Return a numerical value indicating the cumulative probability of xx at \( c_0 \) by empirical CDF method.
FisherZ

Author(s)
Jingqin Luo

See Also

KernelSmoothing.cdf Youden3Grp

Examples

```r
x <- rnorm(100)
res <- empirical.cdf(xx=x,c0=0.5)
```

---

FisherZ Fisher’s Z transformation

Description
Fisher’s Z transformation on an estimator in the range of 0~1.

Usage

```r
FisherZ(x)
```

Arguments

- `x` A numeric value, an estimator within the range of 0 to 1.

Details
Fisher’s Z transformation transform an estimator `x` which is in the range of 0 and 1 to 
\[
z = \frac{1}{2} \log\left(\frac{1+x}{1-x}\right).
\]
The transformation is useful for skewed distribution of a variable between 0 and 1. `z` after transformation will be more normally distributed. Fisher’s Z transformation has been implemented on correlation coefficient, volume under ROC.

Value

Return a numeric value—the transformed estimator `z`.

Note

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)
Jingqin Luo
References

Fisher, R.A. (1915) Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika* 10 (4) 507


See Also

FisherZ.Var

Examples

```r
x <- 0.9
z <- FisherZ(x)
```

---

### FisherZ.Var

*Calculate the variance of an estimator after Fisher’s Z transformation*

**Description**

Calculate the variance of an estimator after Fisher’s Z transformation

**Usage**

FisherZ.Var(x, var.x)

**Arguments**

- `x` A numeric value, an estimator in the range of 0~1.
- `var.x` A numeric value, the variance of the estimator `x`.

**Details**

The variance of an estimator `x` after Fisher’s Z transformation is \( \frac{\text{var}(x)}{(1-x^2)^2} \cdot \text{var}(x)/(1 - x^2)^2 \).

**Value**

Return a numerical value—the variance of an estimator after Fisher’s Z transformation.

**Author(s)**

Jingqin Luo
References

Fisher, R.A. (1915) Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika* **10** (4) 507


See Also

`FisherZ`

Examples

```r
x <- 0.8
var.x <- 0.06
new.var <- FisherZ.Var(x, var.x)
```

**KernelSmoothing.cdf**

*Kernel smoothing cumulative density function (CDF)*

**Description**

Calculate the kernel smoothing cumulative density function (CDF) of a given sample data at a user-specified value.

**Usage**

`KernelSmoothing.cdf(xx, c0, bw)`

**Arguments**

- `xx`: A numeric vector, sample data.
- `c0`: A numeric value, the cumulative probability for which \( \Pr(xx \leq c_0) \) will be estimated.
- `bw`: A numeric value indicating the bandwidth used in the kernel smoothing density approximation.

**Details**

Kernel smoothing is a popular method to approximate a probability density function (PDF) or cumulative density function (CDF). Normal density function is conveniently used in the function as the kernel density and bandwidth is calculated according to the normal reference rule or the Sheather-Jones plug-in method in the package or can be specified arbitrarily by users. For a sample data `xx`, the cumulative CDF \( P(xx \leq c_0) \) can be approximated by kernel smoothing method as \( \frac{1}{n} \sum_{j=1}^{n} \Phi(\frac{xx_j - c_0}{bw}) \) where `n` is the sample size of `xx`. 
NonParametric.VUS

Value

Return a numeric value—the cumulative probability.

Note

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)

Jingqin Luo

References


See Also

bwNref Youden3Grp

Examples

### generate data
x <- rnorm(100, 10, 1.5)

### calculate bandwidth by normal reference rule
bw1 <- KernelSmoothing.cdf(xx=x, c0=6, bw=0.1)

NonParametric.VUS

Nonparametric estimate of volume under ROC surface (VUS)

Description

Nonparametric estimate of volume under ROC surface (VUS)

Usage

NonParametric.VUS(x, y, z, alpha=0.05, NB00T=50, FisherZ=FALSE)
NonParametric.VUS

Arguments

- **x**: A numeric vector, a diagnostic test’s measurements in the \( D^- \) (usually healthy subjects).
- **y**: A numeric vector, a diagnostic test’s measurements in the \( D^0 \) (usually mildly diseased subjects).
- **z**: A numeric vector, a diagnostic test’s measurements in the \( D^+ \) (usually severely diseased subjects).
- **alpha**: A numeric value, \((1-alpha)\times100\%\) percentile Confidence interval of the VUS estimate under normal assumption. Default, alpha=0.05.
- **nboot**: A integer. Total number of bootstrap samples for implementation.
- **FisherZ**: A logic value. Default=FALSE. If TRUE, will transform the nonparametric estimate through Fisher’s Z transformation:
  \[ \theta^* = \frac{1}{2} \log\left(\frac{1+\theta}{1-\theta}\right) \]

Details

The volume under ROC surface (VUS) indicates the probability of correctly ranking a randomly selected triplet \((U, V, W)\) of a diagnostic test’s measurements, each from \(D^- , D^0, D^+\), i.e., \(VUS = Pr\{U < V < W\}\). The nonparametric estimator estimates VUS by empirical CDF method, calculating the proportion of correct orderings among all possible triplets, each from three diagnosis groups.

Value

Return a numeric value as the nonparametric estimate of the VUS.

Note

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)

Jingqin Luo

References


See Also

VUS Normal.VUS NonParametric.VUS.var
**Examples**

```r
data(AL)
group <- AL$group
table(group)

## take the negated FACTOR1 marker measurements
factor1 <- ~ALS$FACTOR1

x <- factor1[group=="D-“]
y <- factor1[group=="D0”]
z <- factor1[group=="D+“]

nonpar.vus <- NonParametric.VUS(x=x,y=y,z=z)
```

---

**NonParametric.VUS.var**  
*Bootstrap variance of the nonparametric estimate on volume under ROC surface (VUS)*

**Description**

Calculate the variance of the nonparametric estimate of volume under ROC surface (VUS) through bootstrapping.

**Usage**

```r
NonParametric.VUS.var(x, y, z, alpha = 0.05, NBOOT = 100, FisherZ=FALSE)
```

**Arguments**

- `x`  
  A numeric vector, a diagnostic test’s measurements in the $D^-$ (usually healthy subjects).

- `y`  
  A numeric vector, a diagnostic test’s measurements in the $D^0$ (usually mildly diseased subjects).

- `z`  
  A numeric vector, a diagnostic test’s measurements in the $D^+$ (usually severely diseased subjects).

- `alpha`  
  A numeric value, confidence level for confidence interval, will result in $(1-alpha)*100\%$; Default alpha=0.05, i.e., 95% confidence interval.

- `NBOOT`  
  A numeric value, total number of bootstrap samples to draw for estimate of variance of the VUS nonparametric estimate.

- `FisherZ`  
  A logic value. Default=FALSE. If TRUE, will transform the nonparametric estimate through Fisher’s Z transformation: $\theta^* = \frac{1}{2}\log\left(\frac{1+\theta}{1-\theta}\right)$

**Details**

Return a numeric value, the variance estimation on the nonparametric VUS estimate obtained through bootstrapping.
**Value**

Return a list of outputs, including,

- **variance**: A numeric value, the estimate of variance of nonparametric VUS through bootstrapping.
- **lower**: A numeric value, the lower limit of bootstrap basic quantile confidence interval for the estimate of variance of nonparametric VUS, i.e., \((\alpha/2) \times 100\%\) quantile. Under default \(\alpha=0.05\), lower = basic 0.25% quantile of bootstrap variance estimates of the nonparametric VUS estimator.
- **upper**: A numeric value, the upper limit of bootstrap basic quantile confidence interval for the estimate of variance of nonparametric VUS, i.e., \((1-\alpha/2) \times 100\%\) quantile. Under default \(\alpha=0.05\), upper = basic 0.975% quantile of bootstrap variance estimates of the nonparametric VUS estimator.

**Warning**

The bootstrap variance estimate may take a minute.

**Note**

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

**Author(s)**

Jingqin Luo

**References**


**See Also**

VUS Normal.VUS Youden3Grp.Variance.Bootstrap

**Examples**

```r
data(AL)
group <- AL$group
table(group)

## take the negated FACTOR1 marker measurements
factor1 <- -AL$FACTOR1

x <- factor1[group=="D-"]
y <- factor1[group=="D0"]
z <- factor1[group=="D+"]

nonpar.vus <- NonParametric.VUS.var (x=x, y=y, z=z, alpha=0.05, NBOOT=10)
```
Normal.VUS

Volume under ROC surface (VUS) and partial VUS estimate and associated variance under normality assumptions

Description

Estimate volume under ROC surface (VUS), partial VUS and associated variance assuming a diagnostic test follows normal distribution in all three ordinal diagnosis groups $D_-, D_0, D_+$.

Usage

Normal.VUS(x, y, z, p = 0, q = 0, alpha = 0.05, subdivisions = 50000, lam.minus = 1/3, lam0 = 1/3, lam.plus = 1/3, typeIerror = 0.05, margin=0.05,FisherZ=FALSE,...)

Arguments

x  A numeric vector, a diagnostic test’s measurements in the $D^-$ (usually healthy subjects).
y  A numeric vector, a diagnostic test’s measurements in the $D^0$ (usually mildly diseased subjects).
z  A numeric vector, a diagnostic test’s measurements in the $D^+$ (usually severely diseased subjects).
p  A numeric value, the minimum required specificity, $0 \leq p < 1$, for calculation partial volume under ROC surface. Default, p=0.
q  A numeric value, the minimum desired sensitivity, $0 \leq q < 1$, for calculation partial volume under ROC surface. Default, q=0. p=q=0 will give the complete VUS estimate, otherwise give the partial VUS estimate satisfying specificity no less than p and sensitivity no less than q.
alpha  A numeric value, the (1-alpha)*100% Confidence interval of the VUS estimate under normal assumption. Default, alpha=0.05.
subdivisions  A numeric value, the number of subintervals for integration using adaptive quadrature in the R function integrate. Default, subdivisions=50000.
lam.minus  A numeric value, the expected population proportion of the $D^-$ group, used for sample size calculation. Default, lam.minus=1/3. The proportions of the three ordinal groups (lam.minus,lam0,lam.plus) should sum to 1.
lam0  A numeric value, the expected population proportion of the $D^0$ group, used for sample size calculation. Default, lam0=1/3. The proportions of the three ordinal groups (lam.minus,lam0,lam.plus) should sum to 1.
lam.plus  A numeric value, the expected population proportion of the $D^+$ group, used for sample size calculation. Default, lam.plus=1/3. The proportions of the three ordinal groups (lam.minus,lam0,lam.plus) should sum to 1.
typeIerror  A numeric value, (1-typeIerror)*100% confidence interval (CI) in sample size calculation. Default typeIerror=0.05, i.e., calculate 95% CI.
margin
A numeric value, the margin of error on the VUS estimates in sample size calculation. Default, margin=0.05. The (1-typeIerror)% CI on VUS estimate under normality is \((VUS - Z_a \times SE(VUS), VUS + Z_a \times SE(VUS))\), then margin=\(Z_a \times SE(VUS)\) or half of the CI’s length, where \(Z_a\) is the normal quantile, \(Z_a=1.95\) given default typeIerror=0.05.

FisherZ
A logic value. Default=FALSE. If set to TRUE, the VUS estimate, associated variance, CI and sample size will be calculated on the logit scale, i.e., after the Fisher Z transformation. In sample size calculation, we place the margin of error "delta" on the CI of the VUS estimate (if FisherZ=FALSE). If VUS is close to 1, the resulting CI on the original scale will have insufficient nominal coverage and under-estimated sample size. Therefore, the VUS estimate, variance, CI and sample size calculation should be implemented on the logit scale via the Fisher Z transformation.

NNN
Other arguments that can be passed to the R function \(\text{integrate}\), e.g., abs.tol,rel.tol,stop.on.error etc.

Details
For three ordinal group diagnostic test, there are two underlying cut-point \(t_−\) and \(t_+\) with \(t_− < t_+\) based on which patients are divided into the three ordinal groups. Patients with a diagnostic test below \(t_−\) will be assigned to \(D_−\); those with the test above \(t_+\) will be assigned to \(D_+\) and the remaining fall into \(D_0\). Following the specificity and sensitivity definition as in diagnostic test for two groups, we call the probability of the first two events as specificity \(x = \Pr_-(T \leq t_-) = F_-(t_-)\) and sensitivity \(y = \Pr_+(T > t_+) = 1 - F_+(t_+) = G_+(t_+)\) where the \(\Pr_i\) and \(F_+\) denotes the probability density function (PDF) and cumulative density function (CDF) of a diagnostic test in \(D_i\), \(i = -, 0, +\) separately. Then, the probability that a patient randomly selected from the \(D_0\) group has the test result between the two cut-points can be expressed as,

\[ z = \int \int_{D_{00}} \left\{ F_0(G_-(y))^1 - F_0(F_-(x)) \right\} dx dy = \int_{-\infty}^{+\infty} F_-(s)G_+(s)f_0(s)ds \]

The integration domain is \(D_{00} = \{0 \leq x \leq 1, 0 \leq y \leq G_+(F_-(x))\}\). Partial VUS requires a marker has specificity at least \(p\) and sensitivity at least \(q\). The equation of partial VUS will be similar to the above but the integration domain is \(D_{pq} = \{p \leq x \leq 1, q \leq y \leq G_+(F_-(x))\}\).

Details on the sample size calculation in the original scale and in the logit scale (after Fisher’s Z transformation) can be found in the reference papers.

Value
A list of components, including

dat
A list of components “x”, “y”, “z”, recording the original marker measurement inputted into the function.
dat.summary A numeric data frame with three rows \((D^-, D^0, D^+\) and three columns (\(\text{"n"}, \text{"mu"}, \text{"sd"})\), summarizing number of observations, mean and SD estimated for each diagnosis group from samples.

estimate A numeric value. A numeric value. If \(p=q=0\), VUS gives the complete VUS estimate, otherwise give the partial VUS estimate satisfying specificity no less than \(p\) and sensitivity no less than \(q\).

variance A numeric value. The variance of VUS estimate under normal assumption.

CI A named numeric vector of length 2. The \((1-\alpha)^\%\) CI on VUS or partial VUS.

sampleSize The sample size on investigating a diagnostic test based on the pilot measurements \(x, y, z\). The sample size calculated such that based on given population proportion of the three diagnosis groups, the margin of error on the \((1-\text{typeIerror})^\%\) confidence interval of the VUS estimate under normality is within given margin of error (by default 0.05).

partialDeriv A numeric data frame with one row and multiple columns, containing relevant parameters \((a,b,c,d)\) and the partial derivatives of VUS estimate w.r.t the relevant parameters which are outputted for performance of statistical tests on markers under normal method or NA under nonparametric method.

Note

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)

Jingqin Luo

References


See Also

VUS NonParametric.VUS NonParametric.VUS.var

Examples

data(AL)
group <- AL$group
table(group)

## take the negated FACTOR1 marker measurements
factor1 <- -AL$FACTOR1
Pairwise.DiagTest3Grp.Test

Do pairwise statistical test to compare the summary measure (VUS or Youden) estimates between a pair of markers among multiple markers assuming that the marker follows normal distribution in each diagnosis group, for both independent samples and paired samples. Resulting p-values can be adjusted for multiple testing. The function returns a nice upper-triangular graphical display and print matrix on pairwise comparisons.

Usage

Pairwise.DiagTest3Grp.Test(dat, paired=FALSE, type=c("VUS","Youden"), p=0, q=0, mu=0, conf.level=0.95, alternative=c("two.sided","less","greater"), p.adjust.method=c("none","holm","hochberg","hommel","bonferroni","BH","BY","fdr"), digits=3,...)

Arguments

dat

For unpaired data (data measured on different independent sets of samples), dat is a list of K (=number of markers) components, each corresponding to a marker. Each component of the list should be a data frame of number of observation by two columns: (1) group (note: must be at the 1st column) with the three levels of $D^-$, $D^0$ and $D^+$ (corresponding to increasing disease severity respectively); (2) numeric marker measurements at the 2nd column. For paired data, dat should
be a data frame with observations at row and group/markers at column (group must be at the 1st column) with the three levels $D^-$, $D^0$ and $D^+$ (corresponding to increasing disease severity respectively), followed by markers at the 2nd to the last column.

**paired**
A logic value TRUE/FALSE, indicating whether all markers are measured on the same set (paired=TRUE) or not (paired=FALSE).

**type**
A character value. type="VUS" or "Youden", indicates which summary measure (VUS or Youden) will be used to indicate the discriminative ability of the three-group diagnostic tests. Allows unique argument matching, i.e., can use unique abbreviations such as "V" or "Y".

**p**
A numeric value. To test on partial VUS satisfying specificity \( \geq p \) and sensitivity \( \geq q \), set \( p, q \) to desired value. default=0 for full VUS.

**q**
A numeric value. To test on partial VUS satisfying specificity \( \geq p \) and sensitivity \( \geq q \), set \( p, q \) to desired value. default=0 for full VUS.

**mu**
A numeric value. The hypothesized true difference between two markers. Default \( \mu=0 \) but can be changed to a nonzero value.

**conf.level**
a numeric value. The confidence interval. default=95%. The confidence interval on point estimate of abs diff between two markers’s summary measures will be returned. For >=3 markers, this will be NA.

**alternative**
A character value. Now is fixed as "two-sided".

**p.adjust.method**
A vector string specifying multiple testing adjustment method. Default="none".

**digits**
A numeric integer. The number of digits for p-value print

**...**
Other arguments that can be input into heatmap.2() function in R package gplots.

### Details

see `p.adjust` for details on multiple testing adjustment on p-values. and see `DiagTest3Grp.Test` for statistical test between two markers.

### Value

Besides a heatmap illustrating the multiplet testing adjustment corrected P-values from pairwise comparisons in a upper-triangular display, the function also returns a list of two components:

**print.matrix**
A data frame. If K markers are analyzed, the data frame will have K-1 row and K-1 columns. The first column contains the IDs of markers 1:(K-1) while the column names are ID for marker 2:K. The entries in column 3 and following are stat (t statistic), raw P (raw P value) or adjusted P from testing two markers. The 2nd column indicates whether the matching row indicate stat (z-statistic), raw P (raw P value) or adjusted P. So this output basically presen the results in an upper triangular matrix format.

**pval.matrix**
More detailed output. A data frame with the columns: index of marker 1, index of marker 2, ID of marker 1,summary measure estimated for marker 1, ID of marker 2, summary measure estimated for marker 1, the lower limit of the resulting CI on the difference between the summary measure estimates (marker1-marker2), the upper limit of the CI, z-statistic, raw P value and adjusted P value.
Warning

The bootstrapping to obtain the variance on the nonparametric VUS estimate may take a while.

Note

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)

Jingqin Luo

References


See Also

VUS Youden3Grp

Examples

data(AL)
new.AL <- data.frame(group=AL$group,-AL[,,-1])
new.AL <- new.AL[,1:4]##use the first 3 markers as an example

##omnibus test on the 3 AL markers
Pairwise.DiagTest3Grp.Test(dat=new.AL, paired=TRUE, type="VUS", p=0,q=0, conf.level=0.95,p.adjust.method="fdr",digits=3)
### plot.DiagTest3Grp

Scatter plot and boxplot on an object of class `DiagTest3Grp`

**Description**

S3 class method plot for a `DiagTest3Grp` object: Scatter-plot and box-plot.

**Usage**

```r
## S3 method for class 'DiagTest3Grp'
plot(x,...)
```

**Arguments**

- `x`: an object of class `DiagTest3Grp`
- `...`: Other arguments for `plot` function

**Value**

Scatter-plot and Box-plot of the data points, with green, blue, red color corresponding to group $D^-$, $D^0$, $D^+$ and summary measure estimate as well as optimal cut points are indicated.

**Author(s)**

Jingqin Luo

**See Also**

- `vus Youden3Grp`

---

### print.DiagTest3Grp

Print an object of class `DiagTest3Grp`

**Description**

S3 class method print for a `DiagTest3Grp` object: Print relevant information on a `DiagTest3Grp` object.

**Usage**

```r
## S3 method for class 'DiagTest3Grp'
print(x,...)
```
Arguments

- `x`: an object of class `DiagTest3Grp`
- `...`: Other arguments for `plot` function

Value

Print: the type of the three-group diagnostic test summary measure (VUS or Youden); the method used for the estimation; the data summary on group sample size, mean and SD; summary measure estimate and associated Confidence interval (CI); optimal cut-points; the three correct classification probability for each diagnosis group; and sample size (if using "Normal" method) to estimate a summary measure within specified precision.

Author(s)

Jingqin Luo

See Also

`vus` `youdenSgrp` `samplesizeNvus` `SampleSize.VUS` `Sample size for volume under surface (VUS)`

Description

Calculates the sample size required to estimate volume under surface (VUS) within a given margin of error when the diagnostic marker follows normal distributions in each diagnosis group.

Usage

```r
SampleSize.VUS(mu.minus, mu0, mu.plus, s.minus, s0, s.plus, p = 0, q = 0, lam.minus = 1/3, lam0 = 1/3, lam.plus = 1/3, typeError = 0.05, margin = 0.05, subdivisions = 50000, ...)
```

Arguments

- `mu.minus`: A numeric value, the sample mean estimate for the mean parameter of normal distribution in $D^-$.  
- `mu.0`: A numeric value, the sample mean estimate for the mean parameter of normal distribution in $D^0$.  
- `mu.plus`: A numeric value, the sample mean estimate for the mean parameter of normal distribution in $D^+$.  
- `s.minus`: A numeric value, the sample standard deviation (SD) estimate for the SD parameter of normal distribution in $D^-$.  
- `s.0`: A numeric value, the sample SD estimate for the SD parameter of normal distribution in $D^0$.  
- `s.plus`: A numeric value, the sample SD estimate for the SD parameter of normal distribution in $D^+$.  
- `p`: A numeric value, the type I error probability.  
- `q`: A numeric value, the type II error probability.  
- `lam.minus`, `lam.0`, `lam.plus`: A numeric value, the optimal cut-points.  
- `margin`: A numeric value, the margin of error.  
- `subdivisions`: A numeric value, the number of subdivisions.
SampleSize.VUS

s.plus A numeric value, the sample SD estimate for the SD parameter of normal distribution in $D^+$.
p A numeric value, the minimum required specificity, $0 \leq p < 1$, for calculation partial volume under ROC surface. Default, $p=0$.
q A numeric value, the minimum desired sensitivity, $0 \leq q < 1$, for calculation partial volume under ROC surface. Default, $q=0$. $p=q=0$ will give the complete VUS estimate, otherwise give the partial VUS estimate satisfying specificity no less than $p$ and sensitivity no less than $q$.
lam.minus A numeric value, the expected population proportion of the $D_-$ group, used for sample size calculation. Default, lam.minus=1/3. The proportions of the three ordinal groups (lam.minus, lam0, lam.plus) should sum to 1.
lam0 A numeric value, the expected population proportion of the $D_0$ group, used for sample size calculation. Default, lam0=1/3. The proportions of the three ordinal groups (lam.minus, lam0, lam.plus) should sum to 1.
lam.plus A numeric value, the expected population proportion of the $D_+$ group, used for sample size calculation. Default, lam.plus=1/3. The proportions of the three ordinal groups (lam.minus, lam0, lam.plus) should sum to 1.
typeIerror A numeric value, (1-typeIerror)*100% confidence interval (CI) in sample size calculation. Default typeIerror=0.05, i.e., calculate 95% CI.
margin A numeric value, the margin of error on the VUS estimates in sample size calculation. Default, margin=0.05.
subdivisions A numeric value, the number of subintervals for integration using adaptive quadrature in the R function integrate. Default, subdivisions=50000.
... Other arguments that can be passed to the R function integrate, e.g., abs.tol, rel.tol, stop.on.error etc.

Details

The (1-typeIerror)% CI on VUS estimate under normality is $(VUS - Z_\alpha * SE(VUS), VUS + Z_\alpha * SE(VUS))$, then the sample size will be calculated such that: margin = $Z_\alpha * SE(VUS)$ or half of the CI's length, where $Z_\alpha$ is the normal quantile, $Z_\alpha = 1.95$ given default typeIerror $a=0.05$.

Value

Returns a numeric value indicating required sample size.

Note

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)

Jingqin Luo
References


See Also

Normal.VUS

Examples

muminus <- 6
mu0 <- 8
muplus <- 13.36

sminus <- 1
s0 <- 1.2
splus <- 1.4

n <- SampleSize.VUS(mu.minus, mu0, mu.plus, s.minus, s0, s.plus, p=0, q=0, type.error=0.05)

SampleSize.Youden3Grp  Sample size for the extended Youden index for three ordinal groups

Description

Calculates the sample size required to estimate the extended Youden index within a given margin of error when the diagnostic marker follows normal distributions in each diagnosis group.

Usage

SampleSize.Youden3Grp(mu.minus, mu0, mu.plus, s.minus, s0, s.plus, t.minus, t.plus, lam.minus, lam0, lam.plus, type.error = 0.05, margin=0.05)

Arguments

mu.minus  A numeric value, the sample mean estimate for the mean parameter of normal distribution in $D^-$.  
mu0  A numeric value, the sample mean estimate for the mean parameter of normal distribution in $D^0$.  
mu.plus  A numeric value, the sample mean estimate for the mean parameter of normal distribution in $D^+$.  
s.minus  A numeric value, the sample standard deviation (SD) estimate for the SD parameter of normal distribution in $D^-$.  

s0 A numeric value, the sample SD estimate for the SD parameter of normal distribution in $D^0$.

s.plus A numeric value, the sample SD estimate for the SD parameter of normal distribution in $D^+$. 

t.minus A numeric value, the optimal lower cut point to attain the optimal extended Youden index under normal assumptions.

t.plus A numeric value, the optimal upper cut point to attain the optimal extended Youden index under normal assumptions.

lam.minus A numeric value, the expected population proportion of the $D_−$ group, used for sample size calculation. Default, lam.minus=1/3. The proportions of the three ordinal groups (lam.minus, lam0, lam.plus) should sum to 1.

lam0 A numeric value, the expected population proportion of the $D_0$ group, used for sample size calculation. Default, lam0=1/3. The proportions of the three ordinal groups (lam.minus, lam0, lam.plus) should sum to 1.

lam.plus A numeric value, the expected population proportion of the $D_+$ group, used for sample size calculation. Default, lam.plus=1/3. The proportions of the three ordinal groups (lam.minus, lam0, lam.plus) should sum to 1.

typeIerror A numeric value, (1-typeIerror)*100% confidence interval (CI) based on which the length of CI for sample size calculation will be obtained. Default typeIerror=0.05, i.e., calculate 95% CI.

margin A numeric value, the margin of error on the extended Youden index estimates in sample size calculation. Default, margin=0.05.

Details

The (1-typeIerror)% CI on the Youden index ($J$) estimate under normality is $(J - Z_a * SE(J), J + Z_a * SE(J))$, then margin= $Z_a * SE(J)$ or half of the CI’s length, where $Z_a$ is the normal quantile, $Z_a = 1.95$ given default typeIerror $a = 0.05$.

Value

Returns a numeric value indicating required sample size.

Note

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)

Jingqin Luo

References

See Also

Youden3Grp

Examples

```r
mu.minus <- 6
mup0 <- 8
mu.plus <- 13.36

s.minus <- 1
s0 <- 1.2
s.plus <- 1.4

t.minus <- 7.02
t.plus <- 10.52

n <- SampleSize.Youden3Grp(mu.minus, mu0, mu.plus, s.minus, s0, s.plus,
t.minus, t.plus)
```

---

Sp.Sm.Se

*Obtain empirical estimation of the three correct classification probabilities*

Description

once identified the optimal cut-points from VUS or Youden3Grp analyses, use the function to obtain the three empirical correct classification probabilities associated with each diagnosis group.

Usage

```r
Sp.Sm.Se(x, y, z, t.minus, t.plus)
```

Arguments

- `x`: A numeric vector, a diagnostic test’s measurements in the D- (usually healthy subjects).
- `y`: A numeric vector, a diagnostic test’s measurements in the D0 (usually mildly diseased subjects).
- `z`: A numeric vector, a diagnostic test’s measurements in the D+ (usually severely diseased subjects).
- `t.minus`: A numeric value, the lower optimal cut-point identified from VUS or Youden3Grp analyses.
- `t.plus`: A numeric value, the upper optimal cut-point identified from VUS or Youden3Grp analyses.
**Details**

Specificity: \( Sp = Pr(x \leq t_-) \) for \( D^- \) group; Sensitivity: \( Se = Pr(z \geq t_+) \) for \( D^+ \) group and the probability of the diagnostic test for the \( D^0 \) group fall between the two cut points: \( Sm = Pr(t_- \leq y \leq t_+) \). These three probabilities will be estimated empirically.

**Value**

Return a numeric vector with three components \( Sp, Sm \) and \( Se \), the three correct classification probabilities.

**Warning**

The bootstrapping to obtain the variance on the nonparametric VUS estimate may take a while.

**Note**

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

**Author(s)**

Jingqin Luo

**References**


**See Also**

VUS Normal.VUS NonParametric.VUS.var

**Examples**

data(AL)
group <- AL$group

###take the negated kfront marker measurements
kfront <- -AL$kfront

x <- kfront[group=="D-"]
y <- kfront[group=="D0"]
z <- kfront[group=="D+"]

class.prob <- Sp.Sm.Se(x=x,y=y,z=z,t.minus=-1.68,t.plus=0.91)
Description

Calculates the variance estimate of a sample estimate for a normal mean parameter.

Usage

```r
var.mu(s, n)
```

Arguments

- `s` A numeric value, the sample estimate for normal standard deviation.
- `n` A numeric value, the sample size of data.

Details

For a sample dataset of size $n$ drawn from a normal distribution $N(\mu, \sigma^2)$, denote the sample estimate on the mean parameter and standard deviation parameter as $\hat{u}$ and $\hat{s}$, Then the variance of the sample mean is $Var(\hat{u}) = \frac{s^2}{n}$.

Value

A numeric value, the variance of a sample estimate for a normal mean parameter.

Note

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)

Jingqin Luo

References


See Also

`var.sigma`
Examples

```r
s <- 1.5
n <- 100
var.mu(s,n)
```

Description

Calculates the variance of a sample estimate for a normal SD parameter.

Usage

```r
var.sigma(s,n)
```

Arguments

- `s`: A numeric value, the sample standard deviation estimate.
- `n`: A numeric value, the sample size of data.

Details

For a sample dataset of size n drawn from a normal distribution \( N(\mu, \sigma^2) \), denote the sample estimate on standard deviation (SD) parameter as s. Then the variance of the sample SD estimate s is

\[
\text{Var}(s) = \frac{s^2}{2(n-1)}.
\]

Value

A numeric value, the variance estimate of the sample estimate for a normal SD parameter.

Note

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)

Jingqin Luo

References


A wrapper function for Volume under Surface (VUS) estimate, variance estimate under normal and nonparametric assumption and sample size calculation

Description

A wrapper function to calculate the Volume under Surface (VUS) estimate, its variance estimate and optimal cut-point, under normal and nonparametric assumption, to provide partial VUS estimate with a minimum requirement on the specificity and sensitivity under normality and to calculate the sample size under normality to achieve a certain estimation precision on VUS estimate.

Usage

```r
vus(x, y, z, method = c("Normal","NonPar"), p = 0, q = 0, alpha = 0.05,
    nboot = 100, subdivisions = 50000, lam.minus = 1/3, lam0 = 1/3,
    lam.plus = 1/3, typeIerror = 0.05, margin = 0.05,FisherZ=FALSE,
    optimalCut=TRUE,cut.seq=NULL,optimize=FALSE,...)
```

Arguments

- **x**: A numeric vector, a diagnostic test’s measurements in the D- (usually healthy subjects).
- **y**: A numeric vector, a diagnostic test’s measurements in the D0 (usually mildly diseased subjects).
- **z**: A numeric vector, a diagnostic test’s measurements in the D+ (usually severely diseased subjects).
- **method**: A character argument, method =“Normal” or “NonPar”, estimate the VUS under normality or nonparametrically.
- **p**: A numeric value, the minimum required specificity, $0 \leq p < 1$, for calculation partial volume under ROC surface. Default, p=0.
- **q**: A numeric value, the minimum desired sensitivity,$0 \leq q < 1$, for calculation partial volume under ROC surface. Default, q=0. p=q=0 will give the complete VUS estimate, otherwise give the partial VUS estimate satisfying specificity no less than p and sensitivity no less than q.
alpha  A numeric value, (1-alpha)*100% confidence interval of the VUS estimate under normal assumption. Default, alpha=0.05.

nBoot  A numeric value, the total number of bootstrapping used to estimate the variance of nonparametric estimate of VUS.

subdivisions  A numeric value, the number of subintervals for integration using adaptive quadrature in the R function integrate. Default, subdivisions=50000.

lam.minus  A numeric value, the expected population proportion of the $D_{-}$ group, used for sample size calculation. Default, lam.minus=1/3. The proportions of the three ordinal groups (lam.minus, lam0, lam.plus) should sum to 1.

lam0  A numeric value, the expected population proportion of the $D_{0}$ group, used for sample size calculation. Default, lam0=1/3. The proportions of the three ordinal groups (lam.minus, lam0, lam.plus) should sum to 1.

lam.plus  A numeric value, the expected population proportion of the $D_{+}$ group, used for sample size calculation. Default, lam.plus=1/3. The proportions of the three ordinal groups (lam.minus, lam0, lam.plus) should sum to 1.

typeError  A numeric value, (1-typeError)*100% confidence interval (CI) in sample size calculation. Default typeError=0.05, i.e., calculate 95% CI.

margin  A numeric value, the margin of error on the VUS estimates in sample size calculation. Default, margin=0.05. The (1-typeError)% CI on VUS estimate under normality is $\left( VUS - Z_a \ast SE(VUS), VUS + Z_a \ast SE(VUS) \right)$, then margin=$Z_a \ast SE(VUS)$ or half of the CI's length, where $Z_a$ is the normal quantile, $Z_a = 1.95$ given default typeError a=0.05.

FisherZ  Reference to the argument in Normal.VUS.

optimalCut  A logic value of TRUE or FALSE. If TRUE, the function will return optimal cut-point from VUS analyses.

cut.seq  A sequence of numeric values from which the optimal cut-point will be selected from, by default=NULL, will use the unique values of the collection of x,y,z.

optimize  A logical value of TRUE or FALSE. If FALSE, take the empirical optimal cut point identified by empirical search within the given cut.seq as final reported optimal cut point. If TRUE, using the empirical optimal cut point as starting point in optimization algorithm for final optimal cut point.

...  Other arguments that can be passed to the R function integrate, e.g., abs.tol, rel.tol, stop.on.error etc.

Details

For three ordinal group diagnostic test, there are two underlying cut-point $t_{-}$ and $t_{+}$ with $t_{-} < t_{+}$ based on which patients are divided into the three ordinal groups. Patients with a diagnostic test below $t_{-}$ will be assigned to $D_{-}$; those with the test above $t_{+}$ will be assigned to $D_{+}$ and the remaining fall into $D_{0}$. Following the specificity and sensitivity definition as in diagnostic test for two groups, we call the probability of the first two events as specificity $x = P_{-}\{T \leq t_{-}\} = F_{-}(t_{-})$ and sensitivity $y = P_{+}\{T > t_{+}\} = 1 - F_{+}(t_{+}) = G_{+}(t_{+})$ where the $P_{i}$ and $F_{i}$ denotes the probability density function and cumulative density function of a diagnostic test in $D_{i}$, $i = -, 0, +$ separately. Then, the probability that a patient randomly selected from the $D_{0}$ group has the test
result between the two cut-points can be expressed as, \( z = P_0\{t_- \leq T \leq t_+\} = F_0(t_+) - F_0(t_-) = F_0(G^{-1}_+(y)) - F_0(F^{-1}_-(x)) \)

where the notation \( H^{-1}(.) \) denotes the inverse function of \( H \) which is a function of the specificity and sensitivity, i.e., \( z = z(x,y) \), which constitutes a ROC surface in the three-dimensional space \((x,y,z)\). The volume under the ROC surface (VUS) defined by \( z \) can be written as,

\[
V_{00} = \int_0^1 \int_0^{G_+ + (F_0 - 1)} F_0(G_+^{-1}(y)) - F_0(F_0^{-1}(x)) \, dy \, dx
\]

Where the notation \( H^{-1}(.) \) denotes the inverse function of \( H \) and \( z \) is a function of the specificity and sensitivity, i.e., \( z = z(x,y) \), which constitutes a ROC surface in the three-dimensional space \((x,y,z)\).

The value of the VUS is defined by the integral over the region \( D_{00} \) as

- The optimal cut-points from VUS analysis are defined as the one.

Value

A object of `DiagTest3Grp` with a list of components.

- **type**
  - A character value, type=dQuoteVUS for VUS and type=dQuoteYouden for the extended Youden index, indicating which summary measure is outputted.

- **method**
  - A character value. For VUS, method can be “Normal” or “NonPar” (nonparametric); for Youden index, choices are “Normal/EMP/KS/EMP”, indicating which method is used to estimate the summary measure.

- **dat**
  - A list of 3 components. Three components have names “x”, “y”, “z”, each recording the inputted marker measurements (after removing NAs) under \( D^{-}, D_{0}, D^{+} \) respectively.

- **dat.summary**
  - A data frame with 3 rows (D-, D0,D+) and 3 columns (number of observations, mean, SD).

- **estimate**
  - A numerical value. Point estimate for the summary measure, either VUS or Youden.

- **variance**
  - A numeric value. Variance on the summary measure estimate. For normal method, output normal variance; for other methods output variance from bootstrapping.

- **CI**
  - A named numeric vector of length 2. Confidence interval on the summary measure estimate, with name like 2.5%, 97.5% if significance level is set to be 5%. For both VUS and the Youden index, when normal method is in use, the CI is normal CI while bootstrap method was used under other methods.

- **cut.point**
  - A named numeric vector of length 2. Optimal cut-points with name “t.minus” for lower optimal cut point and name “t.plus” for upper optimal point.

- **classify.prob**
  - A named numeric vector of 3 values. Estimates on the three group correct classification probabilities. Specificity on \( D^{-} \): \( Sp = Pr(x \leq t_- | D^{-}) \); sensitivity on \( D^{+} \): \( Se = Pr(z \geq t_+ | D^{+}) \); correct classification probability on \( D^{0} \): \( Sm = SPr(t_- < y < t_+ | D^{0}) \). For VUS, it’s empirical estimation. For Youden index, depending on method adopted for the Youden index estimate, the three probabilities will be estimated using specified method.

- **sampleSize**
  - A numeric value. The sample size to estimate the summary measure within given margin of error and type-I error rate. See `SampleSize.VUS` and `SampleSize.Youden3Grp`.
alpha
A numeric value. The significance level for the CI computation, e.g., default=5%.

typeIerror
A numeric value for type-I error rate, e.g., default=5%.

margin
A numeric value. The margin of errors (precision) to estimate the summary measure s.t. the half the length of the resulting CI is equal to the given margin.

partialDeriv
A numeric data frame with one row and multiple columns, containing relevant parameters (a,b,c,d) and the partial derivatives of VUS estimate w.r.t the relevant parameters which are outputted for performance of statistical tests on markers under normal method or NA under nonparametric method.

Warning
The bootstrapping to obtain the variance on the nonparametric VUS estimate may take a while.

Note
Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)
Jingqin Luo

References


See Also
Normal.VUS NonParametric.VUS NonParametric.VUS.var

Examples

data(AL)
group <- AL$group
table(group)

## take the negated kfront marker measurements
kfront <- -AL$kfront

x <- kfront[group=="D-"]
y <- kfront[group=="D0"]
z <- kfront[group=="D+"]

## normal estimate
normal.res <- VUS(x,y,z,method="Normal",p=0,q=0, alpha=0.05)
normal.res
### nonparametric estimate
### Not run:
nonpar.res <- VUS(x,y,z,method="NonPar",p=0,q=0,alpha=0.05,NBOOT=100)
nonpar.res

### End(Not run)

### S3 method for class 'DiagTest3Grp':
print(normal.res)

### S3 method for class 'DiagTest3Grp':
plot(normal.res)

---

**VUS.CutPoint**

Find the optimal cut-points in volume under ROC surface (VUS) analyses

**Description**

Find the optimal cut-points in volume under ROC surface (VUS) analyses which correspond to the coordinate triplet with the shortest distance to the perfect classification coordinates (1,1,1).

**Usage**

VUS.CutPoint(x, y, z, cut.seq=NULL, optimize=F)

**Arguments**

- **x**
  A numeric vector. A diagnostic test’s measurements in the D- (usually healthy subjects).

- **y**
  A numeric vector. A diagnostic test’s measurements in the D0 (usually mildly diseased subjects).

- **z**
  A numeric vector. A diagnostic test’s measurements in the D+ (usually severely diseased subjects).

- **cut.seq**
  A numeric vector. A sequence of user-specified values from which the optimal cut-points will be chosen. If set to NULL, the unique collection of values of x, y, z will be used as the sequence.

- **optimize**
  A logical value of TRUE or FALSE. If FALSE, take the empirical optimal cut point identified by empirical search within the given cut.seq as final reported optimal cut point. If TRUE, using the empirical optimal cut point as starting point in optimization algorithm for final optimal cut point.
Youden3Grp

Details

The optimal cut-point \( t_{\text{minus}}, t_{\text{plus}} \) is determined based on a pair of values \((t_-, t_+), t_- \leq t_+, \) from cut.seq to minimize the goal function: \( (\Pr(x \leq t_-) - 1)^2 + (\Pr(t_- < y < t_+) - 1)^2 + (\Pr(z \geq t_+) - 1)^2 \). The probabilities are estimated by empirical CDF.

Value

Return a data frame with two columns “t.minus” and “t.plus” as optimal pair of cut-points for ROC surface analysis.

Note

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)

Jingqin Luo

See Also

VUS Normal. VUS Youden3Grp

Examples

data(AL)
group <- AL$group
table(group)

## take the negated kfront marker measurements
kfront <- -AL$kfront

x <- kfront[group=="D-"]
y <- kfront[group=="D0"]
z <- kfront[group=="D+"]

vus.cut <- VUS.CutPoint(x=x,y=y,z=z,cut.seq=NULL)

Youden3Grp

The wrapper function for Youden index analysis

Description

The wrapper function provide not only point estimates but also variances for the extended Youden index for three ordinal groups, the associated optimal cut-points under parametric and non-parametric methods and sample size under normal.
Usage

Youden3Grp(x, y, z, method = c("Normal", "TN", "EMP", "KS", "KS-SJ"), randomStart.N = 1, optim.method = NULL, t.minus.start = NULL, t.plus.start = NULL, lam.minus = 1/3, lam0 = 1/3, lam.plus = 1/3, type.error = 0.05, margin = 0.05, NBOOT = 10, seed.seq = NULL, alpha = 0.05, FisherZ = FALSE, ...)

Arguments

x  A numeric vector, a diagnostic test’s measurements for subjects in D- (usually a healthy group).
y  A numeric vector, a diagnostic test’s measurements for subjects in D0 (usually a mildly diseased group).
z  A numeric vector, a diagnostic test’s measurements for subjects in D+ (usually a severely diseased group).

method  A character. Specify a method to be used for estimating the extended Youden index. Choices include ‘Normal’—Estimate the extended Youden index under the assumption of normal distributions of a diagnostic test in the three groups. ‘TN’—Transformed normal. Implement Box-cox transformation to approximate normality and then estimate the extended Youden index under normality. ‘EMP’—Estimate the extended Youden index by using empirical cumulative density function. ‘KS’—Estimate the extended Youden index by using Kernel density estimation with a normal reference rule for bandwidth selection. ‘KS-SJ’—Estimate the extended Youden index by using Kernel density estimation with the Sheather-Jones Plug-in method for bandwidth selection.

randomStart.N  An argument need to be specified when the method “EMP”, “KS” or “KS-SJ” is used to estimate the optimal cut-point and the extended Youden index J. Default, randomStart.N=1.

optim.method  A character argument. The optimization method specified to derive the optimal cut-points when nonparametric methods are used to estimate the extended Youden index. Default, optim.method=L-BFGS-B. See optim.

t.minus.start  A numeric value. The starting points of the lower optimal cut-point (t-) which separate the D- and D0 group. Default t.minus.start=NULL will randomly generate starting value.

t.plus.start  A numeric value. The starting points of the upper optimal cut-point (t+) which separate the D0 and D+ group. Default t.plus.start=NULL will randomly generate starting value.

lam.minus  A numeric value. The expected population proportion of the D- group, used for sample size calculation. Default, lam.minus=1/3. The proportions of the three ordinal groups (lam.minus, lam0, lam.plus) should sum to 1.

lam0  A numeric value. The expected population proportion of the D0 group, used for sample size calculation. Default, lam0=1/3. The proportions of the three ordinal groups (lam.minus, lam0, lam.plus) should sum to 1.

lam.plus  A numeric value. The expected population proportion of the D+ group, used for sample size calculation. Default, lam.plus=1/3. The proportions of the three ordinal groups (lam.minus, lam0, lam.plus) should sum to 1.
Youden3Grp

```

typeIerror  A numeric value. (1-typeIerror)*100% confidence interval (CI) in sample size calculation. Default typeIerror=0.05, i.e., calculate 95% CI.
margin      A numeric value. The margin of error on the Youden index estimate in sample size calculation. Default, margin=0.05.
NBOOT       A numeric value. Total number of bootstrap samples to draw for bootstrap variance and CI.
seed.seq    a numeric vector of length NBOOT. The sequence of random seeds for bootstrapping procedure to obtain bootstrapping variance and percentile confidence interval if methods other than “Normal” is adopted.
alpha       A numeric value. Significance level for CI calculation. Default, alpha=0.05.
fisherZ     Reference to the argument in `normalNvus`.

Details

Let the marker denoted by T and the density function under Dᵢ as fᵢ, i = −, 0, +. The Youden index for three ordinal groups is defined as, J(t−,t+) = \frac{1}{2}(f_−(T \leq t−) + f_0(t− \leq T \leq t+) + f_+(T \geq t+) − 1). The optimal index achieved at an optimal pair will be adopted as the final summary measure. Note that the package assumes that the marker values increase monotonically and stochastically with disease severity (the marker means in the three group D−, D0 and D+ are monotonically increasing). If a monotonically decreasing trend, the marker should be negated for analysis.

For the extended Youden index and optimal cut-point estimates, see reference paper for details.

For sample size calculation, the (1-typeIerror)% CI on the Youden estimate under normality is (J − Z_a * SE(J), J + Z_a * SE(J)), then the sample size will be calculated such that margin=Z_a * SE(J) or half of the CI’s length, where Z_a is the normal quantile, Z_a = 1.95 given default typeIerror a = 0.05.

Value

A object of `DiagTest3Grp` with a list of components.
```

type          A character value, type=dQuoteVUS for VUS and type=dQuoteYouden for the extended Youden index, indicating which summary measure is outputted.
method        A character value. For VUS, method can be “Normal” or “NonPar” (nonparametric); for Youden index, choices are “Normal/TN/EMP/KS/KS-SJ”, indicating which method is used to estimate the summary measure.
dat           A list of 3 components. Three components have names “x”, “y”, “z”, each recording the inputted marker measurements (after removing NAs) under D−, D0, D+ respectively.
dat.summary   A data frame with 3 rows (D−, D0, D+) and 3 columns (number of observations, mean, SD).
estimate      A numerical value. Point estimate for the summary measure, either VUS or Youden.
```
Youden3Grp

variance A numeric value. Variance on the summary measure estimate. For normal method, output normal variance; for other methods output variance from bootstrapping.

CI A named numeric vector of length 2. confidence interval on the summary measure estimate, with name like 2.5%, 97.5% if significance level is set to be 5%. For both VUS and the Youden index, when normal method is in use, the CI is normal CI while bootstrap method was used under other methods.

cut.point A named numeric vector of length 2. optimal cut-points with name “t.minus” for lower optimal cut point and name “t.plus” for upper optimal point.

classify.prob A named numeric vector of 3 values. Estimates on the three group correct classification probabilities. specificity on $D^-: Sp = Pr(x \leq t_- | D^-)$; sensitivity on $D^+: Se = Pr(z \geq t_+ | D^+)$; correct classification probability on $D^0$: $Sm = SpPr(t_- < y < t_+ | D^0)$. For VUS, it’s empirical estimation. For Youden index, depending on method adopted for the Youden index estimate, the three probabilities will be estimated using specified method.

sampleSize A numeric value The sample size to estimate the summary measure within given margin of error and type-I error rate. See SampleSize, VUS and SampleSize.Youden3Grp.

alpha A numeric value. The significance level for the CI computation, e.g., default=5%.

typeIerror A numeric value for type-I error rate, e.g., default=5%.

margin A numeric value. The margin of errors (precision) to estimate the summary measure s.t. the half the length of the resulting CI is equal to the given margin. Note: variance and CI on the optimal cut-points are calculated but not returned in the final DiagTest3Grp object output.

partialDeriv A numeric data frame with one row and multiple columns, containing estimated SD parameters in each diagnosis group and the partial derivatives of Youden estimate w.r.t the relevant mean and SD parameters which are outputted for performance of statistical tests on markers under normal method or NA under non-parametric method.

Note

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)

Jingqin Luo

References


See Also

Examples

data(AL)
group <- AL$group
table(group)

##take the negated FACTOR1 marker measurements
factor1 <- ~AL$FACTOR1

x <- factor1[group=="D-"]
y <- factor1[group=="D0"]
z <- factor1[group=="D+"]

##Normal method
youden.res <- Youden3Grp(x=x, y=y, z=z, method = "Normal")
print(youden.res)
plot(youden.res)

##Kernel smoothing method with Sheather-Jones plug-in method
## Not run:
youden.res.KS <- Youden3Grp(x=x, y=y, z=z, method = "KS-SJ", randomStart.N=3)
## End(Not run)

Youden3Grp.OptimalCutoff.Normal.trueMeanSD

The Youden index and optimal cut point estimation under normal distribution for three ordinal diagnostic groups

Description

For a diagnostic test following normal distributions, given the true or estimated normal distribution mean and standard deviation (SD) parameters for three ordinal diagnostic groups, calculate the Youden index and optimal cut point estimations for three ordinal diagnostic groups (usually healthy D-, mildly diseased D0 and diseased D+).

Usage

Youden3Grp.OptimalCutoff.Normal.trueMeanSD(mu.minus, mu0, mu.plus, s.minus, s0, s.plus)

Arguments

mu.minus A numeric value. The mean parameter of normal distribution in D- group.
mu0 A numeric value. The mean parameter of normal distribution in D0 group.
mu.plus A numeric value. The mean parameter of normal distribution in D+ group.
s.minus A numeric value. The SD parameter of normal distribution in D- group.
Youden3Grp.OptimalCutoff.Normal.trueMeanSD

s0 A numeric value. The SD parameter of normal distribution in D0 group.
s.plus A numeric value. The SD parameter of normal distribution in D+ group.

Details
See references.

Value
A data frame of numeric components,
t.minus The derived optimal lower cut-point t.minus
t.plus The derived optimal upper cut-point t.minus
youden The derived optimal youden index

Note
Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)
Jingqin Luo

References

See Also
Youden3Grp

Examples

```r
mu.minus <- 6
s.minus <- 1

mu0 <- 8
s0 <- 1.2

mu.plus <- 11.05
s.plus <- 1.4

Youden3Grp.OptimalCutoff.Normal.trueMeanSD(mu.minus,mu0,mu.plus,s.minus,s0,s.plus)
```
Provide point estimates for the extended Youden index, the associated optimal cut-points, and sample size for three ordinal groups

Description

Estimate for the extended Youden index, the associated optimal cut-points using normal and non-parametric methods and the associated. The sample size to estimate the extended Youden index of a diagnostic marker within a given margin of error under normal assumptions is also provided.

Usage

Youden3Grp.PointEst(x, y, z, method = "Normal", randomStart.N=1, optim.method = NULL, t.minus.start = NULL, t.plus.start = NULL, lam.minus = 1/3, lam0 = 1/3, lam.plus = 1/3, typeError = 0.05, margin=0.05, FisherZ=FALSE, ...)

Arguments

- **x**: A numeric vector, a diagnostic test’s measurements for subjects in D- (usually a healthy group).
- **y**: A numeric vector, a diagnostic test’s measurements for subjects in D0 (usually a mildly diseased group).
- **z**: A numeric vector, a diagnostic test’s measurements for subjects in D+ (usually a severely diseased group).
- **method**: A character. Specify a method to be used for estimating the extended Youden index. Choices include ‘Normal’—Estimate the extended Youden index under the assumption of normal distributions of a diagnostic test in the three groups. ‘TN’—Transformed normal. Implement Box-cox transformation to approximate normality and then estimate the extended Youden index under normality. ‘EMP’—Estimate the extended Youden index by using empirical cumulative density function. ‘KS’—Estimate the extended Youden index by using Kernel density estimation with a normal reference rule for bandwidth selection. ‘KS-SJ’—Estimate the extended Youden index by using Kernel density estimation with the Sheather-Jones Plug-in method for bandwidth selection.
- **randomStart.N**: An argument need to be specified when the method “EMP”, “KS” or “KS-SJ” is used to estimate the optimal cut-point and the extended Youden index J. Default, randomStart.N=1.
- **optim.method**: A character argument. The optimization method specified to derive the optimal cut-points when nonparametric methods are used to estimate the extended Youden index. Default, optim.method=L-BFGS-B. see optim.
- **t.minus.start**: A numeric value. The starting points of the lower optimal cut-point (t-) which separate the $D^-$ and $D^0$ group. Default t.minus.start=NULL will randomly generate starting value.
Youden3Grp.PointEst

\textbf{t.plus.start} \quad A numeric value. The starting points of the upper optimal cut-point (t+) which separate the \(D^0\) and \(D^+\) group. Default \(t\text{.plus.start} = \text{NULL}\) will randomly generate starting value.

\textbf{lam.minus} \quad A numeric value. The expected population proportion of the \(D^-\) group, used for sample size calculation. Default, \(\text{lam.minus} = 1/3\). The proportions of the three ordinal groups (\(\text{lam.minus}\), \(\text{lam0}\), \(\text{lam.plus}\)) should sum to 1.

\textbf{lam0} \quad A numeric value. The expected population proportion of the \(D^0\) group, used for sample size calculation. Default, \(\text{lam0} = 1/3\). The proportions of the three ordinal groups (\(\text{lam.minus}\), \(\text{lam0}\), \(\text{lam.plus}\)) should sum to 1.

\textbf{lam.plus} \quad A numeric value. The expected population proportion of the \(D^+\) group, used for sample size calculation. Default, \(\text{lam.plus} = 1/3\). The proportions of the three ordinal groups (\(\text{lam.minus}\), \(\text{lam0}\), \(\text{lam.plus}\)) should sum to 1.

\textbf{typeIerror} \quad A numeric value. \((1-\text{typeIerror})\times 100\%\) confidence interval (CI) in sample size calculation. Default \(\text{typeIerror} = 0.05\), i.e., calculate 95\% CI.

\textbf{margin} \quad A numeric value. The margin of error on the Youden index estimate in sample size calculation. Default, \(\text{margin} = 0.05\).

\textbf{FisherZ} \quad Reference to the argument in \texttt{Normal.VUS}.

... \quad Other options for the R function \texttt{optim}.

\section*{Details}

Let the marker denoted by \(T\) and the density function under \(D^i\) as \(f_i, i = -, 0, +\). The Youden index for three ordinal groups is defined as, \(J(t-, t+) = \frac{1}{2}[f_-(T \leq t-) + f_0(t_+ \leq T \leq t_+) + f_+(T \geq t_+)] - 1\). The optimal index achieved at an optimal pair will be adopted as the final summary measure. Note that the package assumes that the marker values increase monotonically and stochastically with disease severity (the marker means in the three group \(D^-, D^0\) and \(D^+\) are monotonically increasing). If a monotonically decreasing trend, the marker should be negated for analysis.

For the extended Youden index and optimal cut-point estimates, see reference paper for details.

For sample size calculation, the \((1-\text{typeIerror})\%\) CI on the Youden estimate under normality is \((J - Z_a*SE(J), J + Z_a*SE(J))\), then the sample size will be calculated such that \(\text{margin} = Z_a*SE(J)\) or half of the CI’s length, where \(Z_a\) is the normal quantile, \(Z_{0.95} = 1.95\) given default \(\text{typeIerror} = 0.05\).

\section*{Value}

A list of three components. “\text{dat}”: a list of components \(x, y, z\), recording raw data inputted (with NA’s removed); “\text{dat.summary}”: a data frame of three rows (corresponding to \(D^-\), \(D^0\) and \(D^+\)) separately and three columns (number of non-missing observations, mean and SD of raw data); “\text{est}”: a data frame of one row and the following columns numeric value components:

\textbf{mu.minus} \quad The sample mean estimate of the diagnostic’s measurements in the \(D^-\) group, which is sample mean of data if \text{method} = \text{“Normal”}/\text{“EMP”}/\text{“KS”}/\text{“KS-SJ”} while if \text{method} = \text{“TN”}, \text{mu.minus} is the sample mean of the data after Box-cox transformation.
mu0  The sample mean estimate of the a diagnostic’s measurements in the D0 group, which is sample mean of data if method=Normal while if method=TN, mu0 is the sample mean of the data after Box-cox transformation.

mu.plus  The sample mean estimate of the a diagnostic’s measurements in the D+ group, which is sample mean of data if method=Normal while if method=TN, mu.plus is the sample mean of the data after Box-cox transformation.

s.minus  The sample standard deviation (SD) estimate of the a diagnostic’s measurements in the D- group, which is sample SD of data if method=Normal while if method=TN, the s.minus is the sample SD of the data after Box-cox transformation.

s0  The sample standard deviation (SD) estimate of the a diagnostic’s measurements in the D0 group, which is sample SD of data if method=Normal while if method=TN, the s0 is the sample SD of the data after Box-cox transformation.

s.plus  The sample standard deviation (SD) estimate of the a diagnostic’s measurements in the D+ group, which is sample SD of data if method=Normal while if method=TN, the s.plus is the sample SD of the data after Box-cox transformation.

t.minus  The estimate of the lower optimal cut-point.

t.plus  The estimate of the upper optimal cut-point.

Se  Sensitivity, the probability of a randomly selected subject from D+ group has a test result above t.plus.

Sp  Specificity the probability of a randomly selected subject from D- group has a test result below t.minus.

Sm  The probability of a randomly selected subject from D0 group has a test result between t.minus and t.plus.

youden  The extended Youden index for three ordinal groups attained at the optimal cut-point t.minus and t.plus.

youden.z  The Fisher’s Z transformation of the extended Youden index.

lambda.est  The estimated lambda, the parameter used in Box-Cox transformation \( y=(x^\lambda-1)/\lambda \). The estimate will be a numeric value when the method=TN is used to estimate the extended Youden index and NA when other methods were used.

t.minus.TN  The optimal lower cut-point t.minus estimated for the data after Box-Cox transformation. The estimate will be a numeric value when the method=TN is used to estimate the extended Youden index and NA when other methods are used.

t.plus.TN  The optimal upper cut-point t.plus estimated for the data after Box-Cox transformation. The estimate will be a numeric value when the method=TN is used to estimate the extended Youden index and NA when other methods are used.

sampleSize  The sample size on investigating a diagnostic test based on pilot measurements x, y, z. The sample size calculated such that based on given population proportion of the three diagnosis groups, the margin of error on the (1-typeIerror)*100% confidence interval of the extended Youden index estimate under normality is as specified, by default 0.05.
Note

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)

Jingqin Luo

References


See Also

Youden3Grp.Variance.Bootstrap

Examples

data(AL)
group <- AL$group
table(group)

##take the negated FACTOR1 marker measurements
factor1 <- -AL$FACTOR1

x <- factor1[group=="D-"]
y <- factor1[group=="D0"]
z <- factor1[group=="D+"]

##Normal method
youden.res.normal <- Youden3Grp.PointEst(x=x, y=y, z=z, method = "Normal")
youden.res.normal$est$youden

##Kernel smoothing method with Sheather-Jones plug-in method
youden.res.KS <- Youden3Grp.PointEst(x=x, y=y, z=z, method = "KS-SJ",randomStart.N=3)
youden.res.KS$est$youden

Youden3Grp.Variance.Bootstrap

Obtain bootstrap variance and confidence interval on the extended Youden index and optimal cut-off points

Description

Given the diagnostic test measurements x, y, z in the three ordinal groups $D^-, D^0, D^+$ separately, the function repeatedly draws a bootstrap sample each of x, y, z to estimate the extended Youden index for three ordinal groups and the associated optimal cut point and obtain the final bootstrap average estimate on the Youden index and optimal cut-off point and the confidence interval.
Usage

Youden3Grp.Variance.Bootstrap(x, y, z, method = "Normal", seed.seq = NULL, randomStart.N = 5, NBOOT=10, alpha=0.05, t.minus.start = NULL, t.plus.start = NULL, ...)

Arguments

x  A numeric vector, a diagnostic test’s measurements for subjects in D- (usually a healthy group).
y  A numeric vector, a diagnostic test’s measurements for subjects in D0 (usually a mildly diseased group).
z  A numeric vector, a diagnostic test’s measurements for subjects in D+ (usually a severely diseased group).
method  A character argument. Specify a method to be used for estimating the extended Youden index. Choices include ‘Normal’—Estimate the extended Youden index under the assumption of normal distributions of a diagnostic test in the three groups.
       ‘TN’—Transformed normal. Implement Box-cox transformation to approximate normality and then estimate the extended Youden index under normality.
       ‘EMP’—Estimate the extended Youden index by using empirical cumulative density function.
       ‘KS’—Estimate the extended Youden index by using Kernel density estimation with a normal reference rule for bandwidth selection.
seed.seq  A numeric vector. Users can specify a sequence of random seeds for bootstrap sampling of x. Bootstrapping of y and z will use seed0+1 and seed0+2. Default will be the sequence of 1:10.
randomStart.N  A numeric value. An argument need to be specified when the method “EMP”, “KS” or “KS-SJ” is used to estimate the optimal cut-point and the extended Youden index J. Default, randomStart.N=1. See Youden3Grp.
NBOOT  A numeric value. Total number of bootstrapping, default=10.
t.minus.start  The starting points of the lower optimal cut-point (t-) which separate the D- and D0 group. Default t.minus.start=NULL will randomly generate starting value. See Youden3Grp.
t.plus.start  A numeric value. The starting points of the upper optimal cut-point (t+) which separate the D0 and D+ group. Default t.plus.start=NULL will randomly generate starting value. See Youden3Grp.
alpha  A numeric value. The significance level, will provide the basic quantile confidence interval (alpha/2*100, 1-alpha/2*100).
...  Other arguments to be passed to the R function Youden3Grp.

Details

This function is carried to get bootstrap estimates of the extended Youden index and associated cut-points to provide confidence interval. See details in Youden3Grp
Value

Return a list, including the following components

- **var.youden**: The normal-method based variance on the optimal Youden index.
- **var.t.minus**: A numeric value. The variance on the lower optimal cut-point t.minus.
- **var.t.plus**: A numeric value. The variance on the upper optimal cut-point t.plus.
- **var.youden.z**: A numeric value. The variance on the Fisher’s Z transformed optimal Youden index.
- **youden.CI**: A named numeric of length 2. CI for the estimate of youden with the lower and the upper CI limit (under the names alpha/2*100% and (1-alpha/2)*100%).
- **t.minus.CI**: A named numeric of length 2. CI for the estimate of t.minus (the lower optimal cut-point) with the lower and the upper CI limit (under the names alpha/2*100% and (1-alpha/2)*100%).
- **t.plus.CI**: A named numeric of length 2. CI for the estimate of t.plus (the upper optimal cut-point) with the lower and the upper CI limit (under the names alpha/2*100% and (1-alpha/2)*100%).
- **youden.z.CI**: A named numeric of length 2. CI for the estimate of Fisher-Z transformed youden with the lower and the upper CI limit (under the names alpha/2*100% and (1-alpha/2)*100%).

Note

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)

Jingqin Luo

References


See Also

Youden3Grp.BW.ref

Examples

```r
x <- rnorm(50,6,1)
y <- rnorm(60,8,1.2)
z <- rnorm(40,10,1.4)

temp.res <- Youden3Grp.Variance.Bootstrap(x=x, y=y, z=z,
method="Normal", seed.seq=1:10, randomStart.N=3, NBOOT=10)
```
Description

Calculate the variance estimation for the extended Youden index and the variance estimation for the associated lower and upper optimal cut-point assuming that a diagnostic test follows normal distributions in the three ordinal groups \((D^-, D^0, D^+).\)

Usage

Youden3Grp.Variance.Normal(x, y, z, alpha = 0.05)

Arguments

- **x**: A numeric vector. A diagnostic test’s measurements in the \(D_-\) (usually healthy subjects).
- **y**: A numeric vector. A diagnostic test’s measurements in the \(D^0\) (usually mildly diseased subjects).
- **z**: A numeric vector. A diagnostic test’s measurements in the \(D^+\) (usually severely diseased subjects).
- **alpha**: A numeric value. Significance level so that the function calculates the \((1-alpha)*100\%\) confidence interval (CI) on the estimates of the extended Youden index and optimal cut-points under normal assumption. Default, \(alpha=0.05\).

Details

See details in Youden3Grp.

Value

Return a list, including the following components

- **var.youden**: The normal-method based variance on the optimal Youden index.
- **var.t.minus**: A numeric value. The variance on the lower optimal cut-point \(t.minus\).
- **var.t.plus**: A numeric value. The variance on the upper optimal cut-point \(t.plus\).
- **var.youden.z**: A numeric value. The variance on the Fisher’s Z transformed optimal Youden index.
- **youden.CI**: A named numeric of length 2. CI for the estimate of youden with the lower and the upper CI limit (under the names alpha/2*100\% and (1-alpha/2)*100\%).
- **t.minus.CI**: A named numeric of length 2. CI for the estimate of \(t.minus\) (the lower optimal cut-point) with the lower and the upper CI limit (under the names alpha/2*100\% and (1-alpha/2)*100\%) CI
Youden3Grp.Variance.Normal

t.plus CI  A named numeric of length 2. CI for the estimate of t plus (the upper optimal cut-point) with the lower and the upper CI limit (under the names alpha/2*100% and (1-alpha/2)*100%).

youden.z.CI  A named numeric of length 2. CI for the estimate of Fisher-Z transformed youden with the lower and the upper CI limit (under the names alpha/2*100% and (1-alpha/2)*100%).

partialDeriv  A numeric data frame with one row and multiple columns, containing estimated SD parameters in each diagnosis group and the partial derivatives of Youden estimate w.r.t the relevant mean and SD parameters which are outputted for performance of statistical tests on markers under normal method or NA under non-parametric method.

Note

Bug reports, malfunctioning, or suggestions for further improvements or contributions can be sent to Jingqin Luo <rosy@wubios.wustl.edu>.

Author(s)

Jingqin Luo

References


See Also

Youden3Grp Youden3Grp.Variance.Bootstrap

Examples

data(AL)
group <- AL$group
table(group)

## take the negated FACTOR1 marker measurements
factor1 <- -AL$FACTOR1

x <- factor1[group=="D-" ]
y <- factor1[group=="D0" ]
z <- factor1[group=="D+" ]

temp.res <- Youden3Grp.Variance.Normal(x=x, y=y, z=z, alpha=0.05)

## variance of the extended Youden index and optimal cut-points
temp.res[ c("var.youden","var.t.minus","var.t.plus") ]
Index

*Topic datagen
  bootsample, 5
*Topic datasets
  AL, 4
*Topic design
  SampleSize. VUS, 26
  SampleSize. Youden3Grp, 28
  VUS, 34
  Youden3Grp, 39
  Youden3Grp. PointEst, 45
*Topic hplot
  plot.DiagTest3Grp, 25
*Topic htest
  BW. ref, 6
  CI. normal, 7
  DigTest3Grp. Test, 8
  NonParametric. VUS, 15
  NonParametric. VUS.var, 17
  Normal. VUS, 19
  Pairwise. DigTest3Grp. Test, 22
  var. mu, 32
  var. sigma, 33
  VUS, 34
  Youden3Grp, 39
  Youden3Grp. OptimalCutoff. Normal. trueMeanSD, 43
  Youden3Grp. PointEst, 45
  Youden3Grp. Variance. Bootstrap, 48
  Youden3Grp. Variance. Normal, 51
*Topic manip
  FisherZ, 12
  FisherZ. Var, 13
*Topic misc
  FisherZ, 12
  FisherZ. Var, 13
*Topic nonparametric
  empirical.cdf, 11
  KernelSmoothing. cdf, 14
  NonParametric. VUS, 15
NonParametric. VUS. var, 17
VUS, 34
Youden3Grp, 39
Youden3Grp. PointEst, 45
*Topic optimize
  VUS. CutPoint, 38
  Youden3Grp, 39
  Youden3Grp. PointEst, 45
*Topic package
  DigTest3Grp-package, 2
*Topic print
  print. DigTest3Grp, 25
*Topic smooth
  KernelSmoothing. cdf, 14
  Youden3Grp, 39
  Youden3Grp. PointEst, 45
*Topic univar
  BW. ref, 6
  empirical.cdf, 11
  FisherZ. Var, 13
  Normal. VUS, 19
  SampleSize. VUS, 26
  SampleSize. Youden3Grp, 28
  var. mu, 32
  var. sigma, 33
  Youden3Grp. Variance. Bootstrap, 48
  Youden3Grp. Variance. Normal, 51
AL, 4
bootsample, 5
BW. ref, 6, 15, 50
CI. normal, 7
DigTest3Grp, 25, 36, 41
DigTest3Grp (DigTest3Grp-package), 2
DigTest3Grp-package, 2
DigTest3Grp. Test, 8, 23
empirical.cdf, 11
FisherZ, 12, 14
FisherZ.Var, 13, 13

integrate, 19, 20, 27, 35

KernelSmoothing.cdf, 7, 12, 14
NonParametric.VUS, 15, 21, 37
NonParametric.VUS.var, 16, 17, 21, 31, 37
Normal.VUS, 16, 18, 19, 28, 31, 35, 37, 39, 41, 46

optim, 40, 41, 45, 46
Pairwise.DiagTest3Grp.Test, 22
plot, 25, 26
plot.DiagTest3Grp, 25
print.DiagTest3Grp, 25

SampleSize.VUS, 3, 26, 36, 42
SampleSize.Youden3Grp, 3, 28, 36, 42
Sp.Sm.Se, 30

var.mu, 32
var.sigma, 32, 33, 34
VUS, 2, 3, 11, 16, 18, 21, 24–26, 31, 34, 39, 42
VUS.CutPoint, 38

Youden3Grp, 3, 7, 8, 11, 12, 15, 24–26, 30, 39, 39, 44, 49–52
Youden3Grp.OptimalCutoff.Normal.trueMeanSD, 43
Youden3Grp.PointEst, 42, 45
Youden3Grp.Variance.Bootstrap, 3, 8, 18, 42, 48, 48, 52