

Package ‘R2ROC’

July 4, 2023

Title AUC Statistics

Version 0.0.0.9000

Description AUC statistic for significance test. Variance and covariance of AUC values used to assess the 95% CI and p-value of the AUC difference for both nested and non-nested model.

License `use_mit_license()`, `use_gpl3_license()` or friends to pick a license

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.2

Depends R (>= 2.10)

LazyData true

R topics documented:

auc_diff	1
auc_var	4
dat1	5
dat2	5
olkin_auc1	6
olkin_auc12	6
olkin_auc12_1	7
olkin_auc12_13	7
olkin_auc12_3	8
olkin_auc12_34	8
olkin_auc1_2	9

Index	10
--------------	-----------

auc_diff	<i>auc_diff function</i>
----------	--------------------------

Description

This function estimates $\text{var}(\text{AUC}(y \sim x[,v1]) - \text{AUC}(y \sim x[,v2]))$ where AUC is the Area Under ROC curve of the model, y is N by 1 matrix having the dependent variable, and x is N by M matrix having M explanatory variables. v1 or v2 indicates the ith column in the x matrix (v1 or v2 can be multiple values between 1 - M, see Arguments below)

Usage

```
auc_diff(dat, v1, v2, nv, kv)
```

Arguments

dat	N by (M+1) matrix having variables in the order of cbind(y,x)
v1	This can be set as v1=c(1) or v1=c(1,2)
v2	This can be set as v2=c(2), v2=c(3), v2=c(1,3) or v2=c(3,4)
nv	Sample size
kv	Population prevalence

Value

This function will estimate significant difference between two PRS (either dependent or independent and joint or single). To get the test statistics for the difference between $AUC(y \sim x[,v1])$ and $AUC(y \sim x[,v2])$. (here we define $AUC1 = AUC(y \sim x[,v1])$ and $AUC2 = AUC(y \sim x[,v2])$). The outputs are listed as follows.

mean_diff	AUC differences between AUC1 and AUC2
var	Variances of AUC differences
upper_diff	Upper value of the differences
lower_diff	Upper value of the differences
p	two tailed P-value for significant difference between AUC1 and AUC2
p_one_tail	one tailed P-value for significant difference
heller_p	P-value based on Heller's test for significant difference
heller_upper_diff	Upper limit of 95% CI for the difference basedon Heller's test
heller_lower_diff	Lower limit of 95% CI for the difference basedon Heller's test

Examples

```
#To get the test statistics for the difference between AUC(y=x[,v1]) and AUC(y=x[,v2])

dat=dat1 #(this example embedded within the package)
nv=length(dat$V1)
kv=0.2 #proportion of cases (prevalence)
v1=c(1)
v2=c(2)
output=auc_diff(dat,v1,v2,nv,kv)

#R2ROC output
#output$mean_diff (mean difference of AUC1 and AUC2)
#0.1344427

#output$var (variance of AUC difference)
#5.467435e-05

#output$upper_diff (upper limit of 95% CI for difference)
#0.1489353
```

```
#output$lower_diff (lower limit of 95% CI for difference)
#0.11995

#output$p (two-tailed P-value for the differences is
#significantly different from zero)
#7.14694e-74

#output$p_one_tail (one-tailed P-value for the differences
#is significantly different from zero)
#3.57347e-74

#To get the test statistics for the difference between
AUC(y=x[,v1]+x[,v2]) and AUC(y=x[,v2])

dat=dat1 #(this example embedded within the package)
nv=length(dat$V1)
kv=0.2 #proportion of cases (prevalence)
v1=c(1,2)
v2=c(2)
output=auc_diff(dat,v1,v2,nv,kv)

#R2ROC output
#output$mean_diff (mean difference of AUC1 and AUC2)
#0.1374099

#output$var (variance of AUC difference)
#7.037293e-05

#output$upper_diff (upper limit of 95% CI for difference)
#0.1538521

#output$lower_diff (lower limit of 95% CI for difference)
#0.1209678

#output$p (two-tailed P-value for the differences is
#significantly different from zero)
#2.655147e-60

#output$p_one_tail (one-tailed P-value for the differences
#is significantly different from zero)
#1.327574e-60

#output$heller_p (two-tailed P-value based on Hellers test
#for the differences is significantly different from zero)
#2.180027e-235

#output$heller_upper_diff (upper limit of 95% CI for
#difference based on Hellers test)
#0.1543439

#output$heller_lower_diff (lower limit of 95% CI for
#difference based on Hellers test)
#0.1214596
```

auc_var	<i>auc_var function</i>
---------	-------------------------

Description

This function estimates $\text{var}(\text{AUC}(y \sim x[,v1]))$ where AUC is the Area Under ROC curve of the model, y is N by 1 matrix having the dependent variable, and x is N by M matrix having M explanatory variables. $v1$ indicates the i th column in the x matrix ($v1$ can be multiple values between $1 - M$, see Arguments below)

Usage

```
auc_var(dat, v1, nv, kv)
```

Arguments

<code>dat</code>	N by $(M+1)$ matrix having variables in the order of <code>cbind(y,x)</code>
<code>v1</code>	This can be set as <code>v1=c(1)</code> , <code>v1=c(1,2)</code> or possibly with more values
<code>nv</code>	Sample size
<code>kv</code>	Population prevalence

Value

This function will test the null hypothesis for AUC. To get the test statistics for $\text{AUC}(y \sim x[,v1])$. The outputs are listed as follows.

<code>auc</code>	AUC
<code>var</code>	Variance of AUC
<code>upper_auc</code>	Upper limit of 95% CI for AUC
<code>lower_auc</code>	Lower limit of 95% CI for AUC

Examples

```
#To get the test statistics for the difference between AUC(y=x[,v1]) and AUC(y=x[,v2])

dat=dat1 #(this example embedded within the package)
nv=length(dat$V1)
kv=0.2 #proportion of cases (prevalence)
v1=c(1)
output=auc_var(dat,v1,nv,kv)
#R2ROC output

#output$auc (AUC)
#0.6821015

#output$var (variance of AUC)
#4.462817e-05

#output$upper_auc (upper limit of 95% CI for AUC)
#0.6951951

#output$lower_auc (lower limit of 95% CI for AUC)
#0.6690079
```

dat1*Raw phenotypes and 2 sets of discovery PGSs*

Description

A dataset containing phenotypes and multiple PGSs estimated two independent discovery population

Usage

dat1

Format

A data frame with 10000 rows and 3 variables:

V1 Phenotype, raw case-control data

V2 PGS1, for discovery population 1

V3 PGS2, for discovery population 2

dat2*Pre-adjusted phenotypes and 2 sets of discovery PGSs*

Description

A dataset containing phenotypes and multiple PGSs estimated two independent discovery population

Usage

dat2

Format

A data frame with 10000 rows and 3 variables:

V1 Phenotype, preadjusted case-control data

V2 PGS1, for discovery population 1

V3 PGS2, for discovery population 2

olkin_auc1	<i>olkin_auc1 function</i>
------------	----------------------------

Description

olkin_auc1 function

Usage

```
olkin_auc1(omat, nv, kv)
```

Arguments

omat	3 by 3 matrix having the correlation coefficients between y, x1 and x2, i.e. omat=cor(dat) where dat is N by 3 matrix having variables in the order of cbind(y,x1,x2)
nv	Sample size

Value

This function will be used as source code

olkin_auc12	<i>olkin_auc12 function</i>
-------------	-----------------------------

Description

olkin_auc12 function

Usage

```
olkin_auc12(omat, nv, kv)
```

Arguments

omat	3 by 3 matrix having the correlation coefficients between y, x1 and x2, i.e. omat=cor(dat) where dat is N by 3 matrix having variables in the order of cbind(y,x1,x2)
nv	Sample size

Value

This function will be used as source code

olkin_auc12_1	<i>olkin_auc12_1 function</i>
---------------	-------------------------------

Description

olkin_auc12_1 function

Usage

```
olkin_auc12_1(omat, nv, kv)
```

Arguments

omat	3 by 3 matrix having the correlation coefficients between y, x1 and x2, i.e. $\text{omat} = \text{cor}(\text{dat})$ where dat is N by 3 matrix having variables in the order of cbind(y,x1,x2)
nv	Sample size

Value

This function will be used as source code

olkin_auc12_13	<i>olkin_auc12_13 function</i>
----------------	--------------------------------

Description

olkin_auc12_13 function

Usage

```
olkin_auc12_13(omat, nv, kv)
```

Arguments

omat	3 by 3 matrix having the correlation coefficients between y, x1 and x2, i.e. $\text{omat} = \text{cor}(\text{dat})$ where dat is N by 3 matrix having variables in the order of cbind(y,x1,x2)
nv	Sample size

Value

This function will be used as source code

olkin_auc12_3	<i>olkin_auc12_3 function</i>
---------------	-------------------------------

Description

olkin_auc12_3 function

Usage

```
olkin_auc12_3(omat, nv, kv)
```

Arguments

omat	3 by 3 matrix having the correlation coefficients between y, x1 and x2, i.e. $\text{omat} = \text{cor}(\text{dat})$ where dat is N by 3 matrix having variables in the order of cbind(y,x1,x2)
nv	Sample size

Value

This function will be used as source code

olkin_auc12_34	<i>olkin_auc12_34 function</i>
----------------	--------------------------------

Description

olkin_auc12_34 function

Usage

```
olkin_auc12_34(omat, nv, kv)
```

Arguments

omat	3 by 3 matrix having the correlation coefficients between y, x1 and x2, i.e. $\text{omat} = \text{cor}(\text{dat})$ where dat is N by 3 matrix having variables in the order of cbind(y,x1,x2)
nv	Sample size

Value

This function will be used as source code

olkin_auc1_2	<i>olkin_auc1_2 function</i>
--------------	------------------------------

Description

olkin_auc1_2 function

Usage

```
olkin_auc1_2(omat, nv, kv)
```

Arguments

omat	3 by 3 matrix having the correlation coefficients between y, x1 and x2, i.e. $omat = cor(dat)$ where dat is N by 3 matrix having variables in the order of cbind(y,x1,x2)
nv	Sample size

Value

This function will be used as source code

Index

*Topic **R2**

auc_diff, 1

auc_var, 4

*Topic **datasets**

dat1, 5

dat2, 5

*Topic **information**

auc_diff, 1

auc_var, 4

*Topic **matrix**

auc_diff, 1

auc_var, 4

*Topic **source**

olkin_auc1, 6

olkin_auc12, 6

olkin_auc12_1, 7

olkin_auc12_13, 7

olkin_auc12_3, 8

olkin_auc12_34, 8

olkin_auc1_2, 9

*Topic **variance**

auc_diff, 1

auc_var, 4

auc_diff, 1

auc_var, 4

dat1, 5

dat2, 5

olkin_auc1, 6

olkin_auc12, 6

olkin_auc12_1, 7

olkin_auc12_13, 7

olkin_auc12_3, 8

olkin_auc12_34, 8

olkin_auc1_2, 9