

Package ‘R2ROC’

July 5, 2023

Title AUC Statistics

Version 1.0.0

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 GPL (>=3)

URL <https://github.com/mommy003/R2ROC>

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.2

Depends R (>= 2.10)

LazyData true

R topics documented:

auc_diff	2
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

*auc_diff function***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 i th 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

<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> or <code>v1=c(1,2)</code>
<code>v2</code>	This can be set as <code>v2=c(2)</code> , <code>v2=c(3)</code> , <code>v2=c(1,3)</code> or <code>v2=c(3,4)</code>
<code>nv</code>	Sample size
<code>kv</code>	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 $\text{AUC}(y \sim x[,v1])$ and $\text{AUC}(y \sim x[,v2])$ (here we define $\text{AUC1} = \text{AUC}(y \sim x[,v1])$ and $\text{AUC2} = \text{AUC}(y \sim x[,v2])$). The outputs are listed as follows.

<code>mean_diff</code>	AUC differences between AUC1 and AUC2
<code>var</code>	Variances of AUC differences
<code>upper_diff</code>	Upper value of the differences
<code>lower_diff</code>	Upper value of the differences
<code>p</code>	Two tailed P-value for significant difference between AUC1 and AUC2
<code>p_one_tail</code>	One tailed P-value for significant difference
<code>heller_p</code>	P-value based on Heller's test for significant difference
<code>heller_upper_diff</code>	Upper limit of 95% CI for the difference basedon Heller's test
<code>heller_lower_diff</code>	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$V1)
kv=sum(dat$V1)/length(dat$V1) # pop. 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)
#0.1756046

#output$var (variance of AUC difference)
#9.274356e-05

#output$upper_diff (upper limit of 95% CI for difference)
#0.1944801

#output$lower_diff (lower limit of 95% CI for difference)
#0.1567292

#output$p (two-tailed P-value for the differences is
#significantly different from zero)
#2.747031e-74

#output$p_one_tail (one-tailed P-value for the differences
#is significantly different from zero)
#1.373515e-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$V1)
kv=sum(dat$V1)/length(dat$V1)# pop. 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)
#0.1793682

#output$var (variance of AUC difference)
#0.0001190366

#output$upper_diff (upper limit of 95% CI for difference)
#0.2007526

#output$lower_diff (lower limit of 95% CI for difference)
#0.1579839

#output$p (two-tailed P-value for the differences is
#significantly different from zero)
#9.87014e-61

#output$p_one_tail (one-tailed P-value for the differences
#is significantly different from zero)
#4.93507e-61

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

```

```
#4.2085e-237

#output$heller_upper_diff (upper limit of 95% CI for
#difference based on Hellers test)
#0.2013899

#output$heller_lower_diff (lower limit of 95% CI for
#difference based on Hellers test)
#0.1586212
```

auc_var

auc_var function

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

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

auc	AUC
var	Variance of AUC
upper_auc	Upper limit of 95% CI for AUC
lower_auc	Lower limit of 95% CI for AUC

Examples

```
#To get the AUC for  $\text{AUC}(y \sim x[,v1])$ 

dat=dat1 #(this example embedded within the package)
nv=length(dat$V1)
kv=sum(dat$V1)/length(dat$V1) # pop. prevalence
v1=c(1)
output=auc_var(dat,v1,nv,kv)
```

```
#R2ROC output
#output$auc (AUC)
#0.7390354

#output$var (variance of AUC)
#7.193337e-05

#output$upper_auc (upper limit of 95% CI for AUC)
#0.7556589

#output$lower_auc (lower limit of 95% CI for AUC)
#0.7224119
```

dat1

*Raw phenotypes and 2 sets of discovery PGSs***Description**

A dataset containing raw 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 preadjusted 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. $\text{omat} = \text{cor}(\text{dat})$ where dat is N by 3 matrix having variables in the order of cbind(y,x1,x2)
nv	Sample size
kv	Population prevalence

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. $\text{omat} = \text{cor}(\text{dat})$ where dat is N by 3 matrix having variables in the order of cbind(y,x1,x2)
nv	Sample size
kv	Population prevalence

Value

This function will be used as source code

olkin_auc12_1 *olkin_auc12_1 function*

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
kv	Population prevalence

Value

This function will be used as source code

olkin_auc12_13 *olkin_auc12_13 function*

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
kv	Population prevalence

Value

This function will be used as source code

olkin_auc12_3 *olkin_auc12_3 function*

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
kv	Population prevalence

Value
This function will be used as source code

olkin_auc12_34 *olkin_auc12_34 function*

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
kv	Population prevalence

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
kv	Population prevalence

Value

This function will be used as source code

Index

*Topic **R2**

auc_diff, [2](#)

auc_var, [4](#)

*Topic **datasets**

dat1, [5](#)

dat2, [5](#)

*Topic **information**

auc_diff, [2](#)

auc_var, [4](#)

*Topic **matrix**

auc_diff, [2](#)

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, [2](#)

auc_var, [4](#)

auc_diff, [2](#)

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](#)