Package 'R2ROC'

January 20, 2025

Title AUC Statistics

Version 1.0.1

Description Area under the receiver operating characteristic curves (AUC) statistic for significance test. Variance and covariance of AUC values used to assess the 95% Confidence interval (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

RoxygenNote 7.1.2

Depends R (>= 2.10)

LazyData true

NeedsCompilation no

Author Hong Lee [aut, cph],
Moksedul Momin [aut, cre, cph]

Maintainer Moksedul Momin <cvasu.momin@gmail.com>

Repository CRAN

Date/Publication 2023-11-26 03:10:02 UTC

Contents

auc_diff 2
auc_trf 4
auc_var 5
dat1 7
dat2 7
olkin_auc1 8
olkin_auc12 8
olkin_auc12_1 9
olkin_auc12_13 9
olkin_auc12_3 10
olkin_auc12_34 10
olkin_auc1_2 11

Index 12

`auc_diff` *auc_diff function*

Description

This function estimates var(AUC(y~x[,v1]) - AUC(y~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~x[,v1]) and AUC(y~x[,v2]) (here we define AUC1=AUC(y~x[,v1])) and AUC2=AUC(y~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[,1])
#and AUC(y=x[,2])
dat=dat1 #(this example embedded within the package)
nv=length(dat$V1)
kv=sum(dat$V1)/length(dat$V1)# pop. prevalence estimated from data
#R2ROC also allows users to estimate AUC using pre-adjusted phenotype
#In that case, users need to specify kv
#eg. kv=0.10 for dat2 (dat2 embedded within the package)
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[,1]+x[,2]) and AUC(y=x[,2])
dat=dat1 #(this example embedded within the package)
nv=length(dat$V1)
kv=sum(dat$V1)/length(dat$V1)# pop. prevalence estimated from data
#R2ROC also allows users to estimate AUC using pre-adjusted phenotype
#In that case, users need to specify kv
#eg. kv=0.10 for dat2 (dat2 embedded within the package)
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_trf` *auc_trf function*

Description

This function transforms the observed scale predictive ability (R2) and its standard error (SE) to AUC with its SE

Usage

```
auc_trf(R2, se, kv)
```

Arguments

`R2`	R2 or coefficient of determination on the observed scale
`se`	Standard error of R2
`kv`	Population prevalence

Value

This function will transform the observed R2 and its s.e between to AUC. Output from the command is the lists of outcomes.

`auc`	Transformed AUC
`se`	SE of transformed AUC

References

Wray, Naomi R., et al. "The genetic interpretation of area under the ROC curve in genomic profiling." PLoS genetics 6.2 (2010): e1000864.

Lee, Sang Hong, et al. "A better coefficient of determination for genetic profile analysis." Genetic epidemiology 36.3 (2012): 214-224.

Examples

```
#To get the transformed AUC

output=auc_trf(0.04, 0.002, 0.05)
output

#output$auc (transformed AUC)
#0.7522887

#output$se (se of transformed AUC)
#0.005948364
```

`auc_var` *auc_var function*

Description

This function estimates var(AUC(y~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 ith 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 cbind(y,x)
`v1`	This can be set as v1=c(1), v1=c(1,2) 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 AUC(y~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
`p`	two tailed p-value
`p_one_tail`	one tailed p-value

Examples

```
#To get the AUC for AUC(y=x[,1])

dat=dat1 #(this example embedded within the package)
nv=length(dat$V1)
kv=sum(dat$V1)/length(dat$V1)# pop. prevalence estimated from data
#R2ROC also allows users to estimate AUC using pre-adjusted phenotype
#In that case, users need to specify kv
#eg. kv=0.10 for dat2 (dat2 embedded within the package)
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

#output$p
#9.28062e-175 (two tailed P-value for the AUC is significantly
#different from 0.5)

#output$p_one_tail (one tailed P-value for the AUC is significantly
#different from 0.5)
#4.64031e-175
```

`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-caontrol 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, preadjustde case-caontrol data

V2 PGS1, for discovery population 1

V3 PGS2, for discovery population 2

`olkin_auc1` *olkin_auc1 function*

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
`kv`	Population prevalance

Value

This function will be used as source code

`olkin_auc12` *olkin_auc12 function*

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
`kv`	Population prevalance

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. 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 prevalance

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. 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 prevalance

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. 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 prevalance

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. 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 prevalance

Value

This function will be used as source code

`olkin_auc1_2` *olkin_auc1_2 function*

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 prevalance

Value

This function will be used as source code

Index

∗ **AUC**
    auc_trf, 4
∗ **R2**
    auc_diff, 2
    auc_trf, 4
    auc_var, 5
∗ **Transformation**
    auc_trf, 4
∗ **datasets**
    dat1, 7
    dat2, 7
∗ **information**
    auc_diff, 2
    auc_var, 5
∗ **matrix**
    auc_diff, 2
    auc_var, 5
∗ **observed**
    auc_trf, 4
∗ **of**
    auc_trf, 4
∗ **source**
    olkin_auc1, 8
    olkin_auc12, 8
    olkin_auc12_1, 9
    olkin_auc12_13, 9
    olkin_auc12_3, 10
    olkin_auc12_34, 10
    olkin_auc1_2, 11
∗ **to**
    auc_trf, 4
∗ **variance**
    auc_diff, 2
    auc_var, 5

auc_diff, 2
auc_trf, 4
auc_var, 5

dat1, 7
dat2, 7

olkin_auc1, 8
olkin_auc12, 8
olkin_auc12_1, 9
olkin_auc12_13, 9
olkin_auc12_3, 10
olkin_auc12_34, 10
olkin_auc1_2, 11