Package ‘r2redux’

May 27, 2022

Title R2 Statistic

Version 1.0.4

Author Hong Lee, Moksedul Momin

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

Description R2 statistic for significance test.

License GPL (>=3)

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.2

NeedsCompilation no

R topics documented:

`cc_trf` *cc_trf function*

Description

This function transforms the predictive ability (R2) and its standard error (se) between the observed scale and liability scale

Usage

```
cc_trf(R2, se, K, P)
```

Arguments

`R2`	R2 or Coefficient of determination on the observed or liability scale
`se`	Standard error of R2
`K`	Population prevalence
`P`	The ratio of cases in the study samples

References

Lee, S. H., Goddard, M. E., Wray, N. R., and Visscher, P. M. A better coefficient of determination for genetic profile analysis. Genetic epidemiology,(2012). 36(3): p. 214-224.

Examples

```
To get the transformed R2
 output=cc_trf(0.06, 0.002, 0.05, 0.05)

 output$R2l (transformed R2 to the liability scale)
 0.2679337

 output$sel (transformed se to the liability scale)
 0.008931123

 output$R2O (transformed R2 to the observed scale)
 0.01343616

 output$seO (transformed se to the observed scale)
 0.000447872
```

`olkin_beta1_2` *olkin_beta1_2 function*

Description

This function derives Information matrix for beta1^2 and beta2^2 where beta1 and 2 are regression coefficients from a multiple regression model, i.e. y = x1•beta1 + x2•beta2 + e, where y, x1 and x2 are column-standardised, (i.e. in the context of correlation coefficients,see Olkin and Finn 1995).

Usage

```
olkin_beta1_2(omat, nv)
```

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

References

Olkin, I. and J.D. Finn, Correlations redux. Psychological Bulletin, 1995. 118(1): p. 155. Momin, M.M., Lee, S. Wray, N. and S. Lee, S.H. The variance and covariance of the coefficients of determination for genetic profile analysis (will be subbitted soon)

Examples

```
To get information (variance-covariance) matrix of beta1^2 and beta2^2 where
beta1 and 2 are regression coefficients from a multiple regression model.

dat=read.table("test_ukbb_thresholds_scaled") (see example files)
omat=cor(dat)[1:3,1:3]
omat
1.0000000 0.1958636 0.1970060
0.1958636 1.0000000 0.9981003
0.1970060 0.9981003 1.0000000

nv=length(dat$V1)
output=olkin_beta1_2(omat,nv)
output

output$info (2x2 information (variance-covariance) matrix)
0.04146276 0.08158261
0.08158261 0.16111124

output$var1 (variance of beta1^2)
0.04146276

output$var2 (variance of beta2^2)
0.1611112

output$var1_2 (variance of difference between beta1^2 and beta2^2)
0.03940878
```

`olkin_beta_inf` *olkin_beta_inf function*

Description

This function derives Information matrix for beta1 and beta2 where beta1 and 2 are regression coefficients from a multiple regression model, i.e. y = x1•beta1 + x2•beta2 + e, where y, x1 and x2 are column-standardised (see Olkin and Finn 1995).

Usage

```
olkin_beta_inf(omat, nv)
```

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

References

Olkin, I. and J.D. Finn, Correlations redux. Psychological Bulletin, 1995. 118(1): p. 155.

Examples

```
To get information (variance-covariance) matrix of beta1 and beta2 where
beta1 and 2 are regression coefficients from a multiple regression model.

dat=read.table("test_ukbb_thresholds_scaled") (see example files)
omat=cor(dat)[1:3,1:3]
omat
1.0000000 0.1958636 0.1970060
0.1958636 1.0000000 0.9981003
0.1970060 0.9981003 1.0000000

nv=length(dat$V1)
output=olkin_beta_inf(omat,nv)
output

output$info (2x2 information (variance-covariance) matrix)
0.2531406 -0.2526212
-0.2526212  0.2530269

output$var1 (variance of beta1)
0.2531406

output$var2 (variance of beta2)
0.2530269

output$var1_2 (variance of difference between beta1 and beta2)
1.01141
```

---

r2_diff *r2_diff function*

Description

This function estimates var(R2(y~x[,v1]) - R2(y~x[,v2])) where R2 is the R squared value 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

```
r2_diff(dat, v1, v2, nv)
```

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

Examples

```
To get the test statistics for the difference between R2(y~x[,v1]) and
R2(y~x[,v2]). (here we define R2_1=R2(y~x[,v1])) and R2_2=R2(y~x[,v2])))

dat=read.table("test_ukbb_thresholds_scaled") (see example files)
nv=length(dat$V1)
v1=c(1)
v2=c(2)
output=r2_diff(dat,v1,v2,nv)
output

r2redux output

output$rsq1 (R2_1)
0.03836254

output$rsq2 (R2_2)
0.03881135

output$var1 (variance of R2_1)
0.0001437583

output$var2 (variance of R2_2)
0.0001452828

output$var_diff (variance of difference between R2_1 and R2_2)
5.678517e-07

output$r2_based_p (p-value for significant difference between R2_1 and R2_2)
0.5514562

output$mean_diff (differences between R2_1 and R2_2)
-0.0004488044

output$upper_diff (upper limit of 95% CI for the difference)
0.001028172

output$lower_diff (lower limit of 95% CI for the difference)
-0.001925781



To get the test statistics for the difference between R2(y~x[,v1]+x[,v2]) and
R2(y~x[,v2]). (here R2_1=R2(y~x[,v1]+x[,v2]) and R2_2=R2(y~x[,v1]))

dat=read.table("test_ukbb_thresholds_scaled") (see example files)
nv=length(dat$V1)
```

```
v1=c(1,2)
v2=c(1)
output=r2_diff(dat,v1,v2,nv)
output

r2redux output

output$rsq1 (R2_1)
0.03896678

output$rsq2 (R2_2)
0.03836254

output$var1 (variance of R2_1)
0.0001475195

output$var2 (variance of R2_2)
0.0001437583

output$var_diff (variance of difference between R2_1 and R2_2)
2.321425e-06

output$r2_based_p (p-value for significant difference between R2_1 and R2_2)
0.4369177

output$mean_diff (differences between R2_1 and R2_2)
0.0006042383

output$upper_diff (upper limit of 95% CI for the difference)
0.004887989

output$lower_diff (lower limit of 95% CI for the difference)
-0.0005574975
```

r2_enrich_beta *r2_enrich_beta*

Description

This function estimates var(t1/exp) - (t2/(1-exp)), where t1 = beta1^2 and t2 = beta2^2, and beta1 and 2 are regression coefficients from a multiple regression model, i.e. y = x1•beta1 + x2•beta2 +e, where y, x1 and x2 are column-standardised (see Olkin and Finn 1995). y is N by 1 matrix having the dependent variable, and x1 is N by 1 matrix having the ith explanatory variables. x2 is N by 1 matrix having the jth explanatory variables. v1 and v2 indicates the ith and jth column in the data (v1 or v2 should be a single interger between 1 - M, see Arguments below). Note that r2_enrich (above) and r2_enrich_beta is equivalent (identical p-value derived).

Usage

```
r2_enrich_beta(dat, v1, v2, nv, exp1)
```

Arguments

`dat`	N by (M+1) matrix having variables in the order of cbind(y,x)
`nv`	sample size
`exp1`	The expectation of the ratio (e.g. ratio of # SNPs in genomic partitioning)
`v1/v2`	These can be set as v1=1 and v2=2, v1=2 and v2=1, v1=3 and v2=2, or any combination as long as the value is between 1 - M

References

Olkin, I. and J.D. Finn, Correlations redux. Psychological Bulletin, 1995. 118(1): p. 155.

Examples

```
To get the test statistic for the ratio which is significantly
different from the expectation.
var[(t1/exp) - (t2/(1-exp))], where t1 = beta1^2 and t2 = beta2^2.
beta1 and beta2 are regression coefficients from a multiple regression model,
i.e. y = x1•beta1 + x2•beta2 +e, where y, x1 and x2 are column-standardised

dat=read.table("test_ukbb_enrichment_choles") (see example file)
nv=length(dat$V1)
v1=c(1)
v2=c(2)
expected_ratio=0.04
output=r2_enrich_beta(dat,v1,v2,nv,expected_ratio)
output

r2redux output

output$beta1_sq (t1)
0.01118301

output$beta2_sq (t2)
0.004980285

output$var1 (variance of t1)
7.072931e-05

output$var2 (variance of t2)
3.161929e-05

output$var1_2 (variance of difference between t1 and t2)
0.000162113

output$cov (covariance between t1 and t2)
-2.988221e-05

output$enrich_p2 (p-value for testing the difference between t1/exp and t2/(1-exp))
0.1997805

output$mean_diff (difference between t1/exp and t2/(1-exp))
0.2743874

output$var_diff (variance of difference, t1/exp - t2/(1-exp))
```

```
0.04579649

output$upper_diff (upper limit of 95% CI for the mean difference)
0.6938296

output$lower_diff (lower limit of 95% CI for the mean difference)
-0.1450549
```

`r2_var` *r2_var function*

Description

This function estimates var(R2(y~x[,v1])) where R2 is the R squared value of the model, where R2 is the R squared value 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

```
r2_var(dat, v1, nv)
```

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

Examples

```
To get the test statistics for R2(y~x[,v1])

dat=read.table("test_ukbb_thresholds_scaled") (see example file)
nv=length(dat$V1)
v1=c(1)
output=r2_var(dat,v1,nv)

r2redux output

output$rsq (R2)
0.03836254

output$var (variance of R2)
0.0001437583

output$r2_based_p (P-value under the null hypothesis, i.e. R2=0)
1.213645e-10

output$upper_r2 (upper limit of 95% CI for R2)
0.06435214
```

```
output$lower_r2 (lower limit of 95% CI for R2)
0.01763347



To get the test statistic for R2(y~x[,v1]+x[,v2]+x[,v3])

dat=read.table("test_ukbb_thresholds_scaled") (see example file)
nv=length(dat$V1)
v1=c(1,2,3)
outout=r2_var(dat,v1,nv)
output

r2redux output

output$rsq (R2)
0.03917668

output$var (variance of R2)
0.0001499374

output$r2_based_p (R2 based P-value)
7.461267e-11

output$upper_r2 (upper limit of 95% CI for R2)
0.06538839

output$lower_r2 (lower limit of 95% CI for R2)
0.01821657
```

Index